

NBER WORKING PAPER SERIES

EVALUATING THE EFFECT OF AN
ANTIDISCRIMINATION LAW USING
A REGRESSION-DISCONTINUITY DESIGN

Jinyong Hahn
Petra Todd
Wilbert Van der Klaauw

Working Paper 7131
<http://www.nber.org/papers/w7131>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1999

A previous version of this paper has been circulated under the title “Estimation of Treatment Effects with a Quasi-Experimental Regression-Discontinuity Design: with Application to Evaluating the Effect of Federal Antidiscrimination Laws on Minority Employment in Small U.S. Firms.” We thank Joshua Angrist, James Heckman, Guido Imbens, Alan Kruegger, Tony Lancaster, Sendhil Mullainathan, and Ken Wolpin for helpful comments. The paper has also benefited from comments received at the 1997 Midwestern Econometrics Group conference, the 1997 AEA meetings, the 1998 Econometric Society Summer Meetings in Montreal, the University of Michigan, the University of Rochester, the University of Wisconsin, the joint Harvard/MIT econometrics conference, and the joint Brown-Yale-NYU-Penn-JHU labor conference. Van der Klaauw thanks the C.V. Starr Center of Applied Economics at NYU for research support. Petra Todd’s participation was supported by NSF grant #SBR-9730688. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

© 1999 by Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Evaluating the Effect of an Antidiscrimination
Law using a Regression-Discontinuity Design
Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw
NBER Working Paper No. 7131
May 1999
JEL No. C14, C51

ABSTRACT

The regression discontinuity (RD) data design is a quasi-experimental design with the defining characteristic that the probability of receiving treatment changes discontinuously as a function of one or more individual characteristics. This data design occasionally arises in economic and other applications but is only infrequently exploited in evaluating the effects of a treatment. We consider the problem of identification and estimation of treatment effects under a RD data design. We offer an interpretation of the IV or so-called Wald estimator as a regression discontinuity estimator. We propose nonparametric estimators of treatment effects and present their asymptotic distribution theory. Then we apply the estimation method to evaluate the effect of EEOC-coverage on minority employment in small U.S. firms.

Jinyong Hahn
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
hahn@econ.sas.upenn.edu

Petra Todd
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
and NBER
petra@athena.sas.upenn.edu

Wilbert van der Klaauw
Department of Economics
University of North Carolina
Chapel Hill, NC 27599

1 Introduction

Quasi-experimental methods, like the regression-discontinuity method, are only infrequently considered in the evaluation literature as a separate method of evaluation. The focus is usually either on purely experimental or purely observational methods. Data generated by a regression-discontinuity design shares features with both experimental and observational data. In an experiment, treatment is assigned by a randomization device, which guarantees that persons in the treated group and in the control group are comparable. A regression-discontinuity (RD) design is similar to an experiment in that there is also a known rule that influences how persons are assigned to treatment. However, assignment to treatment is not random and persons who receive treatment may differ systematically from those who do not. In this sense, data from a RD design are similar to observational data.

In one of the first applications of the RD method and the first paper to introduce the design, Thistlethwaite and Campbell (1960) estimate the effect that receipt of a National Merit Award has on students' success in obtaining additional college scholarships and on their career aspirations. They observe that the awards are given on the basis of whether a test score exceeds a threshold, so one can take advantage of knowing the cut-off point to learn about treatment effects for persons near the cut-off. Berk and Rauma (1983) take a similar approach in analyzing the effect of extending unemployment benefits to released prisoners on recidivism rates, where benefits are given only to prisoners who worked a minimum number of hours while in prison. Van der Klaauw (1996) examines the effect that college scholarships awarded at the time of admission have on students' decisions to attend a particular college. He uses the fact that the value of an index based on a combination of the student's grade-point average and SAT score partly determines in a discontinuous manner whether a fellowship is awarded as well as the amount of fellowship. In another recent application, Angrist and Lavy (1996) estimate the effect of classroom size on student test scores. One of the factors determining class size in their data is a rule stipulating that another classroom be added whenever the average classroom size crosses a threshold level. Finally, Black (1996) uses an RD approach to estimate parents' willingness to pay for higher quality schools by comparing housing prices near geographic school attendance boundaries. In all these examples, the treatment variable changes discontinuously as a function of one or more underlying variables, which is the defining characteristic of regression discontinuity data designs.

Although there have already been many applications of RD methods, important questions still remain concerning sources of identification and ways of estimating treatment effects under minimal parametric restrictions. Trochim (1984) discusses alternative parametric and semiparametric RD estimators that have been proposed in the statistics literature.¹ Van der Klaauw (1996) considers

¹Trochim (1984) also discusses several applications of these methods in educational research.

identification and estimation in a semiparametric model under a constant treatment effect assumption. In this paper, we consider the more general case which allows for variable treatment effects. We demonstrate that treatment effects are typically nonparametrically unidentified in regression discontinuity models but that a weak form of identification can be achieved through a functional form restriction. The restriction is unusual in that it requires imposing continuity assumptions in order to take advantage of the known discontinuity in the treatment assignment mechanism.² We propose two estimators and apply these to evaluate the effect of an antidiscrimination law.

The paper develops as follows. Section 2 discusses the model, the parameters of interest, and different types of RD designs. Section 3 considers alternative sources of identifying information. Section 4 proposes consistent estimators and provides the distribution theory. This section also draws a comparison between IV estimators and RD estimators by observing that under certain conditions a kernel-based RD estimator is numerically equivalent to a standard IV estimator. Section 5 applies the methods to study the relationship between firm size and minority employment patterns using data from the NLSY (National Longitudinal Survey of Youth). In the U.S., only firms with 15 or more employees are subject to federal antidiscrimination laws.³ We can therefore exploit the discontinuity of this rule to analyze whether smaller firms that are not subject to these regulations tend to hire smaller proportions of minority workers. Section 6 summarizes our main findings.

2 The Regression-Discontinuity Design

The goal of an evaluation is to determine the effect that some binary treatment variable x_i has on an outcome y_i . The evaluation problem arises because persons either receive or do not receive treatment and no individual is observed in both states at the same time. Let y_{1i} denote the outcome with treatment and y_{0i} that in the absence of treatment, and let $x_i = 1$ if treatment is received and $x_i = 0$ otherwise. The model for the observed outcome can be written as

$$y_i = \alpha_i + x_i \cdot \beta_i, \tag{1}$$

where $\alpha_i \equiv y_{0i}$, $\beta_i \equiv y_{1i} - y_{0i}$. The entire distribution of treatment impacts may be of interest in an evaluation, but because α_i is not observed for any $x_i = 1$ person it may only be possible to reliably estimate limited aspects of the impact distribution. Two parameters of interest that receive much attention in the evaluation literature are the mean effect of *treatment on the treated* $\Delta_{x=1} \equiv E[\beta_i | x_i = 1]$ and the *mean effect of treatment on randomly assigned individual* $\Delta =$

²While our approach is unusual in its reliance on a continuity assumption for identification, the type of assumption we make is commonly invoked under other estimation approaches.

³Specifically, to Title VII of the Civil Rights Act of 1965 and to a 1972 Amendment to the Act.

$$E[y_{1i} - y_{0i}] = E[\beta_i].^4$$

If the data are purely observational, then little may be known a priori about the selection into treatment process. With data from a RD design, the analyst has some information about the treatment assignment mechanism. There are two main types of discontinuity designs considered in the literature - the *sharp design* and the so-called *fuzzy design* (see *e.g.* Trochim, 1984). With a sharp design, treatment x_i is known to depend in a deterministic way on some observable variable z_i , $x_i = f(z_i)$, where z takes on a continuum of values and the point z_0 where the function $f(z)$ is discontinuous is assumed to be known. In the empirical work of section 6, we consider data from a sharp design.

With a fuzzy design, x_i is a random variable given z_i , but the conditional probability $f(z) \equiv E[x_i | z_i = z] = \Pr[x_i = 1 | z_i = z]$ is known to be discontinuous at z_0 . For example, in the application of Van der Klaauw (1996), the probability that a student receives financial aid changes discontinuously as a function of a known index of the student's GPA and SAT scores. However, there are other factors, some of which are unobserved, which affect the financial aid decision, so the data fits a fuzzy rather than a sharp design.⁵

Next we consider formally why knowing that the probability of receiving treatment changes discontinuously as a function of an underlying variable is a valuable source of identifying information.

3 Sources of Identification

3.1 Sharp Design

To simplify the exposition of ideas, consider the special case of a simple sharp discontinuity design. Treatment is assigned based on whether z_i crosses a threshold value z_0 :

$$\begin{aligned} x_i &= 1 \text{ if } z_i > z_0 \\ &= 0 \text{ if } z_i \leq z_0. \end{aligned}$$

As z may be correlated with the outcome variable, the assignment mechanism is clearly not random and a comparison of outcomes between persons who received and did not receive treatment will generally be a biased estimator of treatment impacts.⁶ However, we may have reason to believe that persons close to the threshold z_0 are similar. If so, we may view the design as almost experimental near z_0 .

⁴See, *e.g.*, Peters (1941), Belson (1956), Rosenbaum and Rubin (1985) or Heckman and Robb (1985).

⁵The model of Angrist and Lavy (1996) also falls under the fuzzy design. Both Angrist and Lavy, and Van der Klaauw (1996) analyze the case of multiple treatment dose levels, which can be viewed as an extension of the uniform treatment dose case.

⁶Note that, as z is assumed to be observed, assignment in the sharp RD design is a special case of selection on observables.

To make this argument rigorous, let $e > 0$ denote an arbitrary small number. Comparing conditional means for persons who received and did not receive treatment gives

$$E[y_i | z_i = z_0 + e] - E[y_i | z_i = z_0 - e] = E[\beta_i | z_i = z_0 + e] + E[\alpha_i | z_i = z_0 + e] - E[\alpha_i | z_i = z_0 - e].$$

When persons near to the threshold are similar, we would expect $E[\alpha_i | z_i = z_0 + e] \cong E[\alpha_i | z_i = z_0 - e]$. This intuition motivates the following assumptions:

Condition (C1) $E[\alpha_i | z_i = z]$ is continuous in z at z_0 .⁷

Condition (C2) The limit $\lim_{e \rightarrow 0^+} E[\beta_i | z_i = z_0 + e]$ is well defined and of interest.

Under Conditions (C1) and (C2), it is easy to see that

$$\lim_{e \rightarrow 0^+} \{E[y_i | z_i = z_0 + e] - E[y_i | z_i = z_0 - e]\} = E[\beta_i | z_0]. \quad (2)$$

By comparing persons arbitrarily close to the point z_0 who did and did not receive treatment, we can in the limit identify $E[\beta_i | z_i = z_0]$. Without further assumptions such as the “common effect” assumption, treatment effects can only be identified at $z_i = z_0$.⁸ Conditions (C1) and (C2) are all that is required for identification.

It is a limitation of a RD design that we can only learn about treatment effects for persons with z values near the point of discontinuity. An advantage of randomized data is that randomization identifies treatment impacts over the full support of z . With data from a RD design, treatment effects can only be identified over a wider range of the support of z by increasing the number of discontinuity points. As the number of points approaches infinity, a discontinuity design approximates the conditions of a randomized experiment.⁹ If treatment effects were locally constant, say within quintiles of z , then a RD design with a fixed number of discontinuity points would allow identification of the range of treatment impacts. Multiple discontinuity points also allow a test of the common effect assumption.

3.2 Fuzzy Design

The fuzzy design differs from the sharp design in that the treatment assignment is not a deterministic function of z_i ; there are additional variables which are unobserved by the evaluator that

⁷Throughout this paper, we also assume that the density of z_i is positive in the neighborhood containing z_0 .

⁸This notion of identification is similar to the “identification at infinity” idea put forth in Heckman (1990), although it differs from his case in that we do not require that z_i has support over the whole real line. Because of the local nature of our source of identification, the semiparametric information bound is by definition zero: a root- n estimator cannot exist.

⁹The number of points has to go to infinity in such a way that the limit densely covers the range of values of z for which the treatment effect is of interest.

determine assignment to treatment. The common feature it shares with the sharp design is that the probability of receiving treatment, $\Pr[x_i = 1|z_i]$, viewed as a function of z_i , is discontinuous at z_0 .¹⁰ We consider identification of treatment impacts under different assumptions about the heterogeneity of impacts.

3.2.1 Common Treatment Effects

Suppose that the treatment effect is constant across different individuals. Write β for the common value. The mean difference in outcomes for persons above and below the discontinuity point is

$$\beta \cdot \{E[x_i|z_i = z_0 + e] - E[x_i|z_i = z_0 - e]\} + E[\alpha_i|z_i = z_0 + e] - E[\alpha_i|z_i = z_0 - e].$$

Under (C1), we have

$$\lim_{e \rightarrow 0^+} \{E[y_i|z_i = z_0 + e] - E[y_i|z_i = z_0 - e]\} = \beta \cdot \lim_{e \rightarrow 0^+} \{E[x_i|z_i = z_0 + e] - E[x_i|z_i = z_0 - e]\}.$$

Thus, we can identify β by

$$\frac{\lim_{e \rightarrow 0^+} E[y_i|z_i = z_0 + e] - \lim_{e \rightarrow 0^+} E[y_i|z_i = z_0 - e]}{\lim_{e \rightarrow 0^+} E[x_i|z_i = z_0 + e] - \lim_{e \rightarrow 0^+} E[x_i|z_i = z_0 - e]}. \quad (3)$$

The denominator is nonzero because the fuzzy RD design ensures that $\Pr[x_i = 1|z_i = z]$ is discontinuous at z_0 .

3.2.2 Variable Treatment Effects

Now consider the question of identification when treatment effects are heterogeneous. To identify $E[\beta_i|z_i = z_0]$ using the same strategy as in the constant treatment effect case, we assume that conditional on z_i the other variables affecting whether a person receives treatment are independent of the treatment effect:

Condition (C3) x_i is independent of β_i conditional on z_i near z_0 : $x_i \perp \beta_i | z_i$.

Although (C3) allows for more generality than the common treatment impact assumption, it still restricts the form of the heterogeneity in a way that may not be acceptable in application.¹¹ On the other hand, the condition is slightly weaker than the conditional ignorability assumption commonly invoked in the statistics literature on matching (e.g. Rosenbaum and Rubin, 1983) in that we do not require that y_{0i} and y_{1i} are conditionally independent of x_i .¹²

¹⁰This probability is often referred to as the propensity score in the statistics literature. See *e.g.*, Rosenbaum and Rubin (1983).

¹¹Condition (C3) maintains that individuals do not select into treatment on the basis of anticipated gains from the treatment, an assumption that is criticized in Heckman and Smith (1997).

¹²To some extent, this technicality is substantive. For example, the common effect model satisfies (C3) trivially but not necessarily the stronger assumption.

Condition (C3) implies that

$$E[x_i \cdot \beta_i | z_i = z \pm e] = E[\beta_i | z_i = z \pm e] \cdot E[x_i | z_i = z \pm e]$$

Combined with (C1) and (C2), we obtain

$$\begin{aligned} \lim_{e \rightarrow 0^+} \{E[y_i | z_i = z_0 + e] - E[y_i | z_i = z_0 - e]\} \\ = E[\beta_i | z_i = z_0] \cdot \lim_{e \rightarrow 0^+} \{E[x_i | z_i = z_0 + e] - E[x_i | z_i = z_0 - e]\}. \end{aligned}$$

We can therefore again identify $E[\beta_i | z_i = z_0]$ by (3).

To examine the consequence of dropping (C3), we examine an alternative case that is sometimes considered in the evaluation literature. Suppose, as in Imbens and Angrist (1994) or Angrist, Imbens, and Rubin (1994), that for each observation i , treatment assignment is a deterministic function of z , but that the function is different for different persons or groups of persons. For example, consider the case where college fellowships are awarded on the basis of SAT scores, but the threshold cutoffs for awarding fellowships are different for different ethnic groups. If the analyst does not have this information, the assignment mechanism may appear to be random. Consider the following set of assumptions on impacts and treatment assignment:

- (i) $(\beta_i, x_i(z))$ is jointly independent of z_i near z_0 .
- (ii) $x_i(z_0 + e) \geq x_i(z_0 - e)$ for $e > 0$ sufficiently small.

Invoking the reasoning in Imbens and Angrist (1994), we obtain

$$\begin{aligned} E[x_i \cdot \beta_i | z_i = z_0 + e] - E[x_i \cdot \beta_i | z_i = z_0 - e] \\ = E[x_i(z_0 + e) \cdot \beta_i] - E[x_i(z_0 - e) \cdot \beta_i] \\ = \Pr[x_i(z_0 + e) - x_i(z_0 - e) = 1] \cdot E[\beta_i | x_i(z_0 + e) - x_i(z_0 - e) = 1] \\ = \{E[x_i | z_i = z_0 + e] - E[x_i | z_i = z_0 - e]\} \cdot E[\beta_i | x_i(z_0 + e) - x_i(z_0 - e) = 1]. \end{aligned}$$

Therefore, under Conditions (C1) and (C3'), (3) identifies the local average treatment effect (LATE) at z_0 ,

$$\lim_{e \rightarrow 0^+} E[\beta_i | x_i(z_0 + e) - x_i(z_0 - e) = 1].$$

That is, the estimator identifies the local average treatment impact for the subgroup of persons for whom treatment changes discontinuously at z_0 .

We can relax Condition (i) of (C3)' and instead impose a series of continuity restrictions. Because the notation is cumbersome, we defer discussion of this case to the appendix.

3.2.3 Nonparametric Nonidentification

In all the above examples, we have shown how identification of treatment effects can be achieved through a local continuity restriction on $E[\alpha_i|z_i]$ and a known discontinuity in $E[x_i|z_i]$. In this section, we show that these restrictions are necessary and that without them the model is nonparametrically unidentified. We can put the model for outcomes in more familiar econometric notation by writing

$$y_i = \alpha(z_i) + \beta \cdot x_i + v_i$$

where α_i is a function of the z_i variable and of some other unobservables v_i , $\alpha_i = \alpha(z_i) + v_i$ and $\alpha(z_i) = E[\alpha_i|z_i]$.¹³ The selection model can be written as

$$\begin{aligned} x_i &= 1 \text{ if } f(z_i) + u_i > 0 \\ &= 0 \text{ else.} \end{aligned}$$

We argue that the usual conditional mean independence restriction, $E[v_i|z_i] = 0$, is not sufficient for identification of the treatment effect, even for the common treatment effect case. For this purpose consider another DGP, where we have

$$y_i = \alpha^*(z_i) + 0 \cdot x_i + v_i^*,$$

and where

$$\alpha^*(z_i) = \alpha(z_i) - \beta \cdot E[x_i|z_i], \quad v_i^* = v_i + \beta \cdot \{x_i - E[x_i|z_i]\}.$$

These two models are equivalent except that the treatment effect in the former case is β whereas in the latter case it is equal to 0. We cannot distinguish the models in the population if $E[v_i|z_i] = 0$ is the only restriction available.

There are different types of restrictions that can be imposed to overcome this problem. If it is known that $E[\alpha_i|z_i]$ is continuous at a point z_0 but $E[x_i|z_i]$ is discontinuous, then the models can be distinguished ($\alpha(z_i)$ is continuous but $\alpha^*(z_i)$ is not). This is the strategy followed here. Alternative strategies not considered here could be based on standard types of exclusion restrictions - variables known to affect treatment assignment but not the average outcome. In that case, v_i^* would depend on the excluded variable whereas v_i would not.¹⁴ Still another possibility is to impose functional form restrictions on $\alpha(z_i)$ so that the treatment impact parameter could be identified solely off the functional form.

¹³For notational simplicity, we consider the case where z_i is the only observable variable.

¹⁴Imposing continuity on $\alpha(z_i)$ also implicitly imposes an exclusion restriction in that it excludes $1(z_i > z_0)$ as a variable in the outcome equation. We thank a referee for this point.

4 Estimation

An estimation approach that has been adopted in the literature for the sharp design is to assume (in addition to continuity) a flexible parametric specification for $g(z) = E[\alpha_i | z_i]$ and add this as a ‘control function’ to the regression of y_i on x_i . Van der Klaauw (1996), for example, assumes the functional form of g to belong to the class of polynomial functions. For the fuzzy design he proposes a similar approach but where x_i in the control function-augmented regression equation is now replaced by a first stage estimate of $E[x_i | z_i]$. Although this would result in a consistent estimator of β under correct specification, it may be more fragile to misspecification. For this reason, we consider a more robust nonparametric strategy based on (3).

For both the sharp design and fuzzy design, (3) identifies the treatment effect at $z = z_0$. Thus, given consistent estimators of the four one-sided limits in (3), the treatment effect can be consistently estimated. In principle, we can use any nonparametric estimator to estimate the limits. We first consider one-sided kernel estimation and observe that under certain conditions an estimate based on kernel regression will be numerically equivalent to a standard Wald estimator. We then argue that such an estimator may have a poor finite sample property due to the boundary problem and propose to avoid the boundary problem by using local linear nonparametric regression (LLR) methods. We describe the asymptotic distribution theory for the estimator of (3) based on local regression.

4.1 A Wald Estimator based on Kernel Regression

Let $\hat{\beta}$ denote an estimator for the treatment impact based on equation (3), where

$$\hat{\beta} = \frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-},$$

and where \hat{y}^+ , \hat{y}^- , \hat{x}^+ , and \hat{x}^- are estimators for each of the limit expressions. One way to estimate the limits is by one-sided kernel regression.

$$\hat{y}^+ = \frac{\sum_i y_i \cdot 1(z_i > z_0) K\left(\frac{z_i - z_0}{h}\right)}{\sum_i 1(z_i > z_0) K\left(\frac{z_i - z_0}{h}\right)}, \quad \hat{y}^- = \frac{\sum_i y_i \cdot 1(z_i \leq z_0) K\left(\frac{z_i - z_0}{h}\right)}{\sum_i 1(z_i \leq z_0) K\left(\frac{z_i - z_0}{h}\right)},$$

$$\hat{x}^+ = \frac{\sum_i x_i \cdot 1(z_i > z_0) K\left(\frac{z_i - z_0}{h}\right)}{\sum_i 1(z_i > z_0) K\left(\frac{z_i - z_0}{h}\right)}, \quad \hat{x}^- = \frac{\sum_i x_i \cdot 1(z_i \leq z_0) K\left(\frac{z_i - z_0}{h}\right)}{\sum_i 1(z_i \leq z_0) K\left(\frac{z_i - z_0}{h}\right)},$$

where $K(\cdot)$ is a kernel function, $1(\cdot)$ is an indicator function that equals one if the condition in parentheses is true and zero else, and h is a bandwidth parameter.

Consider the special case where we use kernel regression estimators based on one-sided uniform kernels.¹⁵ Given appropriate bandwidths h_+ and h_- , we would estimate the limits by

$$\hat{y}^+ = \frac{\sum_i y_i \cdot 1(z_0 < z_i < z_0 + h_+)}{\sum_i 1(z_0 < z_i < z_0 + h_+)}, \quad \hat{y}^- = \frac{\sum_i y_i \cdot 1(z_0 - h_- < z_i < z_0)}{\sum_i 1(z_0 - h_- < z_i < z_0)},$$

¹⁵For the uniform kernel, $K(u) = 1/2$ if $|u| \leq 1$, $K(u) = 0$ else.

and

$$\hat{x}^+ = \frac{\sum_i x_i \cdot 1(z_0 < z_i < z_0 + h_+)}{\sum_i 1(z_0 < z_i < z_0 + h_+)}, \quad \hat{x}^- = \frac{\sum_i x_i \cdot 1(z_0 - h_- < z_i < z_0)}{\sum_i 1(z_0 - h_- < z_i < z_0)}.$$

Now, let

$$\begin{aligned} w_i &= 1 \text{ if } z_0 < z_i < z_0 + h_+ \\ &= 0 \text{ if } z_0 - h_- < z_i < z_0 \end{aligned}$$

and observe that

$$\frac{\sum_i y_i \cdot 1(z_0 < z_i < z_0 + h_+)}{\sum_i 1(z_0 < z_i < z_0 + h_+)}, \quad \frac{\sum_i x_i \cdot 1(z_0 < z_i < z_0 + h_+)}{\sum_i 1(z_0 < z_i < z_0 + h_+)}$$

are the averages of y_i and x_i such that $w_i = 1$. Similarly,

$$\frac{\sum_i y_i \cdot 1(z_0 - h_- < z_i < z_0)}{\sum_i 1(z_0 - h_- < z_i < z_0)}, \quad \frac{\sum_i x_i \cdot 1(z_0 - h_- < z_i < z_0)}{\sum_i 1(z_0 - h_- < z_i < z_0)}$$

are the averages of y_i and x_i such that $w_i = 0$. Thus, the estimator is *numerically* equivalent to an IV estimator for the regression of y_i on x_i which uses w_i as an instrument, applied to the subsample where $z_0 - h_- < z_i < z_0 + h_+$. Denote this estimator by $\hat{\beta}_w$.

It is interesting to note that the regression discontinuity can ‘justify’ a Wald estimator even when the standard IV assumption that the instrument is uncorrelated with the error term is violated. To see this, put the model in more familiar econometric notation by writing

$$\alpha_i = E[\alpha_i] + v_i = \alpha + v_i.$$

Under the common treatment assumption, this yields a model for outcomes of the form

$$y_i = \alpha + x_i \cdot \beta + v_i.$$

Identification of β does not require that the error term v_i be uncorrelated with z_i . All that is required is smoothness conditions C1 and C2. As long as the researcher is willing to change the bandwidths appropriately as a function of the sample size, $\hat{\beta}_w$ is consistent. Thus the local Wald estimator is motivated by a different principle than is the usual Wald estimator, but for a particular choice of kernel and subsample they are numerically equivalent.

4.1.1 The Boundary Problem

Although $\hat{\beta}_w$ is numerically equivalent to a local Wald estimator, inference based on $\hat{\beta}_w$ will be different from that based on a Wald estimator. $\hat{\beta}_w$ will be asymptotically biased, as are many other nonparametric-regression-based estimators, whereas the Wald estimator itself is asymptotically

unbiased. The bias problem is exacerbated in the regression-discontinuity case due to the bad boundary behavior of the kernel regression estimator: at boundary points, the bias of the kernel regression estimator converges to zero at a slower rate than at interior points. The boundary bias problem arises because the kernel weighting becomes asymmetric near boundary points where there is only a one-sided interval over which to carry out the local averaging.¹⁶¹⁷

For our problem, all the points of estimation are at boundaries. The bias could be substantial in finite samples, especially when the conditional expectations $E[y_i|z_i]$ and/or $E[x_i|z_i]$ have nonzero one-sided derivatives around z_0 .¹⁸ It would be misleading to use the conventional confidence interval based on the asymptotic distribution of the (asymptotically unbiased) Wald estimator as the true coverage probability would be very different from the nominal coverage probability.

4.2 A Local Linear Regression Estimator

Because of the poor boundary performance of standard kernel estimators, we propose instead to estimate the limits by local linear regression (LLR), shown by Fan (1992) to have better boundary properties. Both kernel regression and LLR estimators can be viewed as special cases of more general local polynomial estimators.¹⁹ The local polynomial estimator of degree k for $\lim_{e \rightarrow 0^+} E[y_i|z_i = z_0 + e]$, for example, can be computed from the minimization problem

$$\min_{a,b} \sum_{i=1}^n I_i \cdot \left(y_i - a - b_1(z_i - z_0) - b_2(z_i - z_0)^2 - \dots - b_k(z_i - z_0)^k \right)^2 K\left(\frac{z_i - z_0}{h_n}\right),$$

where $I_i = 1(z_i > z_0)$, $K(\cdot)$ is a kernel function and $h_n > 0$ is a suitable bandwidth which converges to zero as $n \rightarrow \infty$.²⁰ The local polynomial regression estimator of the conditional mean is \hat{a} . The standard kernel regression estimator corresponds to a local polynomial of degree 0 and the LLR estimator to one of degree 1.

¹⁶Boundary points are points within one bandwidth of the boundary. Under conventional assumptions on the kernel function, the order of the bias of the standard kernel estimator is $O(h_n)$ at boundary points and $O(h_n^2)$ at interior points, where n is the sample size and h_n is a bandwidth sequence that goes to 0 as n gets large. See Härdle (1990) or Härdle and Linton (1994) for further discussion of the boundary bias problem.

¹⁷Porter (1998) recently proposed an alternative estimator for the sharp discontinuity design, common effect model for which the boundary bias problem does not exist. The estimator invokes slightly stronger regularity conditions than ours.

¹⁸The boundary bias formula of the kernel estimator reveals that the bias is the smallest when the conditional expectations $E[y_i|z_i]$ and/or $E[x_i|z_i]$ have one-sided derivatives around z_0 all equals to zeros. Then, the propensity score is approximately constant on either side of z_0 , and $\alpha(z)$ is constant near z_0 . We thus obtain the not-so-surprising conclusion that the local Wald estimator has a small bias only for the case where α_i has no correlation with z_i , i.e., the case where z_i is a proper instrument and the Wald assumption is exactly satisfied near the discontinuity.

¹⁹Local polynomial estimators were developed in the early literature by Cleveland (1979) and Stone (1977), and more recently considered in the econometrics literature by Gozalo and Linton (1997) and Heckman et. al. (1998).

²⁰Theorem 1 given below places restrictions on the bandwidth sequence.

Fan (1992) shows that for any odd order local polynomial estimator, the expression for the variance is the same as that of the next lower even order estimator but the bias is different. The odd order estimator has a lower order bias at boundary points. Thus, increasing the degree of the local polynomial regression from 0 to 1 brings an advantage in terms of solving the boundary bias problem without any disadvantage.²¹ The smaller bias associated with the LLR estimator implies that it is more rate-efficient than the kernel-based estimator. Another advantage of odd-order estimators emphasized by Fan is that the bias does not depend on the design density of the data. Because of these advantages, local linear methods are usually a better choice than standard kernel methods for nonparametric regression estimation.²²

The local polynomial estimator is asymptotically normal with a rate of convergence equal to $\sqrt{nh_n}$. The rate is slower than the parametric rate, because estimation is based only on the subsample near the discontinuity.

4.2.1 Distribution Theory

We next present the distribution theory for the LLR estimator $\hat{\beta}$ where the limits are estimated by local linear regression. As before, denote the limit expressions by

$$y^+ = \lim_{e \rightarrow 0^+} E[y_i | z_i = z_0 + e], \quad y^- = \lim_{e \rightarrow 0^+} E[y_i | z_i = z_0 - e],$$

$$x^+ = \lim_{e \rightarrow 0^+} E[x_i | z_i = z_0 + e], \quad x^- = \lim_{e \rightarrow 0^+} E[x_i | z_i = z_0 - e]$$

and the corresponding LLR estimators by $\hat{y}^+, \hat{y}^-, \hat{x}^+$, and \hat{x}^- . Also, define $m(z) = E[y_i | z_i = z]$ and $p(z) = E[x_i | z_i = z]$ and define the limits $\lim_{e \rightarrow 0^+} E[y_i | z_i = z + e]$, $\lim_{e \rightarrow 0^+} E[y_i | z_i = z - e]$, $\lim_{e \rightarrow 0^+} E[x_i | z_i = z + e]$, and $\lim_{e \rightarrow 0^+} E[x_i | z_i = z - e]$ by $m^+(z)$, $m^-(z)$, $p^+(z)$ and $p^-(z)$, respectively.

Additionally, define²³

$$\sigma^{2+}(z_0) = \lim_{e \rightarrow 0^+} \text{Var}[y_i | z_i = z_0 + e], \quad \sigma^{2-}(z_0) = \lim_{e \rightarrow 0^+} \text{Var}[y_i | z_i = z_0 - e]$$

$$\eta^+(z_0) = \lim_{e \rightarrow 0^+} \text{Cov}[y_i, x_i | z_i = z_0 + e], \quad \text{and} \quad \eta^-(z_0) = \lim_{e \rightarrow 0^+} \text{Cov}[y_i, x_i | z_i = z_0 - e].$$

To show the distribution theory for $\hat{\beta}$, we make use of the following assumptions:

²¹The advantage stems from the fact that local linear regression imposes an orthogonality condition between the regressors and the residuals that is not imposed under kernel regression. See Fan (1992).

²²One case where they would not be a better choice is if the function being estimated cannot be assumed to have a first derivative, in which case the kernel regression estimator is feasible but the local linear estimator is not. For most well-behaved functions, a polynomial of degree one suffices for local estimation. For a highly variable function a higher order polynomial regression, like local cubic regression, may provide a better fit. (See Fan (1996) for further discussion).

²³In general, $\text{Cov}[y_i, x_i | z_i] \neq 0$ even in the neighborhood of z_0 .

1. For $z > z_0$, $m(z)$ and $p(z)$ are twice continuously differentiable. There exists some $M > 0$ such that $|m^+(z)|$, $|m'^+(z)|$, $|m''^+(z)|$ and $|p^+(z)|$, $|p'^+(z)|$, $|p''^+(z)|$ are uniformly bounded on $(z_0, z_0 + M]$. Similarly, $|m^-(z)|$, $|m'^-(z)|$, $|m''^-(z)|$ and $|p^-(z)|$, $|p'^-(z)|$, $|p''^-(z)|$ are uniformly bounded on $[z_0 - M, z_0)$.
2. The limits $m^+(z_0)$, $m^-(z_0)$, $m'^+(z_0)$, $m'^-(z_0)$, $m''^+(z_0)$, $m''^-(z_0)$, $p^+(z_0)$, $p^-(z_0)$, $p'^+(z_0)$, $p'^-(z_0)$, $p''^+(z_0)$, and $p''^-(z_0)$ are well-defined and finite.
3. The density of z_i , $f(z)$, is continuous and bounded near z_0 . It is also bounded away from zero near z_0 .
4. $K(\cdot)$ is continuous, symmetric and nonnegative-valued with compact support.
5. $\sigma^2(z_i) = \text{Var}(y_i|z_i)$ is uniformly bounded near z_0 . Similarly, $\eta(z_i) = \text{Cov}(y_i, x_i|z_i)$ is uniformly bounded near z_0 . Furthermore, the limits $\sigma^{2+}(z_0)$, $\sigma^{2-}(z_0)$, $\eta^+(z_0)$, and $\eta^-(z_0)$ are well-defined and finite.
6. $\lim_{e \rightarrow 0^+} E \left[|y_i - m(z_i)|^3 \mid z_i = z_0 \pm e \right]$ is well-defined and finite.
7. The bandwidth sequence satisfies $h_n = \rho \cdot n^{-1/5}$ for some ρ .

Theorem 1: Asymptotic Distribution

Under assumptions 1-7,

$$n^{\frac{2}{5}} \left(\frac{\widehat{y}^+ - \widehat{y}^-}{\widehat{x}^+ - \widehat{x}^-} - \frac{y^+ - y^-}{x^+ - x^-} \right) \rightarrow \mathcal{N}(\mu_f, \Omega_f),$$

where

$$\mu_f \equiv \frac{1}{x^+ - x^-} (\rho^+ m''^+(z_0) - \rho^- m''^-(z_0)) - \frac{y^+ - y^-}{(x^+ - x^-)^2} (\rho^+ p''^+(z_0) - \rho^- p''^-(z_0)),$$

and

$$\begin{aligned} \Omega_f \equiv & \frac{1}{(x^+ - x^-)^2} (\omega^+ \sigma^{2+}(z_0) + \omega^- \sigma^{2-}(z_0)) \\ & - 2 \frac{y^+ - y^-}{(x^+ - x^-)^3} (\omega^+ \eta^+(z_0) + \omega^- \eta^-(z_0)) \\ & + \frac{(y^+ - y^-)^2}{(x^+ - x^-)^4} (\omega^+ p^+(z_0) (1 - p^+(z_0)) + \omega^- p^-(z_0) (1 - p^-(z_0))), \end{aligned}$$

and where

$$\rho^+ \equiv \frac{(\int_0^\infty u^2 K(u) du)^2 - (\int_0^\infty u^3 K(u) du) (\int_0^\infty u K(u) du) \varrho^2}{(\int_0^\infty u^2 K(u) du) (\int_0^\infty K(u) du) - (\int_0^\infty u K(u) du)^2} \frac{\varrho^2}{2},$$

$$\omega^+ \equiv \frac{\int_0^\infty ((\int_0^\infty s^2 K(s) ds) - (\int_0^\infty s K(s) ds) \cdot u)^2 K(u)^2 du}{f(z_0) \varrho \cdot [(\int_0^\infty u^2 K(u) du) (\int_0^\infty K(u) du) - (\int_0^\infty u K(u) du)^2]^2}.$$

with ρ^- and ω^- similarly defined but now with the integral in the limits of integration over $(-\infty, 0)$.

P roof. See Appendix A. ■

Theorem 1': Asymptotic Distribution

Under assumptions 1-7,

$$n^{\frac{2}{5}} (\widehat{y}^+ - \widehat{y}^- - (y^+ - y^-)) \rightarrow \mathcal{N}(\mu_s, \Omega_s),$$

where

$$\mu_s \equiv \rho^+ m''^+(z_0) - \rho^- m''^-(z_0), \quad \text{and} \quad \Omega_s \equiv \omega^+ \sigma^{2+}(z_0) + \omega^- \sigma^{2-}(z_0).$$

P roof. See Appendix A. ■

When there are multiple discontinuity points, then it is possible to test for whether the treatment effect is constant by testing the equivalence of the locally estimated treatment effects. As long as the test points are at least two bandwidths apart, treatment effect estimates are independent and a joint test of the null that treatment effects are equivalent across points can be based on a standard chi-square test.

5 Empirical Application: Evaluating the Effect of EEOC-coverage on Minority Employment Patterns in Small Firms

We now apply the methods developed above to analyze the effect of a federal antidiscrimination law on firm employment of minority workers. We use the fact that by U.S. law, firms with at least 15 employees are covered by Title VII of the Civil Rights Act of 1964 and by the 1972 amendment to the act, while firms with fewer than 15 employees are not.²⁴

While there have been a large number of studies of the effect of Title VII legislation over the last few decades, researchers still disagree as to how effective government legislation and litigation has been in increasing employment and earnings opportunities for minority workers. The lack of a

²⁴The 1972 amendment to the Civil Rights Act, also called the 1972 Equal Employment Opportunity Act, expanded coverage of Title VII from firms with 25 or more employees to firms with 15 or more employees.

consensus on this question probably stems in part from the difficulty of assessing the impact of laws that have near universal coverage. Two broad schools of thought have emerged from the literature. One emphasizes the importance of long-term secular trends in education, educational quality and migration as the predominant sources of black economic progress. (See *e.g.* Smith and Welch, 1986) The other assigns a greater role to government policies. (See *e.g.* Donohue and Heckman, 1991, Heckman and Payner, 1989, Chay, 1995, and Carrington, McCue, and Pierce, 1995.)

Many of the existing empirical studies identify the effects of discrimination laws either through a “before-after” or a “difference-in-difference” approach. For example, some studies establish a relationship between the timing of the introduction of a law and the timing of changes in employment practices. A potential problem with this kind of approach is that of disentangling the effect of legislation from other year effects and demographic trends. When data are available on a control group that is not subject to the law, then it is possible to control for time effects through a difference-in-difference approach. However, the validity of a difference-in-differences strategy depends crucially on whether the experiences of the control group accurately represent how the treated group (the group subject to the law change) would have fared in the absence of the legal intervention.

A recent study that examines a question of interest similar to ours using a difference-in-differences approach is that of Chay (1995). Chay analyzes the impact of the 1972 Amendment to the Civil Rights Act on minority employment and earnings outcomes in firms that were affected by the change in coverage (firms with 15-24 employees). Chay uses variation across industries in the fraction of employees in small establishments and across states in the employer coverage of Fair Employment Practice (FEP) laws to define treatment and control groups.²⁵ Changes over time in employment and earnings outcomes are then compared in the treatment and control groups. Chay’s study concludes that the change in the law led to annual increases in the percentage of black employment in small firms of about 0.5 percentage points, resulting in an accumulated increase of about 3.5 percentage points in the 1980 employment shares.

Another related study is that of Carrington, McCue, and Pierce (1995), which examines whether the introduction of Title VII of the Civil Rights Act and of affirmative action led to a migration of blacks and women towards larger firms (defined as firms with 20 or more employees) who were covered under the Act. Their study finds evidence that such migration did occur, leading over 10 years to substantial changes in employment patterns in small and large establishments.

To analyze whether the anti-discrimination law has had and continues to have an effect on

²⁵Because the CPS data Chay used did not allow workers employed in establishments with less than 15 employees to be distinguished from those working in establishment with 15-24 employees, he simply compares small firms (defined as those with less than 25 employees) with larger firms. While most states in the South did not have FEP laws at the time of the 1972 Amendment, so that firms with less than 25 employees were not covered by any law before 1972, most other states already had FEP laws in place covering firms smaller than those subject to Title VII.

the relative share of minority employment, we adopt a related but different estimation approach. Like Chay (1995) and Carrington, McCue, and Pierce (1995) we exploit variation in employment between employers covered and not covered by the law as determined by their number of employees. However, we use only cross-sectional variation. Because the law applies discontinuously as a function of firm size, an RD approach provides a straightforward way of evaluating the effect that coverage under the law has on minority employment. There are several advantages of using an RD estimator instead of a more traditional evaluation estimator. It can be justified under very weak assumptions and it bypasses many questions concerning model specification, as we discuss further below.

5.1 The Model and the Data

We obtain data on individuals and the sizes of the firms at which they work from the National Longitudinal Survey of Youth (NLSY). This panel data set contains standard demographic information as well as information on employment, earnings and education histories. In most years, the NLSY survey asks about the size of the firm at which the individual is employed and asks whether the firm operated at other locations.²⁶

The variables that we use are the race/ethnicity of the individual and the size of the firm at which the individual is employed. We define “minority” as black or Hispanic and combine women and men in the analysis. Only working persons are included in our sample and individuals at firms with more than one location are excluded.²⁷ The final sample consists of approximately 2000 observations followed over the years 1979 to 1993.²⁸ Table 1 provides descriptive statistics of the variables used in our study.

In our application, treatment x_i is a dummy variable for whether a firm is covered by the law. The dependent variable y_i is the fraction of minority workers hired by the firm. Let z_i denote the size of the firm. “Treatment” is received if the size of the firm has at least 15 employees:

$$\begin{aligned} x_i &= 1 \text{ if } z_i \geq 15, \\ &= 0 \text{ otherwise.} \end{aligned}$$

²⁶We would have preferred a dataset with more observations than the moderate number available in the NLSY, but most other datasets report firm size in intervals, which precludes the application of an RD evaluation method. These include the CPS (Current Population Survey), PSID (Panel Survey of Income Dynamics) and NLS (National Longitudinal Survey) datasets.

²⁷The law applies to the total number of employees in all locations combined. We do not know the exact number of employees at all other locations, only whether there were more or less than 1000 employees at the other locations. Since we can not determine whether the law applies, we excluded these persons.

²⁸The number of observations varies slightly from year to year because of attrition and as a consequence of our sample selection criteria. The question about firm size was not asked in years 1981 through 1985. See Appendix B for additional information on omitted observations.

Since treatment assignment depends only on z_i , the model fits within the sharp RD design framework.

We do not have data on individual firms but rather data on individual workers and the sizes of the firms at which they work.²⁹ From the data set, however, we can infer the fraction of minorities in firms of a given size, and the number of firms of a given size. By using the number of observations of a given firm size as weights in the local linear regression of the fraction of minorities on the firm sizes, we can recover the estimates that would have been obtained had we had access to individual firm data. For example, the local linear regression estimator of $\lim_{e \rightarrow 0^+} E(y_i | z_0 + e)$ that incorporates sampling weights c_i is given by

$$\frac{\sum_{i=1}^n c_i y(z_i) W_i}{\sum_{i=1}^n c_i W_i},$$

where

$$W_i = \frac{I_i K_i \sum_{j=1}^n I_j K_j^2 - I_i K_i \sum_{k=1}^n I_k K_k}{\sum_{k=1}^n I_k K_i \sum_{j=1}^n I_j K_j^2 - \sum_{k=1}^n I_k K_k \sum_{j=1}^n I_j K_j},$$

$I_i = 1(z_i \geq 15)$, $K_i = K((z_i - z_0)h^{-1})$, $y(z_i)$ is the mean fraction of minorities for observations with firm size z_i and c_i is the number of individual level observations reporting firm size z_i .³⁰ The estimators for the lower limit expression can be defined analogously. To estimate the treatment effect, we take the difference between the estimated conditional mean fraction of minorities employed at firms just above and just below the threshold value of 15.

There are many different variables that one would expect to affect the proportion of minorities hired by a firm. For example, there may be differential costs with more selective hiring. The size of the firm may be correlated with other variables such as the industrial sector to which the firm belongs and the geographic region in which it is located, both of which may in turn be correlated with the demographic composition of the applicant pool (see Brown and Medoff, 1989, and/or Holzer, 1996). Furthermore, in 20 states (mostly northern states) firms are also subject to state fair employment laws which are similar to the federal laws.

A key advantage of adopting an RD approach is that it does not require building a model of all the determinants of minority hirings. This is because the effects of all these other unobserved determinants are not discontinuous at 15 and so enter through α_i in (1). As long as their effect is continuous at 15, their effect differences out in estimation. In the case of state fair employment

²⁹This implies that our data set is subject to size-biased sampling, since it oversamples bigger firms and under-samples smaller ones. However, size-biased sampling does not pose any particular problem in our study since we are interested in the percentage of minorities conditional on a given firm size.

³⁰Sampling weights provided in the NLSY dataset to account for nonrandom sampling are used to reweight the data back to random proportions in constructing each of the $y(z_i)$ means. See Todd (1996) for discussion of local linear regression methods in nonrandom sampling situations.

laws, it is worthwhile to note that the cut-off size, which varies across states (from 1 to 100), is never equal to 15.³¹

5.2 Empirical Results

We estimate the conditional means nonparametrically by local linear regression, which requires that we choose a kernel and bandwidth. Previous research has shown that nonparametric estimates can be particularly sensitive to the choice of bandwidth—more so than to the choice of kernel. (see *e.g.* Silverman 1982, or Hardle and Linton, 1994). We specify the kernel to be the biweight kernel and consider several alternative choices of the bandwidth to check the robustness of our findings.³²

Figure 1 plots mean firm size against the average fraction minority for a representative year in the panel (1987) and shows the number of observations at each data point.³³ Estimates of the percentage minorities conditional on the firm size are obtained by local linear regression weighted means, taken separately over the data sets defined by $z_j < 15$ and $z_j \geq 15$. Figure 2 plots the estimates for each of the years in our panel. The estimated effect of the change in the law for firms of size 15 is the difference in the estimated conditional means evaluated at 15.³⁴ Table 2 reports the estimated treatment effect in each year, obtained under several different bandwidth choices.³⁵

Figure 2 and the estimates in Table 2 provide support for the hypothesis that the law has a positive effect on the percentage of minority workers employed at small firms.³⁶ In most of the years, the figure shows a positive change in the conditional mean at firm size equal to 15. However, the table reveals that often the estimated changes are not significantly different from zero, which we conjecture is due primarily to the moderate size samples available to us. For some of the years, the estimates are sensitive to the choice of bandwidth. In two of the years, 1987 and 1991, effects of a magnitude of around 10% (implying an increase of 10 percentage points in minorities' share

³¹It may be, however, that the federal law may have a smaller effect in states with stringent fair employment laws. The treatment effect we estimate is the impact of the federal law conditional on any existing state laws.

³²The biweight kernel (also called the quartic kernel) is $K(s) = \frac{15}{16}(s^2 - 1)^2$ for $|s| < 1$, = 0 else.

³³In our data, numbers such as 10, 15, 20, 25 and 50 tend to be reported more often, suggesting that individuals may round in reporting the firm size. It appears from the frequencies of reported firm sizes that individuals at larger firms are more likely to round. If employees are rounding in reporting firm size, then this creates a measurement error problem: Some employees giving a size below 15 may still be at a firm that reported to the EEOC and other individuals giving a size greater than 15 may be at firms that did not report. We do not yet know how to address the measurement error problem in the regression discontinuity design.

³⁴A bandwidth equal to 12 was used in generating the figure.

³⁵Estimation with bandwidths of size 7 or less gave estimates with high variability and resulted in negative estimated percentages minority for at least one year. Therefore, 8 is the smallest bandwidth size that we report in the table. We report results for bandwidths equal to 8, 10, 12, and 14 but bandwidths equal to 9, 11, and 13 yielded similar results.

³⁶Table B.2, which reports estimates omitting with exactly 15 employees from the sample firms, reveals that our point estimates are rather sensitive.

of employment) are statistically significant at conventional levels for all the bandwidth choices. At least for these two years, we can conclude that EEOC coverage had a positive effect of the percentage of minorities employed in small firms. The effect may be due to discrimination on the part of small firms in hiring minorities or to greater efforts on the part of minority workers to seek employment at firms that are covered by the law. We cannot distinguish between these two sources, but because of the RD design we can at least be relatively confident that this estimated effect is due to the law and not to other factors. On the other hand, the use of 5% significance levels may be a little misleading given the number of years used. Such significance levels may result in over-rejection when the ten tests are regarded as components of a single null hypothesis " H_0 : The law did not have any effect in any given year". It may well be the case that the significance of the two years, 1987 and 1991, is purely due to such problem. Bonferroni type modification then make it impossible to reject H_0 .

The second to last row of the table reports the estimated treatment effect that would be obtained simply by taking the difference between the percentage of minorities for firms with 15 or more employees and for firms with fewer than 15 employees. This would correspond to a simple IV estimator using an indicator for size ≥ 15 as the instrument. We find that such an estimator would lead to similar inference in terms of the signs of the estimates. On the other hand, the 'reported' standard errors based on the usual IV asymptotics are much smaller for this estimator. If the Wald assumption is satisfied, the IV estimator uses all the data instead of only the ones near cut-off point and, hence, is more accurate—as reflected in the smaller 'reported' standard errors. If the Wald assumption is not satisfied, then the IV estimator, viewed as an RD estimator, uses a bandwidth large enough to encompass all the data, and is therefore likely to suffer from severe bias. The last row of the table considers a local IV estimator, that only uses data in a neighborhood of the cut-off (observations within a bandwidth equal to 12). Here the effects are also positive but of somewhat smaller magnitude.³⁷

6 Summary

A RD evaluation provides an answer to a precise question: what is the treatment effect for a particular subgroup of individuals? The most attractive feature of the method is that it is justified with very minimal assumptions and bypasses many of the questions concerning model specification: both the question of which variables to include in the model and of their functional forms. A limitation of the RD method is that there are many aspects of the treatment impact distribution that may be of interest in any evaluation. Estimating these other parameters usually requires imposing more structure on the problem.

³⁷The estimated effects are very similar to those reported in Chay's (1995) study.

In this paper, we examine how the regression discontinuity (RD) data design can be used to nonparametrically estimate treatment effects. We consider the question of identification and estimation under two salient cases, the sharp design and the fuzzy design. The estimator we propose uses recently developed local linear nonparametric regression techniques that avoid the poor boundary behavior of the kernel regression estimator. We also demonstrate that the regression-discontinuity design sometimes provides a possible justification for the Wald estimator, even when the zero correlation condition is violated.

We apply the proposed methods to estimate the effect that EEOC coverage has on firms' minority employment patterns. Our estimator takes advantage of the fact that firms with fewer than 15 employees are not covered under the law. For two years in our panel, we find empirical support for a statistically significant positive effect of EEOC coverage on the employment of minority workers, although Bonferroni type consideration renders even such finding statistically insignificant.

References

- [1] ANGRIST, J. AND LAVY, V. (1996): "Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement," Unpublished manuscript, MIT.
- [2] ANGRIST, J., G. IMBENS AND D. RUBIN (1994): "Identification of Causal Effects using Instrumental Variables," forthcoming in *Journal of the American Statistical Association*.
- [3] BELSON, W. A. (1956): "A Technique for Studying the Effects of a Television Broadcast," *Applied Statistics*, Vol V, 195-202.
- [4] BERK, R. A. AND RAUMA, D. "Capitalizing on Nonrandom Assignment to Treatments: A Regression-Discontinuity Evaluation of a Crime-Control Program," *Journal of the American Statistical Association*, 78, 381, 21-28.
- [5] BLACK, S. (1996): "Do 'Better' Schools Matter? Parents Think So!" Unpublished manuscript, Harvard University.
- [6] BROWN, C. AND J. MEDOFF (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97, 1027-59.
- [7] CARRINGTON, W. J., K. MCCUE AND B. PIERCE, "Using Establishment Size to Measure the Impact of Title VII and Affirmative Action," Working Paper 347, Johns Hopkins University.
- [8] CHAY, K. Y. (1995), "The Impact of Federal Civil Rights Policy on Black Economic Progress: Evidence from the Equal Employment Opportunity Act of 1972", Working paper 346, Industrial Relations Section, Princeton University.
- [9] CLEVELAND, W. (1979): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829-836.
- [10] DONOHUE, J. AND HECKMAN, J. (1991): "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, 29, 1603-1643.
- [11] FAN, J. (1992): "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998-1004.
- [12] FAN, J. (1996): *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.
- [13] FAN, J., T. HU, AND Y. TRUONG (1994): "Robust Nonparametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433-446.

- [14] GOZALO, P. AND LINTON, O. (1997): "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression," unpublished manuscript, Yale University.
- [15] HÄRDLE, W. (1990), *Applied Nonparametric Regression*. New York: Cambridge University Press.
- [16] HÄRDLE, W. AND LINTON, O. (1994), "Applied Nonparametric Methods," in *Handbook of Econometrics, Volume 4*, ed. by D.F. McFadden and R.F. Engle. Amsterdam: North Holland, 2295 - 2339.
- [17] HECKMAN, J. (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313-318.
- [18] HECKMAN, J., H. ICHIMURA, J. SMITH AND P. TODD (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017 - 1098.
- [19] HECKMAN, J., AND J. SMITH (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, ed. by S. Strom. Cambridge: Cambridge University Press for Econometric Monograph Series.
- [20] HECKMAN, J. AND PAYNER, B. (1989): "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina," *American Economic Review*, 79, 138-177.
- [21] HOLZER, H. J. (1996): "Employer Hiring Decisions and Antidiscrimination Policy", in *Demand Side Strategies for Low-Wage Labor Markets*, ed. by R. Freeman and P. Gottschalk. New York: Russell Sage Foundation.
- [22] IMBENS, G., AND J. ANGRIST (1994): "Identification of Local Average Treatment Effects," in *Econometrica*, 62, 467-475.
- [23] PETERS, C. C. (1941): "A Method of Matching Groups for Experiments With no Loss of Population," *Journal of Educational Research*, 34, 606-612.
- [24] PORTER, J. (1998): "Estimation of Regression Discontinuities", *Seminar Notes*.
- [25] ROSENBAUM, P., AND D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.
- [26] ROSENBAUM, P., AND D. RUBIN (1985): "The Bias Due to Incomplete Matching," *Biometrics*, 41, 103 - 116.
- [27] SMITH, J. AND WELCH, F. (1989): "Black Economic Progress After Myrdal," *Journal of Economic Literature*, 27, 519-564.

- [28] STONE, C. (1977): "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 595-645.
- [29] THISTLETHWAITE, D., AND D. CAMPBELL (1960) : "Regression-discontinuity Analysis: An alternative to the ex post facto experiment", *Journal of Educational Psychology*, 51, 309-317.
- [30] TODD, P. (1995): "Local Linear Approaches to Program Evaluation Using a Semiparametric Propensity Score," unpublished manuscript, University of Pennsylvania.
- [31] TROCHIM, W. (1984): *Research Design for Program Evaluation: the Regression-Discontinuity Approach*. Beverly Hills: Sage Publications.
- [32] VAN DER KLAUW, W. (1996): "A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment," Unpublished manuscript, New York University.

Appendix A: Generalizing Condition (i) of (C3')

Instead of Condition (i) of (C3'), we assume the following continuity assumptions:

$$\lim_{e \rightarrow 0^+} \Pr [x_i(z_0 - e) = 1 | z_i = z_0 + e] = \lim_{e \rightarrow 0^+} \Pr [x_i(z_0 - e) = 1 | z_i = z_0 - e], \quad (4)$$

$$\lim_{e \rightarrow 0^+} \Pr [x_i(z_0 + e) - x_i(z_0 - e) = 1 | z_i = z_0 + e] = \Pr [x_i(z_0+) - x_i(z_0-) = 1 | z_i = z_0]. \quad (5)$$

$$\begin{aligned} & \lim_{e \rightarrow 0^+} E[\beta_i(z_0 + e) | z_i = z_0 + e, x_i(z_0 - e) = 1] \\ &= \lim_{e \rightarrow 0^+} E[\beta_i(z_0 - e) | z_i = z_0 - e, x_i(z_0 - e) = 1]. \end{aligned} \quad (6)$$

$$\begin{aligned} & \lim_{e \rightarrow 0^+} E[\beta_i(z_0 + e) | z_i = z_0 + e, x_i(z_0 + e) - x_i(z_0 - e) = 1] \\ &= E[\beta_i(z_0) | z_i = z_0, x_i(z_0+) - x_i(z_0-) = 1]. \end{aligned} \quad (7)$$

Remark: Consider the case where college scholarships are awarded on the basis of SAT scores (z_i). Assume that (around the discontinuity) (i) individuals in Group A (Always Takers) get the scholarship no matter what; (ii) individuals in Group C (Compliers) get the scholarship only if their SAT scores exceed the threshold; and (iii) individuals in Group N (Never Takers) never get the scholarship. Based on this interpretation, we may note the following:

- The event $x_i(z_0 - e) = 1$ is the same event that the i th individual belongs to Group A. Hence, $\Pr [x_i(z_0 - e) = 1 | z_i = z_0 + e]$ denotes the proportion of Group A among those whose SAT scores are equal to $z_0 + e$. Similarly, $\Pr [x_i(z_0 - e) = 1 | z_i = z_0 - e]$ is the proportion of Group A among those whose SAT score is $z_0 - e$. Therefore, (4) simply means that the proportion of Group A near the discontinuity is continuous as a function of SAT score.
- The event $x_i(z_0 + e) - x_i(z_0 - e) = 1$ is the same event that the i th individual belongs to Group C. Therefore, (5) simply means that the proportion of Group C among people with SAT score equal to z is a continuous function near the discontinuity.
- $E[\beta_i(z_0 \pm e) | z_i = z_0 \pm e, x_i(z_0 - e) = 1]$ denotes the average treatment effects among Group A whose SAT score is equal to $z_0 \pm e$. Therefore, (6) simply means that the average treatment effects for Group A is continuous near the discontinuity as a function of SAT score.
- $E[\beta_i(z_0 + e) | z_i = z_0 + e, x_i(z_0 + e) - x_i(z_0 - e) = 1]$ denotes the average treatment effects among Group C whose SAT score is equal to $z_0 + e$. Therefore, (7) simply means that the average treatment effects of Group C is a continuous function of SAT score near discontinuity.

After some algebra, it can be shown that

$$\begin{aligned}
E[x_i \beta_i | z_i = z_0 + e] &= E[\beta_i(z_0 + e) | z_i = z_0 + e, x_i(z_0 + e) - x(z_0 - e) = 1] \\
&\quad \cdot \Pr[x_i(z_0 + e) - x(z_0 - e) = 1 | z_i = z_0 + e] \\
&\quad + E[\beta_i(z_0 + e) | z_i = z_0 + e, x_i(z_0 + e) = x(z_0 - e) = 1] \\
&\quad \cdot \Pr[x(z_0 - e) = 1 | z_i = z_0 + e],
\end{aligned}$$

and

$$E[x_i \beta_i | z_i = z_0 - e] = E[\beta_i(z_0 - e) | z_i = z_0 - e, x_i(z_0 - e) = 1] \cdot \Pr[x_i(z_0 - e) = 1 | z_i = z_0 - e].$$

Therefore, we obtain

$$\begin{aligned}
&\lim_{e \rightarrow 0^+} \{E[y_i | z_i = z_0 + e] - E[y_i | z_i = z_0 - e]\} \\
&= E[\beta_i(z_0) | z_i = z_0, x_i(z_0+) - x_i(z_0-) = 1] \cdot \Pr[x_i(z_0+) - x(z_0-) = 1 | z_i = z_0]. \quad (8)
\end{aligned}$$

Now note that

$$\begin{aligned}
&E[x_i | z_i = z_0 + e] - E[x_i | z_i = z_0 - e] \\
&= E[x_i(z_0 + e) | z_i = z_0 + e] - E[x_i(z_0 - e) | z_i = z_0 + e] \\
&\quad + E[x_i(z_0 - e) | z_i = z_0 + e] - E[x_i(z_0 - e) | z_i = z_0 - e] \\
&= \Pr[x_i(z_0 + e) - x_i(z_0 - e) = 1 | z_i = z_0 + e] \\
&\quad + \Pr[x_i(z_0 - e) | z_i = z_0 + e] - \Pr[x_i(z_0 - e) | z_i = z_0 - e].
\end{aligned}$$

Therefore, we have

$$\lim_{e \rightarrow 0^+} \{E[x_i | z_i = z_0 + e] - E[x_i | z_i = z_0 - e]\} = \Pr[x_i(z_0 + e) - x_i(z_0 - e) = 1 | z_i = z_0 + e]. \quad (9)$$

Combining (8) and (9) along with (C1) and (ii) of (C3'), we obtain

$$\lim_{e \rightarrow 0^+} \frac{E[y_i | z_i = z_0 + e] - E[y_i | z_i = z_0 - e]}{E[x_i | z_i = z_0 + e] - E[x_i | z_i = z_0 - e]} = E[\beta_i(z_0) | z_i = z_0, x_i(z_0+) - x_i(z_0-) = 1].$$

Appendix B: Asymptotic Distribution Theory of the Estimator of Treatment Impacts based on Local Linear Regression

This appendix establishes a series of lemmas used to derive the distribution theory of the estimator of treatment effects based on local linear regression, discussed in sections 4.1 and 4.2 of the paper. We use the notation defined in the text and invoke assumptions 1-7 stated in section 4.2.1. For simplicity of notation, we use the same bandwidth h_n in describing the distribution theory for all the limit estimators, although the bandwidth need not be the same in implementation.

We first discuss the asymptotic behavior of \hat{y}^+ and \hat{x}^+ . The corresponding asymptotic theory of \tilde{y}^- and \tilde{x}^- are not discussed because they are analogous. The structure of our proof is an extension of Fan, Hu and Truong (1994).

The local linear regression estimators \hat{y}^+ and \hat{x}^+ are obtained by solving two minimization problems

$$\begin{aligned} \min_{a,b} \sum_{i=1}^n I_i \cdot (y_i - a - b(z_i - z_0))^2 K(h_n^{-1}(z_i - z_0)), \\ \min_{c,d} \sum_{i=1}^n I_i \cdot (x_i - c - d(z_i - z_0))^2 K(h_n^{-1}(z_i - z_0)), \end{aligned}$$

where $I_i = 1(z_i > z_0)$.

Define

$$y_i^* = y_i - a_0 - b_0 \cdot (z_i - z_0) \text{ and } x_i^* = x_i - c_0 - d_0 \cdot (z_i - z_0),$$

where $a_0 \equiv m^+(z_0)$, $b_0 \equiv m'^+(z_0)$, $c_0 \equiv p^+(z_0)$, $d_0 \equiv p'^+(z_0)$. In this notation, the objective functions can be written as

$$\begin{aligned} \sum_{i=1}^n (y_i^* - (a - a_0) - (b - b_0) \cdot (z_i - z_0))^2 \cdot I_i \cdot K(h_n^{-1}(z_i - z_0)), \\ \sum_{i=1}^n (x_i^* - (c - c_0) - (d - d_0) \cdot (z_i - z_0))^2 \cdot I_i \cdot K(h_n^{-1}(z_i - z_0)). \end{aligned}$$

Letting Z_i equal the vector $\left(1 \quad h_n^{-1}(z_i - z_0) \right)'$, first order conditions yield

$$\begin{pmatrix} \hat{a} - a_0 \\ h_n(\hat{b} - b_0) \end{pmatrix} = \left[\sum_{i=1}^n Z_i Z_i' \cdot K_i \right]^{-1} \sum_{i=1}^n Z_i y_i^* \cdot K_i$$

and

$$\begin{pmatrix} \hat{c} - c_0 \\ h_n(\hat{d} - d_0) \end{pmatrix} = \left[\sum_{i=1}^n Z_i Z_i' \cdot K_i \right]^{-1} \sum_{i=1}^n Z_i x_i^* \cdot K_i,$$

where $K_i \equiv I_i \cdot K(h_n^{-1}(z_i - z_0))$.

We next present several lemmas used to establish the asymptotic distribution of our estimator. Lemma 1 shows convergence in mean square of the denominator term in the above expressions ($\sum_{i=1}^n Z_i Z_i' \cdot K_i$). Lemmas 2 and 3 show convergence of the unconditional and conditional expectation of the numerator terms. Lemmas 4 and 5 show convergence of the unconditional and conditional variance of the numerator terms and Lemmas 6 and 7 show convergence in distribution of the numerator terms. The lemmas invoke assumptions 1-7 stated in the text.

Lemma 1 (Denominator)

$$(nh_n)^{-1} \sum_{i=1}^n Z_i Z_i' \cdot K_i \rightarrow f(z_0) \cdot \Gamma,$$

where

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_2 \end{bmatrix} \quad \text{and} \quad \gamma_l = \int_0^\infty u^l K(u) du.$$

P roof. Let

$$\theta_l \equiv (nh_n)^{-1} \sum_{i=1}^n (h_n^{-1}(z_i - z_0))^l \cdot I_i \cdot K(h_n^{-1}(z_i - z_0)), \quad l = 0, 1, 2.$$

Note that

$$\begin{aligned} E[\theta_l] &= h_n^{-1} E \left[(h_n^{-1}(z_i - z_0))^l \cdot I_i \cdot K(h_n^{-1}(z_i - z_0)) \right] \\ &= h_n^{-1} \int_{z_0}^\infty (h_n^{-1}(z - z_0))^l \cdot K(h_n^{-1}(z - z_0)) \cdot f(z) dz. \end{aligned}$$

With change of variables $u = h_n^{-1}(z - z_0)$, we obtain

$$E[\theta_l] = \int_0^\infty u^l K(u) f(z_0 + h_n u) du = f(z_0) \gamma_l + o(1),$$

where the second equality follows by the Dominated Convergence Theorem.

As for variance, we have

$$\begin{aligned} \text{Var}(\theta_l) &\leq (nh_n^2)^{-1} E \left[1(z_i > z_0) \cdot (h_n^{-1}(z_i - z_0))^{2l} \cdot K(h_n^{-1}(z_i - z_0))^2 \right] \\ &= (nh_n)^{-1} \int_0^\infty u^{2l} \cdot K(u)^2 \cdot f(z_0 + h_n \cdot z) du \\ &= o(1) \end{aligned}$$

again by the Dominated Convergence Theorem. ■

We now consider the asymptotic behaviors of

$$(nh_n)^{-1/2} \sum_{i=1}^n Z_i y_i^* \cdot K_i \quad \text{and} \quad (nh_n)^{-1/2} \sum_{i=1}^n Z_i x_i^* \cdot K_i.$$

For this purpose, define

$$\begin{aligned} \zeta(z) &\equiv m(z) - a_0 - b_0(z - z_0) - \frac{1}{2}m''^+(z_0)(z - z_0)^2 \\ \text{and } \xi(z) &\equiv p(z) - c_0 - d_0(z - z_0) - \frac{1}{2}p''^+(z_0)(z - z_0)^2, \end{aligned}$$

and observe that

$$\sup_{z_0 < z \leq z_0 + Mh_n} |\zeta(z)| = o(h_n^2), \quad \sup_{z_0 < z \leq z_0 + Mh_n} |\xi(z)| = o(h_n^2),$$

where M is defined in Assumption 1.

Lemma 2 (Numerator: Expectation)

$$\begin{aligned} E \left[(nh_n)^{-1} \sum_{i=1}^n Z_i y_i^* \cdot K_i \right] &\rightarrow \frac{1}{2} f(z_0) m''^+(z_0) h_n^2 \cdot (\delta + o(1)), \\ E \left[(nh_n)^{-1} \sum_{i=1}^n Z_i x_i^* \cdot K_i \right] &\rightarrow \frac{1}{2} f(z_0) p''^+(z_0) h_n^2 \cdot (\delta + o(1)), \end{aligned}$$

where

$$\delta = \begin{bmatrix} \delta_0 \\ \delta_1 \end{bmatrix} = \begin{bmatrix} \int_0^\infty u^2 K(u) du \\ \int_0^\infty u^3 K(u) du \end{bmatrix}.$$

P proof. We only prove the first claim because the second can be similarly established. Let

$$U_l = (nh_n)^{-1} \sum_{i=1}^n (h_n^{-1}(z_i - z_0))^l y_i^* K_i, \quad l = 0, 1.$$

Observe that

$$\begin{aligned} E[U_l] &= h_n^{-1} E \left[(h_n^{-1}(z_i - z_0))^l (m(z_i) - a_0 - b_0 \cdot (z_i - z_0)) \cdot K_i \right] \\ &= h_n^{-1} E \left[(h_n^{-1}(z_i - z_0))^l \left(\frac{1}{2} m''^+(z_0) \cdot (z_i - z_0)^2 + \zeta(z_i) \right) \cdot K_i \right]. \end{aligned}$$

Without loss of generality, we can assume that $[-M, M]$ contains the support of K . We then have

$$\begin{aligned} E[U_l] &= \frac{1}{2} m''^+(z_0) \cdot h_n^{-1} \int_{z_0}^\infty (h_n^{-1}(z_i - z_0))^l (z - z_0)^2 K(h_n^{-1}(z_i - z_0)) f(z) dz \\ &\quad + o(h_n^2) \cdot \frac{1}{h_n} \int_{z_0}^\infty (h_n^{-1}(z - z_0))^l K(h_n^{-1}(z - z_0)) f(z) dz \end{aligned}$$

We now note that

$$\begin{aligned}
& h_n^{-1} \int_{z_0}^{\infty} (h_n^{-1}(z - z_0))^l (z - z_0)^2 K(h_n^{-1}(z - z_0)) f(z) dz \\
&= h_n^2 \int_0^{\infty} u^l u^2 K(u) f(z_0 + h_n u) du \\
&= f(z_0) \delta_l h_n^2 + o(h_n^2),
\end{aligned}$$

by the Dominated Convergence Theorem. We also note that, because

$$\int_0^{\infty} u^l K(u) f(z_0 + h_n u) du$$

is finite, we have

$$\begin{aligned}
h_n^{-1} \int_{z_0}^{\infty} (h_n^{-1}(z - z_0))^l K(h_n^{-1}(z - z_0)) f(z) dz &= \int_0^{\infty} u^l K(u) f(z_0 - h_n u) du \\
&= O(1).
\end{aligned}$$

■

Lemma 3 (Numerator: Conditional Expectation)

$$\begin{aligned}
(nh_n)^{-1} \sum_{i=1}^n E[Z_i y_i^* \cdot K_i | z_i] &= E \left[(nh_n)^{-1} \sum_{i=1}^n Z_i y_i^* \cdot K_i \right] + o_p(h_n^2), \\
(nh_n)^{-1} \sum_{i=1}^n E[Z_i x_i^* \cdot K_i | z_i] &= E \left[(nh_n)^{-1} \sum_{i=1}^n Z_i x_i^* \cdot K_i \right] + o_p(h_n^2).
\end{aligned}$$

P roof. Again, we only prove the first claim. We have

$$\begin{aligned}
(nh_n)^{-1} \sum_{i=1}^n E[Z_i y_i^* \cdot K_i | z_i] &= (nh_n)^{-1} \sum_{i=1}^n Z_i K_i \cdot (m(z_i) - a_0 - b_0 \cdot (z_i - z_0)) \\
&= (nh_n)^{-1} \sum_{i=1}^n Z_i K_i \cdot \left(\frac{1}{2} m''^+(z_0) (z_i - z_0)^2 + \zeta(z_i) \right)
\end{aligned}$$

Observe that

$$\begin{aligned}
& \text{Var} \left[(nh_n)^{-1} \sum_{i=1}^n (h_n^{-1}(z_i - z_0))^l K_i \cdot \left(\frac{1}{2} m''^+(z_0) \cdot (z_i - z_0)^2 + \zeta(z_i) \right) \right] \\
&= (nh_n^2)^{-1} \text{Var} \left[(h_n^{-1}(z_i - z_0))^l K_i \cdot \left(\frac{1}{2} m''^+(z_0) \cdot (z_i - z_0)^2 + \zeta(z_i) \right) \right] \\
&\leq (nh_n^2)^{-1} C E \left[(h_n^{-1}(z_i - z_0))^{2l} K_i^2 \cdot \left((z_i - z_0)^4 + \zeta(z_i)^2 \right) \right] \\
&= C \cdot h_n^4 (nh_n^2)^{-1} h_n \int_0^{\infty} u^{2l+4} \cdot K(u)^2 \cdot f(z_0 + h_n \cdot u) du \\
&\quad + C \cdot o(h_n^2) (nh_n)^{-1} h_n \int_0^{\infty} u^{2l} \cdot K(u)^2 \cdot f(z_0 + h_n \cdot u) du
\end{aligned}$$

(The C after the inequality denotes a generic positive and finite constant.) By the Dominated Convergence Theorem, we have

$$\begin{aligned}
& \text{Var} \left[(nh_n)^{-1} \sum_{i=1}^n (h_n^{-1}(z_i - z_0))^l K_i \cdot \left(\frac{1}{2} m''_+(z_0) (z_i - z_0)^2 + \zeta(z_i) \right) \right] \\
&= O(n^{-1} h_n^3) + o(n^{-1} h_n^2) \\
&= h_n^2 \cdot [O(n^{-1} h_n) + o(n^{-1})] \\
&= h_n^2 \cdot o(1).
\end{aligned}$$

The conclusion follows. ■

Lemma 4 (Numerator: Conditional Variance)

$$\begin{aligned}
& \text{Var} \left[(nh_n)^{-1} \sum_{i=1}^n (Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i]) \right] \\
&= (nh_n)^{-1} \sigma^{2+}(z_0) f(z_0) \begin{bmatrix} v_0 + o(1) & v_1 + o(1) \\ v_1 + o(1) & v_2 + o(1) \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
& \text{Var} \left[(nh_n)^{-1} \sum_{i=1}^n (Z_i x_i^* \cdot K_i - E[Z_i x_i^* \cdot K_i | z_i]) \right] \\
&= (nh_n)^{-1} p^+(z_0) (1 - p^+(z_0)) f(z_0) \begin{bmatrix} v_0 + o(1) & v_1 + o(1) \\ v_1 + o(1) & v_2 + o(1) \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
& \text{Cov} \left[(nh_n)^{-1} \sum_{i=1}^n (Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i]), (nh_n)^{-1} \sum_{i=1}^n (Z_i x_i^* \cdot K_i - E[Z_i x_i^* \cdot K_i | z_i]) \right] \\
&= (nh_n)^{-1} \eta^+(z_0) f(z_0) \begin{bmatrix} v_0 + o(1) & v_1 + o(1) \\ v_1 + o(1) & v_2 + o(1) \end{bmatrix},
\end{aligned}$$

where

$$v_l = \int_0^\infty u^l K(u)^2 du.$$

P proof. Again, only the first claim is proved. Because

$$\begin{aligned}
& (nh_n)^{-1} \sum_{i=1}^n (Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i]) \\
&= (nh_n)^{-1} \sum_{i=1}^n Z_i K_i \cdot (y_i - a_0 - b_0 \cdot (z_i - z_0) - (m(z_i) - a_0 - b_0 \cdot (z_i - z_0))) \\
&= (nh_n)^{-1} \sum_{i=1}^n Z_i K_i \cdot (y_i - m(z_i)),
\end{aligned}$$

we only consider

$$(nh_n)^{-1} \sum_{i=1}^n K_i \cdot (h_n^{-1}(z_i - z_0))^l (y_i - m(z_i)), \quad l = 0, 1, 2.$$

Notice that its variance equals

$$\begin{aligned} & (nh_n^2)^{-1} E \left[K_i^2 \cdot (h_n^{-1}(z_i - z_0))^{2l} \cdot \sigma^2(z_i) \right] \\ &= (nh_n^2)^{-1} \int_{z_0}^{\infty} (h_n^{-1}(z - z_0))^{2l} \cdot K(h_n^{-1}(z - z_0))^2 \cdot \sigma^2(z) \cdot f(z) \, dz \\ &= (nh_n)^{-1} \int_0^{\infty} u^{2l} \cdot K(u)^2 \cdot \sigma^2(z_0 + h_n \cdot u) \cdot f(z_0 + h_n \cdot u) \, du \\ &= (nh_n)^{-1} \left(\sigma^{2+}(z_0) f(z_0) \cdot v_l + o(1) \right). \end{aligned}$$

The conclusion follows. ■

Lemma 5 (Numerator: Conditional CLT)

$$\begin{aligned} & (nh_n)^{-1/2} \sum_{i=1}^n \begin{pmatrix} Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i] \\ Z_i x_i^* \cdot K_i - E[Z_i x_i^* \cdot K_i | z_i] \end{pmatrix} \\ & \rightarrow f(z_0)^{-1} \mathcal{N} \left(0, \begin{bmatrix} \sigma^{2+}(z_0) S & \eta^+(z_0) S \\ \eta^+(z_0) S & p^+(z_0) (1 - p^+(z_0)) S \end{bmatrix} \right), \end{aligned}$$

where

$$S = \begin{bmatrix} v_0 & v_1 \\ v_1 & v_2 \end{bmatrix}.$$

P proof. We only consider

$$(nh_n)^{-1/2} \sum_{i=1}^n (Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i]).$$

By Cramér-Wold device, it suffices to establish that

$$\frac{\lambda'}{\sigma^{2+}(z_0) f(z_0) \cdot \lambda' S \lambda} (nh_n)^{-1/2} \sum_{i=1}^n \lambda' (Z_i y_i^* \cdot K_i - E[Z_i y_i^* \cdot K_i | z_i]) \rightarrow \mathcal{N}(0, 1)$$

for every $\lambda \in \mathbb{R}^2$. For simplicity, we consider CLT term by term, and ignore joint distribution. So we only consider

$$(nh_n)^{-1/2} \sum_{i=1}^n K_i \cdot (h_n^{-1}(z_i - z_0))^l (y_i - m(z_i)), \quad l = 0, 1.$$

We apply Lyapounov with third absolute moment: We need to establish

$$\left((nh_n)^{-1} \sigma^{2+}(z_0) f(z_0) v_{2l} \right)^{-3/2} (nh_n)^{-3} \sum_{i=1}^n E \left[K_i^3 \cdot (h_n^{-1}(z_i - z_0))^{3l} |y_i - m(z_i)|^3 \right] = o(1)$$

or

$$(nh_n)^{-1/2} h_n^{-1} E \left[K_i^3 \cdot (h_n^{-1}(z_i - z_0))^{3l} |y_i - m(z_i)|^3 \right] = o(1).$$

Let

$$\{ (z) \equiv E \left[|y_i - m(z_i)|^3 \mid z_i \right].$$

Because

$$\begin{aligned} & h_n^{-1} E \left[K_i^3 \cdot (h_n^{-1}(z_i - z_0))^{3l} |y_i - m(z_i)|^3 \right] \\ &= h_n^{-1} \int_{z_0}^{\infty} (h_n^{-1}(z - z_0))^{3l} \cdot K(h_n^{-1}(z - z_0))^3 \cdot \{ (z) \cdot f(z) \, dz \\ &= \int_0^{\infty} u^{3l} \cdot K(u)^3 \cdot \{ (z_0 + h_n \cdot u) \cdot f(z_0 + h_n \cdot u) \, du, \end{aligned}$$

we have

$$h_n^{-1} E \left[K_i^3 \cdot (h_n^{-1}(z_i - z_0))^{3l} |y_i - m(z_i)|^3 \right] = O(1)$$

and the conclusion follows. ■

Lemma 6 (Numerator: Unconditional CLT)

$$\begin{aligned} & (nh_n)^{-1/2} \sum_{i=1}^n \begin{pmatrix} Z_i y_i^* \cdot K_i \\ Z_i x_i^* \cdot K_i \end{pmatrix} - \frac{(nh_n)^{1/2} h_n^2}{2} f(z_0) \begin{pmatrix} m_+''(z_0) \cdot \delta \\ p_+''(z_0) \cdot \delta \end{pmatrix} \\ & \rightarrow f(z_0)^{1/2} \cdot \mathcal{N} \left(0, \begin{bmatrix} \sigma^{2+}(z_0) S & \eta^+(z_0) S \\ \eta^+(z_0) S & p^+(z_0) (1 - p^+(z_0)) S \end{bmatrix} \right). \end{aligned}$$

P proof. Due to Lemmas 2 and 3, we have

$$(nh_n)^{-1/2} \sum_{i=1}^n E \left[\begin{pmatrix} Z_i y_i^* \cdot K_i \\ Z_i x_i^* \cdot K_i \end{pmatrix} \right] = \frac{1}{2} (nh_n)^{1/2} h_n^2 f(z_0) \begin{pmatrix} m_+''(z_0) \cdot (\delta + o(1)) \\ p_+''(z_0) \cdot (\delta + o(1)) \end{pmatrix}$$

and

$$E \left[(nh_n)^{-1/2} \sum_{i=1}^n \begin{pmatrix} Z_i y_i^* \cdot K_i \\ Z_i x_i^* \cdot K_i \end{pmatrix} \middle| z_i \right] = (nh_n)^{-1/2} \sum_{i=1}^n E \left[\begin{pmatrix} Z_i y_i^* \cdot K_i \\ Z_i x_i^* \cdot K_i \end{pmatrix} \right] + (nh_n)^{1/2} o_p(h_n^2).$$

It follows that

$$E \left[(nh_n)^{-1/2} \sum_{i=1}^n \begin{pmatrix} Z_i y_i^* \cdot K_i \\ Z_i x_i^* \cdot K_i \end{pmatrix} \middle| z_i \right] - \frac{1}{2} (nh_n)^{1/2} h_n^2 f(z_0) \begin{pmatrix} m_+''(z_0) \cdot (\delta + o(1)) \\ p_+''(z_0) \cdot (\delta + o(1)) \end{pmatrix} = o_p(nh_n^5).$$

The conclusion follows. ■

The following lemma gives the joint distribution for $\widehat{a} - a_0$, $h_n(\widehat{b} - b_0)$, $\widehat{c} - c_0$ and $h_n(\widehat{d} - d_0)$:

Lemma 7

$$\begin{aligned}
& (nh_n)^{1/2} \left(\widehat{a} - a_0, h_n (\widehat{b} - b_0), \widehat{c} - c_0, h_n (\widehat{d} - d_0) \right)' \\
&= \frac{1}{2} (nh_n)^{1/2} h_n^2 \begin{bmatrix} \Gamma^{-1} & 0 \\ 0 & \Gamma^{-1} \end{bmatrix} \cdot \begin{pmatrix} m''_+(z_0) \cdot \delta \\ p''_+(z_0) \cdot \delta \end{pmatrix} \\
&+ f(z_0)^{-1/2} \begin{bmatrix} \Gamma^{-1} & 0 \\ 0 & \Gamma^{-1} \end{bmatrix} \cdot \mathcal{N} \left(0, \begin{bmatrix} \sigma_+^2(z_0) S & \eta_+(z_0) S \\ \eta_+(z_0) S & p_+(z_0) (1 - p_+(z_0)) S \end{bmatrix} \right) + o_p(1).
\end{aligned}$$

P proof. Follows easily from Lemmas 1 and 6. ■

Proof of Theorem 1 and 1'

As a consequence of lemmas 1-8, we obtain

$$n^{\frac{2}{5}} \begin{pmatrix} \widehat{y}^+ - y^+ \\ \widehat{x}^+ - x^+ \end{pmatrix} \rightarrow \mathcal{N} \left(\rho^+ \begin{bmatrix} m''^+(z_0) \\ p''^+(z_0) \end{bmatrix}, \omega^+ \begin{bmatrix} \sigma^{2+}(z_0) & \eta^+(z_0) \\ \eta^+(z_0) & p^+(z_0) (1 - p^+(z_0)) \end{bmatrix} \right).$$

Analogous expressions can be obtained for $(\widehat{y}^- - y^-, \widehat{x}^- - x^-)'$. Since observations are independent, $n^{\frac{2}{5}} (\widehat{y}_+ - y_+, \widehat{x}_+ - x_+)'$ and $n^{\frac{2}{5}} (\widehat{y}_- - y_-, \widehat{x}_- - x_-)'$ are independent of each other. At this point, the result stated in theorem 1 of the text follows straightforwardly by the Delta Method so the rest of the proof is omitted. ■

Data Appendix

>From the original 1979-1993 NLSY waves containing information on 12682 individuals, a sample was extracted that included all person-year observations for the years in which the question on firm size was asked (1979, 1980 and 1986 to 1993). Only individuals who reported to be employed at the time of each interview at a firm with only 1 location and who reported a positive number of employees at this location were included, leaving 25059 person-year observations on 8151 different individuals.

Subsequently all individuals with missing information on race and those who reported to be self-employed or to be working without pay were excluded to result in a final sample of 21383 person-year observations on 8033 different individuals.

The following table gives information similar to that shown in Figure 1 for all the years of the survey considered here combined.

Table A1

Firm size	1	2	3	4	5	6	7	8	9	10
No. of obs.	1124	1038	1113	1087	1066	937	681	638	405	801
Mean y	0.139	0.145	0.152	0.147	0.155	0.140	0.125	0.121	0.131	0.151
Firm size	11	12	13	14	15	16	17	18	19	20
No. of obs.	197	503	219	187	644	144	107	148	33	714
Mean y	0.104	0.141	0.145	0.106	0.160	0.168	0.177	0.136	0.089	0.175
Firm size	21	22	23	24	25	26	27	28	29	30
No. of obs.	53	88	39	63	587	43	41	55	12	744
Mean y	0.149	0.147	0.132	0.184	0.168	0.175	0.085	0.074	0.159	0.172

Table 1
Descriptive Statistics^(a)
Means with Standard Errors reported in parentheses
(NLSY weights used to account for nonrandom sampling)

variable	Year									
	1979	1980	1986	1987	1988	1989	1990	1991	1992	1993
black	0.08 (0.006)	0.09 (0.006)	0.12 (0.006)	0.12 (0.006)	0.12 (0.007)	0.12 (0.007)	0.11 (0.006)	0.12 (0.007)	0.12 (0.008)	0.10 (0.007)
white	0.86 (0.008)	0.84 (0.008)	0.81 (0.008)	0.80 (0.008)	0.80 (0.008)	0.81 (0.008)	0.82 (0.008)	0.82 (0.009)	0.81 (0.009)	0.83 (0.009)
Hispanic	0.05 (0.005)	0.06 (0.005)	0.06 (0.005)	0.07 (0.005)	0.06 (0.005)	0.06 (0.005)	0.06 (0.005)	0.05 (0.005)	0.05 (0.005)	0.06 (0.006)
female	0.42 (0.01)	0.46 (0.01)	0.47 (0.01)	0.44 (0.01)	0.43 (0.01)	0.43 (0.01)	0.44 (0.01)	0.45 (0.01)	0.44 (0.011)	0.43 (0.012)
firm size^(b)	131 (17)	165 (29)	178 (20)	250 (27)	199 (24)	442 (109)	4304 (412)	3592 (416)	2738 (368)	687 (178)

(a) Sample includes all men and women in the NLSY who do not report working at firms with multiple locations, who do not have missing data on race, sex or firm size, and who are not self-employed or working without pay. See Table B.1 for number of observations omitted due to missing data.

(b) The firm size mean is a simple mean over the observations in the sample and is unadjusted for size-biased sampling.

Table 2
Estimated Effects of EEOC-reporting on Percentage Minority in the Firm^(a)
(asymptotic standard errors reported in parentheses)

Bandwidth	Year									
	1979	1980	1986	1987	1988	1989	1990	1991	1992	1993
8	-8.7 (9.1)	-26.8 (11.3)	9.8 (6.7)	10.9 (5.3)	12.2 (8.9)	9.8 (11.1)	4.7 (5.9)	11.1 (4.1)	3.7 (10.4)	7.6 (9.1)
10	-2.1 (5.9)	-7.5 (9.0)	5.1 (6.9)	10.1 (4.6)	3.7 (7.5)	1.5 (6.0)	4.3 (7.3)	9.4 (4.6)	2.7 (10.0)	6.1 (8.8)
12	3.3 (4.5)	1.9 (13.2)	3.3 (8.7)	10.5 (4.7)	-1.3 (8.3)	5.5 (6.7)	7.4 (9.0)	9.6 (4.8)	2.6 (12.0)	4.7 (12.0)
14	8.2 (5.4)	8.8 (18.6)	1.9 (11.2)	10.6 (5.1)	-2.7 (8.0)	8.0 (8.4)	10.3 (8.8)	11.5 (4.7)	5.3 (13.8)	-1.5 (14.0)
global IV	3.3 (0.3)	7.1 (0.4)	8.4 (0.3)	6.7 (0.3)	8.3 (0.3)	8.2 (0.3)	7.2 (0.3)	7.9 (0.4)	11.0 (0.5)	7.4 (0.4)
local IV with bw=12	0.92 (0.3)	3.03 (0.3)	2.8 (0.2)	4.0 (0.1)	3.7 (0.3)	3.8 (0.2)	2.6 (0.2)	3.0 (0.3)	3.5 (0.3)	-1.0 (0.3)

(a) The first four rows report results from using the estimator based on local linear regression described in section 4.2 of the text. The last two rows report results using a global Wald estimator (i.e. one that uses all the data) and a local Wald estimator (one that only uses data within a bandwidth).

Table B.1
Number of Observations Omitted^(a)

	Year									
	1979	1980	1986	1987	1988	1989	1990	1991	1992	1993
number of observations	1893	1935	2874	2962	2784	2788	2758	2395	2357	2313
missing race	11	15	12	19	13	10	13	15	18	16
missing class of worker or self-employed or working without pay	0	0	339	372	411	507	441	466	499	528
number after omitting persons with missing data	1182	1935	2528	2575	2362	2275	2305	1918	1845	2297

(a) The number of observations in the second row are all men and women in the NLSY who do not reporting at firms with multiple location. None of the observations were missing data on sex.

Figure 1: Avg. Fraction Minority by Firm Size in 1987 (Number of Observations shown at each point)

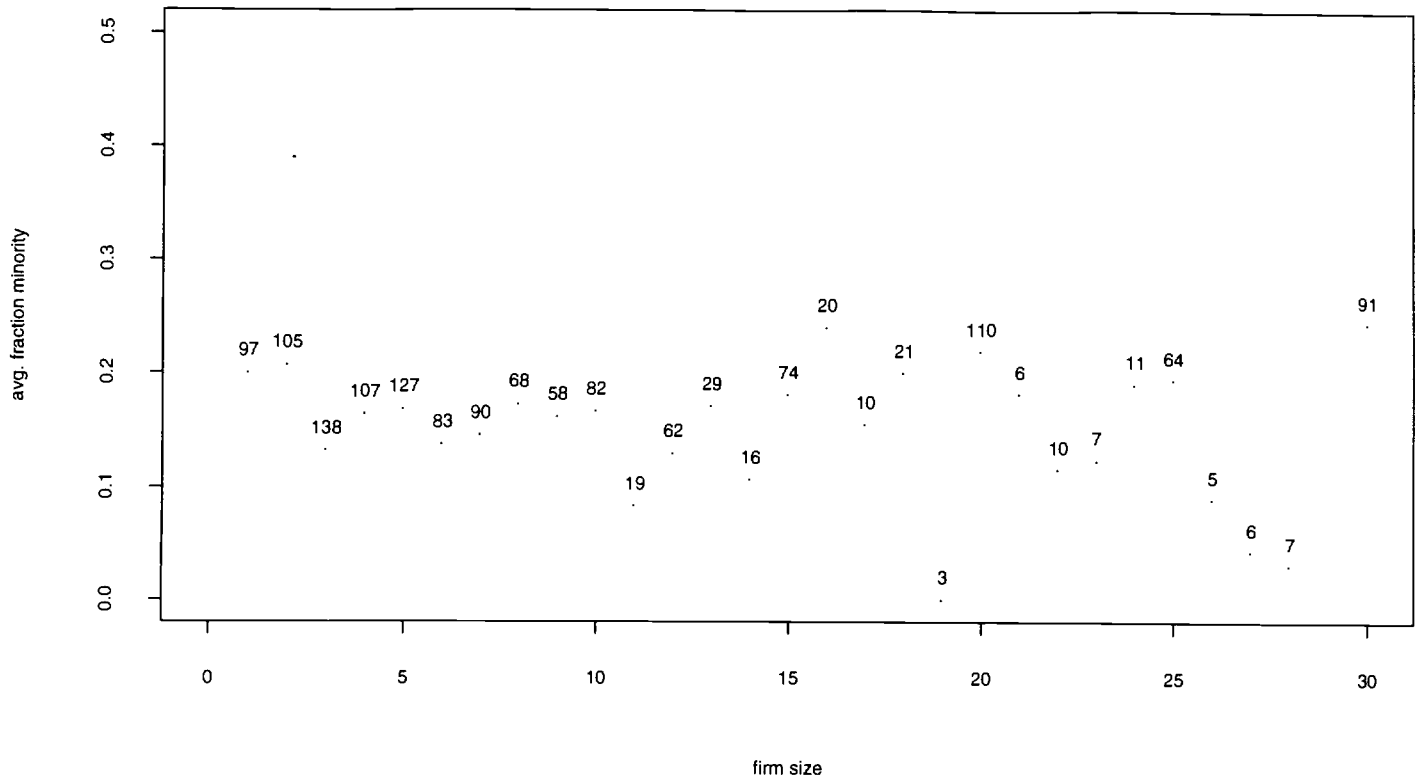


Table B.2
Estimated Effects of EEOC-reporting on Percentage Minority in the Firm
omitting from the Sample Firms with Exactly 15 employees^(a)
 (asymptotic standard errors reported in parentheses)

Bandwidth	Year									
	1979	1980	1986	1987	1988	1989	1990	1991	1992	1993
8	45.0 (35.8)	-30.7 (12.9)	31.2 (18.9)	16.2 (10.8)	5.1 (14.5)	8.8 (24.9)	-3.2 (7.4)	-7.1 (6.8)	5.8 (14.9)	-3.6 (12.6)
10	15.0 (14.0)	-7.7 (9.2)	14.8 (10.4)	14.0 (5.2)	-6.6 (9.0)	1.1 (7.8)	0.7 (7.7)	0.3 (0.5)	5.0 (11.7)	-3.5 (9.3)
12	10.9 (10.7)	3.4 (13.3)	10.9 (10.7)	13.4 (5.0)	-12.3 (8.9)	5.3 (7.5)	6.2 (9.4)	3.2 (5.5)	2.3 (12.9)	-2.8 (12.1)
14	13.2 (9.9)	11.3 (18.9)	8.3 (12.4)	13.1 (5.2)	-12.9 (8.2)	7.9 (8.9)	10.3 (9.3)	7.1 (5.5)	4.4 (14.4)	0.2 (14.1)
global IV	3.5 (0.4)	7.5 (0.4)	9.0 (0.3)	6.9 (2.9)	8.2 (0.4)	8.5 (0.4)	7.3 (0.4)	0.8 (0.4)	11.2 (0.5)	8.1 (4.4)
local IV with bw= 12	1.3 (0.4)	4.0 (0.4)	4.2 (0.3)	4.3 (0.1)	1.3 (0.3)	4.0 (0.2)	2.4 (0.2)	2.5 (0.3)	4.4 (0.3)	-0.0 (0.4)

(a) The first four rows report results from using the estimator based on local linear regression described in section 4.2 of the text. The last two rows report results using a global Wald estimator (i.e. one that uses all the data) and a local Wald estimator (one that only uses data within a bandwidth).

Figure 2: Estimated Percentage Minority Conditional on Firm Size (bw=12)

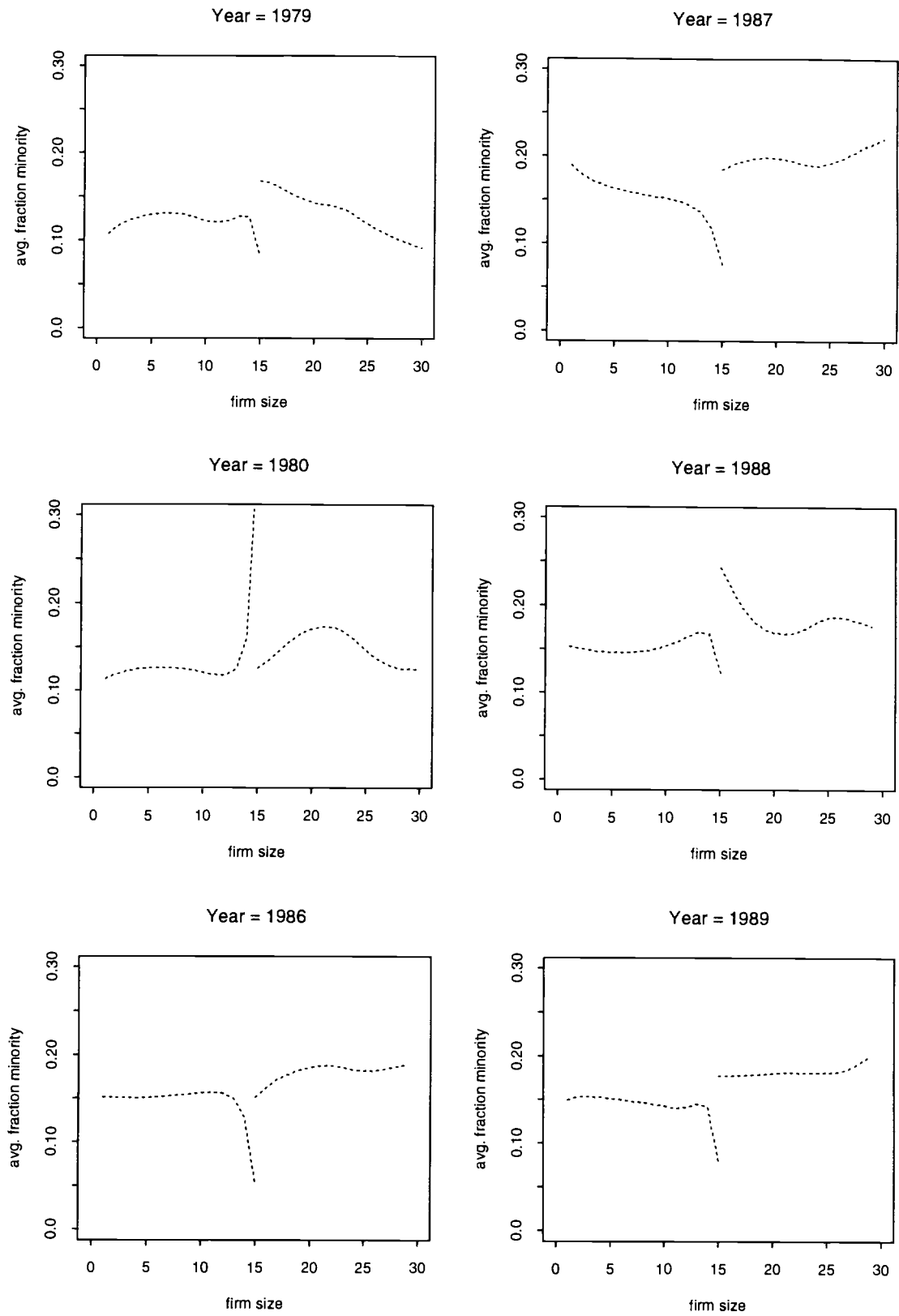


Figure 2: Estimated Percentage Minority Conditional on Firm Size (bw=12)

