HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 11/17

Extending Regression Discontinuity Models Beyond the Jump Point

Chunrong Ai
Edward C. Norton
Zhou Yang

July 2011

york.ac.uk/res/herc/hedgwp

Extending Regression Discontinuity Models Beyond the Jump Point[1]

Chunrong Ai
Department of Economics, University of Florida
Gainesville, FL 32611, USA

Edward C. Norton
Department of Health Management and Policy and
Department of Economics, University of Michigan
Ann Arbor, MI 48109, USA

Zhou Yang
Department of Health Policy and Management, Emory University
Atlanta, GA 30322, USA

This verion is still under revision; please do not quote.

[1]We have benefited greatly from conversation with David Lee and the comments made by seminar participants at Maryland, Ohio State, NYU-Columbia joint econometric seminar. The Warrington College of Business at the University of Florida provided partial support to Chunrong Ai through 2006 and 2007 Summer research grant.

Abstract

This paper proposes a new estimation method for regression discontinuity models, allowing for estimation of a treatment effect beyond the jump point (with additional assumptions). The proposed procedure consistently estimates the treatment effect function, as well as the average outcome in the absence of treatment. The treatment effect estimator is root-N consistent. We apply the method to an important question in health economics—what is the effect of having Medicare insurance on admissions and health care costs after age 65? Our preferred models shows an increase in both admissions and costs after age 65 due to Medicare.

KEYWORDS: Regression discountiuity; Medicare; mortality; health expenditures.

1 Introduction

Regression discontinuity (RD) has recently been suggested as a framework for evaluating treatment programs in a quasi-experimental design. The key assumption is that the probability of treatment is a discontinuous function of a continuous variable. Small changes in the value of this continuous variable therefore produce a large jump in the probability of treatment, thus providing identification of the treatment effect. In many cases this provides a powerful statistical tool to identify the treatment effect of interest. However, standard RD says nothing about the treatment effect for any other value of the continuous variable away from the discontinuity. For many treatments, however, the effect of the treatment may continue to change over time. The policy interest is in estimating the treatment effect over an extended range of the identifying continuous variable.

Our new approach extends regression discontinuity to allow for estimated effects beyond the jump point. Our approach is easy to estimate because it only requires running a series of least squares regressions. It is flexible, being semi-parametric, and therefore consistently estimates the parameters for a variety of underlying functions. We also explain the limitations and underlying assumptions of this approach. Therefore, an important part of his paper is not only showing how to extend regression discontinuity beyond the jump point, but also in understanding the limitations inherent in the assumptions necessary to estimate the model.

We apply the methods to data on near-elderly who then become eligible for Medicare. We follow them for several years after age 65 so that we can estimate the effects of insurance on mortality and health care expenditures beyond the jump point. Our preferred model shows a modest improvement in mortality at age 65, but, surprisingly, a worse effect over time such that the net effect after 5 years (age 70) is about zero. Our results are not sensitive to the identifying assumptions.

We start with the basic regression discontinuity model and then build upon it by allowing the treatment effect to vary over time beyond the jump point. We next lay out the assumptions needed to identify such an effect. The following section explains how to estimate the parameters of the model for either a sharp or fuzzy design. We also discuss mispecification bias. The last part of the paper implements the model to estimate the effect of having Medicare health insurance on admissions and on health care expenditures after age 65.

2 Model

We formulate the model by imposing restrictions on the following dummy variable equation (see Ai (2008)):

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

where i indexes individual, y_i denotes the observed outcome, x_i denotes the observed treatment status: $x_i = 1$ if treatment is received and $x_i = 0$ if treatment is not received, α_i denotes the outcome if treatment is not received, and $\alpha_i + \beta_i$ denotes the outcome if treatment is received (so that β_i is the treatment effect). The variables (α_i, β_i, x_i) are often related to some underlying variables z_i (which could be more than one variables) through the probability of receiving treatment $p(z) = E\{x_i | z_i = z\}$, the average baseline effect $\alpha(z) = E\{\alpha_i | z_i = z\}$, and the average treatment effect $\beta(z) = E\{\beta_i | z_i = z\}$. The key restriction is that the probability of receiving treatment, $p(z)$, is a discontinuous function. Specifically, for some exclusive decomposition $\mathcal{Z} = \mathcal{Z}_- \cup \mathcal{Z}_+$, there exists some z_0 on the boundary of both subsets such that

$$\begin{aligned}
\lim_{z \to z_0^+} p(z) &= \lim_{t \to 0} p(z^+(t)) \text{ for some path } z^+(t) \in \mathcal{Z}_+ \text{ for } t > 0 \text{ and } z^+(0) = z_0, \\
\lim_{z \to z_0^-} p(z) &= \lim_{t \to 0} p(z_(t)) \text{ for some path } z_(t) \in \mathcal{Z}_- \text{ for } t > 0 \text{ and } z_(0) = z_0.
\end{aligned}$$

Discontinuity means that

$$\lim_{z \to z_0^+} p(z) \neq \lim_{z \to z_0^-} p(z)$$

holds. In economic applications, the subset $\mathcal{Z}_+$ defines eligibility criterion for the treatment program with z_0 as the minimum criterion and $\mathcal{Z}_-$ consists of individuals who do not meet the minimum criterion. Since ineligible individuals are not allowed to participate in the program, we have $p(z) = 0$ for all $z \in \mathcal{Z}_-$. Thus, the discontinuity restriction is satisfied if there is a nontrivial probability of receiving treatment for any eligible individual: $p(z) > 0$ for any $z \in \mathcal{Z}_+$. It is worth noting that the eligibility criterion $\mathcal{Z}_+$, though determined by a single underlying variable in most economic applications, may depend on several underlying variables. For instance, a perspective student's eligibility for a need-based scholarship requires her SAT score to be above certain threshold and her family income is below certain level. If eligibility is determined by more than one variable, then z_0 is not unqiue. There may exist other points at which $p(z)$ is discontinuous. The proposed approach in this paper applies to any discontinuity point so we should treat z_0 as a generic discontinuity point.

Under the conditional mean independence condition:

$$E\{\beta_i x_i | z_i = z\} = E\{\beta_i | z_i = z\} E\{x_i | z_i = z\} = \beta(z) * p(z)$$

and the condition that $\alpha(z)$ and $\beta(z)$ are both continuous at $z = z_0$, Hahn, Todd, and Van der

Klaauw (2001) show that the coefficient $\beta(z_0)$ is identified as

$$\beta(z_0) = \frac{\lim_{z \to z_0^+} E\{y_i | z_i = z\} - \lim_{z \to z_0^-} E\{y_i | z_i = z\}}{\lim_{z \to z_0^+} p(z) - \lim_{z \to z_0^-} p(z)}.$$

They then proceed to propose an estimator for $\beta(z_0)$ by replacing the pathwise limits with consistent nonparametric estimates. Porter (2003) derives the asympptotic distribution of this and other similar estimators. Using the same proof as in Hahn, Todd, and Van der Klaauw (2001), if there exist other discontinuity points, we can show that the values of $\beta(z)$ at those discontinuity points are also identified.

The attraction of Hahn's et. al. procedure is that it is quite simple to compute and has desirable asymptotic properties. The limitation of their procedure is that it says nothing about the treatment effect on non-threshold eligible individuals. It is possible that the threshold individuals receive no effect: $\beta(z_0) = 0$ but other eligible individuals receive positive effect: $\beta(z) > 0$ for $z \in \mathcal{Z}_+$. It is also possible that the threshold individuals receive negative effect: $\beta(z_0) < 0$ but well qualified individuals receive positive effect: $\beta(z) > 0$ for large $z \in \mathcal{Z}_+$, or that $\beta(z_0) > 0$ and $\beta(z) < 0$ for large $z \in \mathcal{Z}_+$. Thus, to evaluate the effectiveness of the treatment program, it is important to estimate the whole function $\beta(z)$, not just $\beta(z_0)$. Unfortunately, under Hahn's et. al. conditions, $\beta(z_0)$ is the only parameter that is identified. The identification difficulty arises from the fact that there are no individuals with the same underlying variables z in both the ineligible and eligible group and that the eligible individuals alone cannot distinguish the treatment effect $\beta(z)$ from the baseline effect $\alpha(z)$.

Clearly, to identify the treatment effect function, some restrictions must be imposed on either the baseline effect $\alpha(z)$ or the treatment effect $\beta(z)$ or both. Since the treatment effect function is the focus of most empirical studies, it is natural to impose restrictions on the treatment effect function but leave the baseline effect function $\alpha(z)$ unspecified. It is worth noting that any parameterization of $\beta(z)$ such as $h(z, \theta_0)$ is not necessarily identified without some restrictions on $\alpha(z)$. This follows because we can always write

$$h(z, \theta_0) + \alpha(z) = h(z, \theta) + (\alpha(z) + h(z, \theta_0) - h(z, \theta)) = h(z, \theta) + \widetilde{\alpha}(z), \ z \in \mathcal{Z}_+$$

for any θ. With restriction that $\alpha(z)$ is continuous at $z = z_0$, we can identify $h(z_0, \theta_0)$ but not θ_0 if θ_0 is not univariate.

Some other form of restriction must be imposed to identify $\beta(z)$. We now show that the following

functional form restriction on $\beta(z)$:

$$\frac{\partial^m \beta(z)}{\partial v^m} = 0 \text{ holds over } z \in \mathcal{Z}_+ \text{ almost everywhere,} \tag{2}$$

where m is some known integer (e.g., $m = 2$) and $z = (v, w')'$ with v a continuous scalar, is enough to identify $\beta(z)$ and $\alpha(z)$. To see this, suppose that $\beta(z)$ and $\alpha(z)$ have up to m^{th} derivatives with respect to v almost everywhere. Consider the sharp design $x = 1\{z \in \mathcal{Z}_+\}$. Condition (2) allows us to eliminate the treatment effect $\beta(z)$ by differentiation:

$$\frac{\partial^m E\{y_i | z_i = z\}}{\partial v^m} = \frac{\partial^m \alpha(z)}{\partial v^m} \text{ over } z \in \mathcal{Z} \text{ almost everywhere.}$$

Since we can compute $E\{y_i|z_i = z\}$ for almost all z from observed data, we can compute the derivative $h(z) = \frac{\partial^m \alpha(z)}{\partial v^m}$ almost everywhere. Notice that condition (2) implies

$$\beta(z) = b(w)' t^J(v),$$

for some $J \times 1$ vector of known functions $t^J(v)$ satisfying

$$\frac{d^m t^J(v)}{dv^m} = 0 \text{ almost everywhere,}$$

and $b(w)$ is a vector of unknown functions of the other underlying variables w. Examples of the known functions $t^J(v)$ include power functions:

$$t^J(v) = p^J(v) = (1, v, ..., v^{J-1})'$$

with v_0 as the known threshold, the step functions (e.g., $m = 1$):

$$t^J(v) = (1, 1\{e_1 \leq v < e_2\}, ..., 1\{e_J \leq v\})'$$

for some known $e_1, e_2, ..., e_J$, and the piecewise linear functions (e.g., $m = 2$):

$$t^J(v) = (1, 1\{e_1 \leq v < e_2\}, ..., 1\{e_J \leq v\})' \otimes (1, v)'.$$

In most applications, it is preferrable to specify $t^J(v)$ as low order spline basis functions. The component $b(w)$ is either parameterized (such as linear form) or left unspecified. We will discuss both cases. But for demonstration purpose, we will use the simplest case: $z = v$, $t^J(v) = 1$ and

$b(w) = \theta_0$ as example.

Compute the indefinite integration of $h(z)$ with respect to v for m times, we obtain

$$\begin{aligned}\alpha(z) &= \int \left(\int \cdots \int h(v,w)dv \cdots \right) dv + c(w)'p^J(v) \\ &= g(z) + c(w)'p^J(v)\end{aligned}$$

for almost all z, where $c(w)$ is a vector of unknown functions and $g(z)$ is the known part of the indefinite integration. This means that $\alpha(z)$ is identified if $c(w)$ is identified. Let $\mathcal{W}$ denote the support of w and let $\mathcal{V}$ denote the support of v. Suppose that $\mathcal{Z}_+ = \{v \in \mathcal{V}, v \geq v_0\} \times \mathcal{W}$ and $\mathcal{Z}_- = \{v \in \mathcal{V}, v < v_0\} \times \mathcal{W}$. Then, $c(w)$ is identified by ineligible individuals as $c(w)$ solves

$$\min_{c(\cdot)} E\left\{(y_i - g(z_i) - c(w_i)'p^J(v_i))^2 | w_i = w, x_i = 0\right\}.$$

After computing $\alpha(z) = g(z) + c(w)'p^J(v)$, $b(w)$ is identified by eligible individuals:

$$\min_{b(\cdot)} E\left\{y_i - \alpha(z_i) - b(w_i)'t^J(v_i))^2 | w_i = w, x_i = 1\right\},$$

provided that

$$E\{t^J(v_i)t^J(v_i)' | w_i = w, x_i = 1\} \text{ is nonsingular for almost all } w.$$

We summarize these results in the following assumption and lemma.

Assumption 1. *(i) The observations* $\{(y_i, x_i, z_i), i = 1, 2, ..., N\}$ *are identically distributed; (ii) the following conditional independence condition*

$$E\{\beta_i x_i | z_i = z\} = E\{\beta_i | z_i = z\}E\{x_i | z_i = z\} = \beta(z) * p(z)$$

holds for all z*; (iii) for some known basis functions* $t^J(v)$, $\beta(z) = b(w)'t^J(v)$*; (iv) for some known integer* m, $t^J(v)$ *and* $\alpha(z)$ *have up to* m^{th} *derivatives with respect to* v *almost everywhere and*

$$\frac{d^m t^J(v)}{dv^m} = 0 \text{ holds almost everywhere;}$$

(v) $\mathcal{Z}_+ = \{v \in \mathcal{V}, v \geq v_0\} \times \mathcal{W}$ *and* $\mathcal{Z}_- = \{v \in \mathcal{V}, v < v_0\} \times \mathcal{W}$*; (vi)*

$$E\{t^J(v_i)t^J(v_i)' | w_i = w, x_i = 1\}$$

is nonsingular for every w; (vii) $x = 1\{v \geq v_0\}$ and $p(z) = x$.

Lemma 1. *For the sharp design, Assumption 1 identifies $\alpha(z)$ for $z \in \mathcal{Z}$ and $\beta(z) = b(w)'t^J(v)$ for $z \in \mathcal{Z}_+$.*

Before we turn to the case of fuzzy design, we illustrate the identification through the simplest example: $z = v$, $t^J(v) = 1$ and $b(w) = \theta_0$. In this example, we have

$$\begin{aligned} h(v) &= \frac{dE\{y|v, x=0\}}{dv} \text{ for } v < v_0 \text{ and } h(v) = \frac{dE\{y|v, x=1\}}{dv} \text{ for } v \geq v_0, \\ \alpha(v) &= \int h(v)dv + c = g(v) + c. \end{aligned}$$

The unknown constant c solves

$$\min_c E\left\{(y_i - g(v_i) - c)^2 | x_i = 0\right\},$$

which yields

$$c = E\left\{y_i - g(v_i) | x_i = 0\right\}.$$

The treatment effect solves

$$\min_\theta E\left\{y_i - g(v_i) - c - \theta)^2 | x_i = 1\right\},$$

which yields

$$\theta_0 = E\left\{y_i - g(v_i) - c) | x_i = 1\right\}.$$

It is clear from the above expression that we use the whole treatment group, not just the threshold individuals, to estimate the average treatment effect. Thus, the proposed procedure yields a root-N consistent estimator for the average treatment effect. The proposed procedure is criticized, however, for being multistep (e.g., step I eliminates the treatment effect through differentiation; step II recovers the baseline effect from the ineligible group; step III estimates the treatment effect from eligible group) and not better than Porter's (2003) dummy variable approach. Again, in our simplest example, Porter's dummy variable approach is the model

$$y_i = \alpha(v_i) + x_i\theta_0 + u_i, i = 1, 2, ..., N.$$

Applying Robinson's partialling out procedure, we obtain

$$y_i - E\{y_i|v_i\} = (x_i - E\{x_i|v_i\})\,\theta_0 + u_i.$$

The problem here is that $x_i - E\{x_i|v_i\}$ is identically zero for all v_i. Porter suggests using the approximation $x_i - \widetilde{E}\{x_i|v_i\}$ and estimate the treatment effect by

$$\widehat{\theta}_{Porter} = \frac{\sum_{i=1}^{N}\left(x_i - \widetilde{E}\{x_i|v_i\}\right)y_i}{\sum_{i=1}^{N}\left(x_i - \widetilde{E}\{x_i|v_i\}\right)^2}.$$

This estimator appears using all observations, not just those near threshold. However, looking deeper, we find this estimator is no better than Hahn's estimator which use observations near threshold. The reason is that the approximation error $x_i - \widetilde{E}\{x_i|v_i\}$, though not identically zero, converges to zero at differential rate. The approximation error converges to zero for all v outside a neighborhood of v_0 faster than for v in the neighborhood of v_0. Since the slower rate terms dominate the faster rate terms in both the numerator and denominator of Porter's estimator, his estimator essentially uses individuals near threshold. Moreover, because of the approximation error converges to zero at rate slower than root-N, Porter's estimator is not root-N consistent. We are able to achieve root-N consistentcy because we exploit information on derivatives, not just the discontinuity, wheras Porter (2003) does not.

We now turn to the fuzzy design. Under our assumption, we have $E\{x|z\} = E\{x|v\}$. So the probability of receiving treatment is $p(v) = E\{x|v\}$. The difference between fuzzy and sharp design is that $p(v)$ is no longer an indicator function. Our approach is to transform the model so that the transformed model is identical to the sharp design. Let $\widetilde{p}(v)$ denote any positive and continuously differentiable probability function for all v and satisfying $\widetilde{p}(v) = p(v)$ for all $v \geq v_0$. Define

$$\widetilde{y}_i = \frac{y_i}{\widetilde{p}(v_i)} \text{ and } \widetilde{\alpha}(z_i) = \frac{\alpha(z_i)}{\widetilde{p}(v_i)}.$$

It is easy to show that

$$E\{\widetilde{y}_i|z_i = z\} = \widetilde{\alpha}(z) + \beta(z)\widetilde{x}$$

where $\widetilde{x} = 1\{v \geq v_0\}$ is the eligibility status variable. If we treat any eligible individuals as "treated" regardless of whether they actually receive the treatment or not, the above equation is the RD with sharp design with $\widetilde{x}$ as the "treatment status" variable, $\widetilde{\alpha}(z)$ as the baseline effect, and $\widetilde{y}$ as the observed effect. Applying Lemma 1 to this model, we identify $\widetilde{\alpha}(z)$ and $\beta(z) = b(w)'t^J(v)$. Since $p(v)$ is identified by regressing x on v, we obtain:

Lemma 2. *For the fuzzy design, suppose that $\widetilde{p}(z)$ has up to m^{th} derivatives with respect to v almost*

everywhere. Assumption 1 identifies $\alpha(z)$ for $z \in \mathcal{Z}$ and $\beta(z)$ for $z \in \mathcal{Z}_+$.

To illustrate, we again use the simplest case. The probability of receiving treatment can be estimated parametrically by probit or logit regression of x on v and v^2 or nonparametrically. Denote

$$\begin{aligned}\widetilde{h}(v) &= \frac{dE\{\widetilde{y}|v,\widetilde{x}=0\}}{dv} \text{ for } v < v_0 \text{ and } \widetilde{h}(v) = \frac{dE\{\widetilde{y}|v,\widetilde{x}=1\}}{dv} \text{ for } v \geq v_0,\\ \widetilde{\alpha}(v) &= \int \widetilde{h}(v)dv + \widetilde{c} = \widetilde{g}(v) + \widetilde{c},\end{aligned}$$

$$\widetilde{c} = E\left\{\widetilde{y}_i - \widetilde{g}(v_i)|\widetilde{x}_i = 0\right\}.$$

The treatment effect is given by

$$\theta_0 = E\left\{\widetilde{y}_i - \widetilde{g}(v_i) - \widetilde{c})|\widetilde{x}_i = 1\right\}.$$

The paper is organized as follows: Section 2 details an estimation strategy, Section 3 discusses the consequence if the treatment effect function is misspecified, and Section 4 concludes the paper.

3 Estimation

As we discussed above, the key step for estimating the treatment effect is to estimate the baseline effect. To estimate the baseline effect, we need to estimate the conditional mean function. We will use the sieve regression which is also applied by Gallant and Nychka (1987), Andrews (1994), Newey (1997), Ai and Chen (2003), Ai (2005a,b) and others. Specifically, let $s(w) = (s_1(w), s_2(w), \cdots)$ denote a sequence of known basis functions that can approximate any measurable function $c(w)$ arbitrarily well in the sense that there exist coefficients $(\pi_1, \pi_2 \cdots)$ such that

$$\left\| c(w) - \sum_{j=1}^{K_2} s_j(w)\pi_j \right\|_\infty = \sup_{z\in\mathcal{Z}} \left| c(w) - \sum_{j=1}^{K_2} s_j(w)\pi_j \right| \to 0 \text{ as } K_2 \to +\infty.$$

Examples of the basis functions include power series and B-splines. Let $t(v) = (t^J(v)', t_{J+1}(v), t_{J+2}(v), \cdots)$ denote the known basis functions that can approximate any measurable function of v arbitrarily well under the norm $\|\cdot\|_\infty$. Here we include $t^J(v)$ in the basis functions for a technical convenience. For some integers $K_1(> J)$ and K_2, denote $t^{K_1}(v) = (t^J(v)', t_{J+1}(v), ..., t_{K_1}(v))'$ and

$s^{K_2}(w) = (s_1(w), s_2(w), \cdots, s_{K_2}(w))'$. With $K = K_1 K_2$, denote

$$q^K(z) = t^{K_1}(v) \otimes s^{K_2}(w),$$

where $\otimes$ is the Kronecker product. Then, for any measurable function $f(z)$, there exist coefficients π_K such that

$$\left\| f(z) - q^K(z)'\pi_K \right\|_\infty = \sup_{z \in \mathcal{Z}} \left| f(z) - q^K(z)'\pi_K \right| \to 0 \text{ as } K_1, K_2 \to +\infty.$$

Moreover, if $f(z)$ is continuously differentiable with respect to v, we will assume

$$\left\| \frac{\partial^s f(z)}{\partial v^s} - \frac{\partial^s q^K(z)'\pi_K}{\partial v^s} \right\|_\infty = O(K^{-r}) \text{ uniformly in } K_1, K_2 \text{ and for } s = 0, 1, ..., m.$$

These approximation results permit us to use the truncated series $q^K(z)$ and $s^{K_2}(w)$ to approximate the unknown functions, and use $\frac{\partial^m q^K(z)}{\partial v^m}$ to approximate the derivative.

Let $\{(y_i, z_i, x_i), i = 1, 2, \cdots, N\}$ denote a sample of observations. To simplify exposition and without loss of generality, we assume that the first N_c individuals are ineligible for treatment while the last $N - N_c$ individuals are eligible for treatment.

3.1 Sharp design

Regressing y_i on $q^K(z_i)$ for the ineligible group, we obtain the regression coefficients

$$\widehat{\pi}_{Kc} = (Q_c' Q_c)^{-1} Q_c' Y_c,$$

where $Q_c = (q^K(z_1), q^K(z_2), \cdots, q^K(z_{N_c}))'$ and $Y_c = (y_1, \cdots, y_{N_c})'$. The conditional mean function for ineligible individuals is estimated by

$$\widehat{E}\left\{y_i | z_i = z\right\} = q^K(z)'\widehat{\pi}_{Kc} = q_1^K(z)'\widehat{\pi}_{1Kc} + q_2^K(z)'\widehat{\pi}_{2Kc} \text{ for } v < v_0,$$

where $q_1^K(z) = t^J(v) \otimes s^{K_2}(w)$, and the derivative is estimated by

$$\frac{\partial^m \widehat{E}\left\{y_i | z_i = z\right\}}{\partial v^m} = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2Kc}}{\partial v^m} \text{ for } v < v_0. \tag{3}$$

Regressing y_i on $q^K(z_i)$ for the eligible group, we obtain the regression coefficients

$$\widehat{\pi}_{Kt} = (Q_t'Q_t)^{-1}Q_t'Y_t,$$

where $Q_t = (q^K(z_{N_c+1}), q^K(z_{N_c+2}), \cdots, q^K(z_N))'$ and $Y_t = (y_{N_c+1}, \cdots, y_N)'$. The conditional mean function for eligible individuals is estimated by

$$\widehat{E}\left\{y_i|z_i = z\right\} = q^K(z)'\widehat{\pi}_{Kt} = q_1^K(z)'\widehat{\pi}_{1Kt} + q_2^K(z)'\widehat{\pi}_{2Kt} \text{ for } v \geq v_0,$$

and the derivative is estimated by

$$\frac{\partial^m \widehat{E}\left\{y_i|z_i = z\right\}}{\partial v^m} = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2Kt}}{\partial v^m} \text{ for } v \geq v_0. \tag{4}$$

Notice that the derivative estimators in (3) and (4) should estimate the same function $\frac{\partial^m \alpha(z)}{\partial^m v}$ but the estimates $\widehat{\pi}_{2Kc}$ and $\widehat{\pi}_{2Kt}$ are not necessarily identical. Our proposal is to fit a smoothed curve

$$\widehat{h}(z) = \frac{\partial^m q_2^K(z)'\widehat{\pi}_{2K}}{\partial v^m},$$

where $\widehat{\pi}_{2K}$ solves

$$\widehat{\pi}_{2K} : \min_{\pi_2} \sum_{i=1}^{N_c} (\frac{\partial^m q_2^K(z_i)'\widehat{\pi}_{2Kc}}{\partial v^m} - \frac{\partial^m q_2^K(z_i)'\pi_2}{\partial v^m})^2 + \sum_{i=N_c+1}^{N} (\frac{\partial^m q_2^K(z_i)'\widehat{\pi}_{2Kt}}{\partial v^m} - \frac{\partial^m q_2^K(z_i)'\pi_2}{\partial v^m})^2.$$

The baseline effect is estimated by

$$\widehat{\alpha}(z) = q_2^K(z)'\widehat{\pi}_{2K} + (s^{K_2}(w) \otimes p^J(v))'\widehat{c}$$

where $\widehat{c}$ is obtained by

$$\text{regressing } y_i - q_2^K(z_i)'\widehat{\pi}_{2K} \text{ on } s^{K_2}(w_i) \otimes p^J(v_i), i = 1, 2, ..., N_c.$$

After computing the baseline effect, we estimate the treatment effect function by

$$\text{regressing } y_i - q_2^K(z_i)'\widehat{\pi}_{2K} - (s^{K_2}(w_i) \otimes p^J(v_i))'\widehat{c} \text{ on } t^J(v_i)' \otimes s^{K_2}(w_i)', i = N_c + 1, ..., N.$$

we obtain the OLS estimates $\widehat{\gamma}$ and

$$\widehat{\beta}(z) = \left(t^J(v)' \otimes s^{K_2}(w)' \right) \widehat{\gamma}.$$

The consistency of $\widehat{\alpha}(z)$ and $\widehat{\beta}(z)$ can be easily established using arguments similar to those of Newey (1997) and Ai and Chen (2003).

To illusrate this step, we consider the case $z = v$ and $\beta(z) = \beta_0 + \beta_1 v$. So $J = 2$ and $t^J(v) = (1, v)$. For some $K_1 > 2$, denote $t^{K_1}(v) = (1, v, v^2, ..., v^{K_1})'$. Obtain the least squares estimates $\widehat{\pi}_{Kc} = (\widehat{\pi}_{Kc0}, \widehat{\pi}_{Kc1}, ..., \widehat{\pi}_{KcK_1})$ from

$$\text{regressing } y_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } i = 1, 2, ..., N_c.$$

Then

$$\frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} = 2\widehat{\pi}_{Kc2} + 6v\widehat{\pi}_{Kc3} + ... + K_1(K_1 - 1)v^{K_1 - 2}\widehat{\pi}_{KcK_1} \text{ for } v < v_0.$$

Obtain the least squares estimates $\widehat{\pi}_{Kt} = (\widehat{\pi}_{Kt0}, \widehat{\pi}_{Kt1}, ..., \widehat{\pi}_{KtK_1})$ from

$$\text{regressing } y_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } i = N_c + 1, ..., N.$$

Then

$$\frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} = 2\widehat{\pi}_{Kt2} + 6v\widehat{\pi}_{Kt3} + ... + K_1(K_1 - 1)v^{K_1 - 2}\widehat{\pi}_{KtK_1} \text{ for } v \geq v_0.$$

Obtain $\widehat{\pi}_{2K}$ from

$$\text{regressing } \frac{\partial^2 \widehat{E}\left\{y_i | v_i = v\right\}}{\partial v^2} \text{ on } 2, 6v_i, ..., K_1(K_1 - 1)v_i^{K_1 - 2} \text{ for } i = 1, ..., N.$$

Then we estimate the baseline effect by

$$\widehat{\alpha}(v) = \widehat{c}_0 + \widehat{c}_1 v + (v^2, ..., v^{K_1})\widehat{\pi}_{2K},$$

where $\widehat{c}_0, \widehat{c}_1$ are obtained by

$$\text{regressing } y_i - (v_i^2, ..., v_i^{K_1})\widehat{\pi}_{2K} \text{ on } 1, v_i \text{ for } i = 1, 2, ..., N_c.$$

After computing the baseline effect, we estimate the treatment effect function by

$$\text{regressing } y_i - (v_i^2, ..., v_i^{K_1})\widehat{\pi}_{2K} - (1, v_i)\widehat{c} \text{ on } 1, v_i, i = N_c + 1, ..., N.$$

we obtain the OLS estimates $\widehat{\gamma}$ and

$$\widehat{\beta}(v) = (1, v)\,\widehat{\gamma}.$$

3.2 Fuzzy design

For the fuzzy design, we need to estimate the probability of receiving treatment. A simplest approach is to fit a probit model by

$$\text{probit regression of } x_i \text{ on } 1, v_i, ..., v_i^{K_1} \text{ for } x_i = 1.$$

Denote the fitted values by

$$\widehat{P}_i = \Phi((1, v_i, ..., v_i^{K_1})\widehat{\delta}), i = 1, 2, ..., N$$

where $\widehat{\delta}$ is probit estimate. Denote

$$\widetilde{y}_i = \frac{y_i}{\widehat{P}_i}, i = 1, 2, ..., N.$$

The treatment effect function can be estimated by applying the procedure of sharp design with y_i replaced by $\widetilde{y}_i$.

4 Misspecification Bias

The proposed procedure produces consistent estimates under the functional form restriction (2). Since the true treatment effect function is unknown, it is a legitimate concern that the proposed procedure may produce biased estimates if (2) is not satisfied. In the case that the functional form restriction (2) is false, what does the proposed procedure actually estimate? To investigate this question, consider the sharp design. Notice that differentiation now yields

$$\begin{aligned}
\frac{\partial^m E\{y_i | z_i = z\}}{\partial v^m} &= \frac{\partial^m \alpha(z)}{\partial v^m} \text{ over } z \in \mathcal{Z}_- \text{ almost everywhere,} \\
\frac{\partial^m E\{y_i | z_i = z\}}{\partial v^m} &= \frac{\partial^m [\alpha(z) + \beta(z)]}{\partial v^m} \text{ over } z \in \mathcal{Z}_+ \text{ almost everywhere.}
\end{aligned}$$

The smoothing step in the proposed procedure, in this case, finds an approximation $h_K(z) = q^K(z)'\pi_K$ to $\frac{\partial^m E\{y_i|z_i=z\}}{\partial v^m}$ in the sense that

$$\sup_{z\in\mathcal{Z}_-}\left|\frac{\partial^m\alpha(z)}{\partial v^m} - h_K(z)\right| \to 0 \text{ and } \sup_{z\in\mathcal{Z}_+}\left|\frac{\partial^m[\alpha(z)+\beta(z)]}{\partial v^m} - h_K(z)\right| \to 0.$$

Suppose that $\frac{\partial^m E\{y_i|z_i=z\}}{\partial v^m}$ is bounded everywhere except for the threshold (i.e. $v = v_0$). Then $h_K(z)$ is bounded for all z and all K. Let $b^*(w)'t^J(v)$ denote the least squares projection of $\beta(z)$ and satisfy

$$b^*(w)'t^J(v_0) = \beta(v_0, w) \text{ for all } w, \tag{5}$$

in the sense that

$$b^*(w) : \arg\min E\{[\beta(z) - b(w)'t^J(v)]^2|w\}$$

subject to the constraint (5). Write

$$\beta(z) = b^*(w)'t^J(v) + r_J(z).$$

Let $g_K(z)$ denote the integration of $h_K(z)$ and let $c_K(w)$ solve

$$\min_{c(\cdot)} E\left\{(y_i - g_K(z_i) - c(w_i)'t^J(v_i))^2|w_i = w, x_i = 0\right\}.$$

Then $\alpha_K(z) = g_K(z) + c_K(w)'t^J(v)$ is an approximation to $f_J(z) = \alpha(z)$ over $z \in \mathcal{Z}_-$ and $f_J(z) = \alpha(z) + r_J(z)$ over $z \in \mathcal{Z}_+$. It is worth pointing out that constraint (5) is critical for this approximation result. If this constraint is not satisfied, $f_J(z)$ is discontinuous at the threshold. From the approximation theory, a discontinuous function cannot be approximated by a sequence of functions with bounded derivatives. However, the derivative of $\alpha_K(z)$ is $h_K(z)$ which is bounded for all z and K.

Let $b_K(w)$ solve:

$$\min_{b(\cdot)} E\left\{y_i - \alpha_K(z_i) - b_K(w_i)'t^J(v_i))^2|w_i = w, x_i = 1\right\}.$$

Then, $b_K(w)$ is an approximation to $b^*(w)$. Hence the estimator $\widehat{\beta}(z)$ is a consistent estimator of $\beta^*(z) = b^*(w)'t^J(v)$.

There are two implications of this approximation result. First, it implies that the average treat-

ment effect for threshold individuals can be estimated consistently by

$$\widehat{\mu} = \frac{1}{N - N_c} \sum_{i=N_c+1}^{N} \widehat{\beta}(v_0, w_i)$$

regardless of whether (2) holds. Moreover, because the average treatment effect estimator uses all eligible individuals, it is root-N consistent. The root-N consistency is achieved by exploiting the derivative information which is not utilized by the existing estimators. The second implication is that the approximation error can be small for large J if $t^J(v)$ is part of some approximating basis functions. Thus, for the non-threshold individuals, $\widehat{\beta}(z)$ is an estimate for $\beta(z)$ with small bias if we choose J large enough.

5 Empirical example

5.1 Motivation

Two of the fundamental questions in health economics are how does health insurance affect health and how does it affect health care expenditures? These questions are important to answer because of the large amounts of money spent on both public and private health insurance. For example, during the debates about how best to implement the Patient Protection and Affordable Care Act (ACA), proponents of the ACA argued that extending health insurance would improve population health. Opponents, however, raised the expected higher costs. The RAND Health Insurance Experiment found the expected negative relationship between out-of-pocket payments and health care expenditures in a non-elderly population, but no substantial effect on health (Manning et al., 1987). Other studies that have found health effects of insurance have not convincingly controlled for the endogeneity of insurance.

A number of recent papers have returned to these questions by examing the experience of people who become eligible for Medicare at age 65. This population is interesting because any effects of insurance on morbidity and mortality are expected to be larger than for the non-elderly population and because Medicare is such a large and influential purchaser of health care. In recent papers, Card, Dobkin, and Maestas (2008, 2009) use a regression discontinuity approach to estimate the narrow effect of Medicare insurance on mortality following emergency admissions for certain conditions at the age where most people become eligible—age 65. They find a statistically-significant effect of nearly a 1-percentage point drop in the seven-day mortality, equivalent to a 20% reduction. Furthermore, the results cannot be explained by those who had previously been uninsured; Medicare seems to have an

effect even on those previously insured.

Card, Dobkin, and Maestas (2008, 2009) find small but statistically-significant effects of having Medicare insurance on mortality for a specific group of people: those admitted to the emergency room for severe conditions. Because they used regression discontinuity, with the probability of insurance jumping right at age 65, their estimated effect is defined right at the threshold where eligibility changes. Their estimated effects are precise, immediate, and narrowly defined.

Another view, though, of health insurance is that it has a larger value beyond the immediate effect right at age 65. Having health insurance may improve access to care, provide better continuity of care, and increase prevention. None of these effects are immediate, such as estimated by Card, Dobkin, and Maestas. Instead, they would be cumulative over time. Improving access to care, for example, affects health care use and outcomes in the current period but also in the future. This dynamic view of the importance of health insurance has been shown to be important in assessing the effects of prescription drug insurance (Yang, Gilleskie, Norton, 2009). A regression discontinuity design cannot reveal whether health insurance has such longer-term effects. Traditional regression discontinuity can only estimate the immediate effect of health insurance right at the discontinuity (Hahn, Todd, Van der Klaauw, 2000). This is why it is important to find models that allow for estimation of treatment effects beyond the jump point.

Polsky and colleagues (2009) take a different approach to answering a similar question about the health effects of Medicare insurance. They focus on those who are not insured just before age 65. They estimate Markov transition models to estimate the transition probabilities between several states of health, including death, and compare those insured before age 65 to those uninsured, and what happens after becoming Medicare eligible. They find no statistically significant effects of Medicare on health trajectories, including the probability of death by age 73.

5.2 Simple example

Consider a simple specific example, based on Medicare eligibility at age 65. Age is the continuous variable z. To focus on the mechanics of estimating the model, assume that Medicare has both a discrete effect on expenditures at age 65 and a linear effect over time. Therefore, $\beta(z)$ is a linear function of z, and $m = 2$. The reason that $m = 2$ is because m is one more than the order of the treatment function. So for linear treatment, $m = 2$, and for quadratic treatment, $m = 3$. Furthermore, assume that $\alpha(z)$ is a cubic equation. Under these assumptions, here is the recipe for applying our regression discontinuity method. We ignore other covariates in this example; they can easily be added.

1. Regress y on $(age-65)$, $(age-65)^2$, and $(age-65)^3$ for the sample of those under age 65.

$$y = \alpha_0 + \alpha_1 (age-65) + \alpha_2 (age-65)^2 + \alpha_3 (age-65)^3$$

2. Generate a new variable y_{2deriv} equal to the second derivative of the estimated y with respect to $(age-65)$ for the sample of those under age 65. The reason to estimate the second derivative is that we assumed that $m = 2$.

$$y_{2deriv} = 2\hat{\alpha}_2 + 6\hat{\alpha}_3 | age < 65$$

3. Regress y on $(age-65)$, $(age-65)^2$, and $(age-65)^3$ for the sample of those at least age 65.

$$y = \beta_0 + \beta_1 (age-65) + \beta_2 (age-65)^2 + \beta_3 (age-65)^3$$

4. Generate a new variable y_{2deriv} equal to the second derivative of the estimated y with respect to $(age-65)$ for the sample of those at least age 65.

$$y_{2deriv} = 2\hat{\beta}_2 + 6\hat{\beta}_3 | age \geq 65$$

5. Regress y_{2deriv} on just $(age-65)$ for the full sample to estimate a constant term and a slope.

$$y_{2deriv} = \gamma_0 + \gamma_1 (age-65)$$

6. Generate a new variable equal to y minus the double indefinite integral of y_{2deriv} with respect to age for the full sample.

$$y_{new} = y - \frac{\hat{\gamma}_0}{2} (age-65)^2 + \frac{\hat{\gamma}_1}{6} (age-65)^3$$

7. Regress y_{new} on $(age-65)$ for the sample of those under age 65 to estimate the two terms not defined by the double indefinite integral.

$$y_{new} = \delta_0 + \delta_1 (age-65) | age < 65$$

8. Generate a new variable y_α that is the difference between y and the double integral of y_{2deriv}

with respect to age, including information from the previous two regressions for the full sample.

$$y_\alpha = y - \hat{\delta}_0 - \hat{\delta}_1 (age - 65) - \frac{\hat{\gamma}_0}{2} (age - 65)^2 + \frac{\hat{\gamma}_1}{6} (age - 65)^3$$

9. Run a regression of y_α on $(age - 65)$ for the sample at least age 65.

$$y_\alpha = \theta_0 + \theta_1 (age - 65)$$

It is important to compare our method with just running a simple regression with least squares regression on a cubic equation of age. The difference is the convergence rate (not efficiency; of course the estimator with faster convergence rate is more accurate). If both the treatment effect and the base outcome are correctly parameterized, then it is a parametric problem. If the base outcome is incorrectly parameterized—as is likely—then our estimator converges at square root of sample size, while the simple regression method converges at much slower rate.

Before running any models, it is important to check the specification of the polynomial for age, the variable of interest, for the baseline effect (the polynomial in $\alpha(z)$). One way to do that is to run a cross-validation. In ordinary least squares, including more polynomials always improves the R-squared, but can lead to over-fitting. If there is not a strong theoretical reason to choose a particular specification, we recommend running a cross-validation to check the specification in the following way. Start with a quadratic function in age. Run N models, each time leaving out one observation. Compute the fitted value and then the prediction error for the dropped observation. Sum the squared prediction errors over all observations. Do this again with a cubic function. Continue until adding more polynomials does not improve the fit as measured by the sum of squared prediction errors. Use the model specification that minimizes the sum of squared prediction errors.

6 Data

We used hospital discharge data from the Nationwide Inpatient Sample (NIS) from 2001 through 2003. The NIS is part of the Healthcare Cost and Utilization Project (HCUP), a federal-state-industry partnership sponsored by the Agency for Healthcare Research and Quality. The NIS provides key data elements for this study. The NIS has a large sample with information about admissions and health care expenditures across many states and several years. It is the largest all-payer inpatient care database in the United States—with data from approximately 8 million hospital stays from up 38 states each year (the participant states vary from year to year). The NIS represents approximately

a 20% stratified sample of U.S. community hospitals. Each NIS record contains information regarding the primary and secondary diagnoses, procedures, the admission and discharge status, patient demographics, length of stay, total charge and expected payment source. Hospital characteristics (ownership, size, teaching status, etc) are also recorded.

The unit of analysis is at the state-year-age level. Information from the NIS is aggregated by state and year for each age between ages 45 and 85. In our final sample for analysis, there are 3,680 observations from 35 states; this is less than 4,305 = (41 ages)×(35 states)×(3 years) because of missing data for certain states in some years. There are 1,120 observations in 2001, 1,200 in 2002, and 1,360 in 2003, States included in our analysis are: AZ, CA, CO, CT, FL, GA, HI, IL, IN, IA, KS, KY, ME, MD, MA, MI, MN, MO, NV, NH, NJ, NY, NC, OH, OR, RI, SC, SD, TN, UT, VT, VA, WA, WV, and WI.

The two dependent variables of admissions and hospital costs are constructed as per capita rates, where the per capita rates are specific to the state and year and age group.. By merging the NIS sample from HCUP with Census data at state level, we calculated the admission rate per 10,000 people and hospital care cost per 1,000 capita as the dependent variables. Because the NIS does not provide actual dollar values of the cost as paid by the payers, we used the cost-to-charge ratio file provided by HCUP to convent the inpatient total charges to total cost. Such ratio is based on the cost-to-charge ratio at hospital level. The cost is measured in 2003 dollars, based on the Consumer Price Index of medical goods published by Bureau of Labor Statistics.

We control for state-level variables that may affect supply and demand by merging the NIS with variables from the Area Resource Files. These variables include the average income per capita, unemployment rate, and high school graduation rate as state-level factors that affect demand. We also included number of hospital beds per 1,000 capita and the number of physicians per 1,000 capita as state-level factors that affect supply. The summary statistics are presented in Table 1.

7 Results

Initial regression results show that admissions increase with age at an increasing rate both before and after age 65, but the magnitidudes indicate that the marginal effect of age is even larger after age 65.. Admissions also decline with income and education, consistent with health being positively related to income and education. Perhaps surprisingly, the unemplyment rate—which is usually related to the level of private health insurance among non-elderly adults—does not affect admissions. In terms of supply-side variables, admissions are higher in areas with more beds per capita. Doctors per capita are unrelated to admissions. The signs and significance of the results are similar for those at least

age 65 and for those under age 65, however the magnitudes appear generally large for the elderly.

For costs, the linear effect of age is positive and significant. The quadratic term is positive and significant only for those under age 65; for the elderly the coefficient on the quadratic term is negative and insignificant. For costs, income (positive) and education (negative) now have opposite signs. More dotors per capita decrease costs while more beds increase costs. Again, the results are generally similar in sign but not magnitude for the two age groups.

We next turn out attention to the main results from the newly proposed model.

The cross validation shows that the best fitting model for admissions has a squared term for age (see Table 5). In constrast, a fourth-order polynomial fits best for costs (although cubic is nearly as good). Therefore, we focus on models in which the baseline effect $\alpha(z)$ of age is squared for admissions (model "21" in Table 4) and fourth-order for costs (models "41 and "42" in Table 4).

We ran five different versions of the model, with different specifications for age (see Table 4). The model specifications are distinguished by two numbers, the first is the polynomial for age for the baseline effect $\alpha(z)$ and the second is the polynomial in age for the treatement effect $\beta(z)$. For example, model "21" means that the baseline effect is quadratic and the treatment effect is linear. Identification requires that the baseline effect have at least one more polynomial in age than the treatment effect.

The results imply that the treatment effect of Medicare is positive on both admissions and costs for individuals beyond age 65 (see Table 4). The coefficient on age in the "21" admissions model is positive, indicating that admissions per capita increase with age due to Medicare beyond the trending effect of age modeled prior to age 65. For costs, the coefficient on age in model "41" and both age and age squared in model "42" again indicate that costs increase with age beyond the trending effect of age. To get a sense of the magnitudes, we graphed the results for all five models. The preferred models are graphed in bold lines. Age is on the x-axis and either admissions or costs are on the y-axis. The results imply that Medicare has an effect that changes over time, in this case getting stronger. As discussed before, this is not surprising due to the cumulative effects of access to care.

8 Conclusions

In this paper, we propose a new estimation method for regression discontinuity model under the additional assumption that the treatment effect function satisfies some functional form restriction, and the average outcome in the absence of treatment has higher order of continuous derivatives. These additional assumptions allow us to remove the treatment effect through differentiation and to use all observations, not just the few near the threshold, to estimate the model. Thus, our procedure

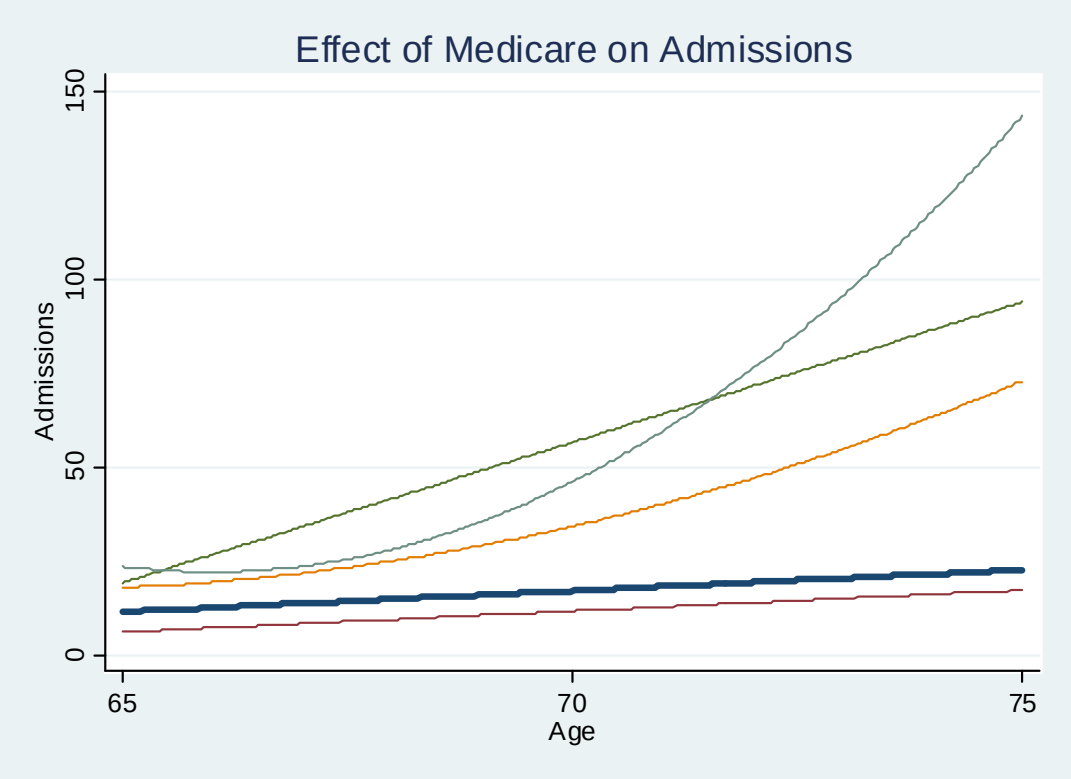
Effect of Medicare on Admissions
Admissions
0
50
100
150
65
70
75
Age

Figure 1:

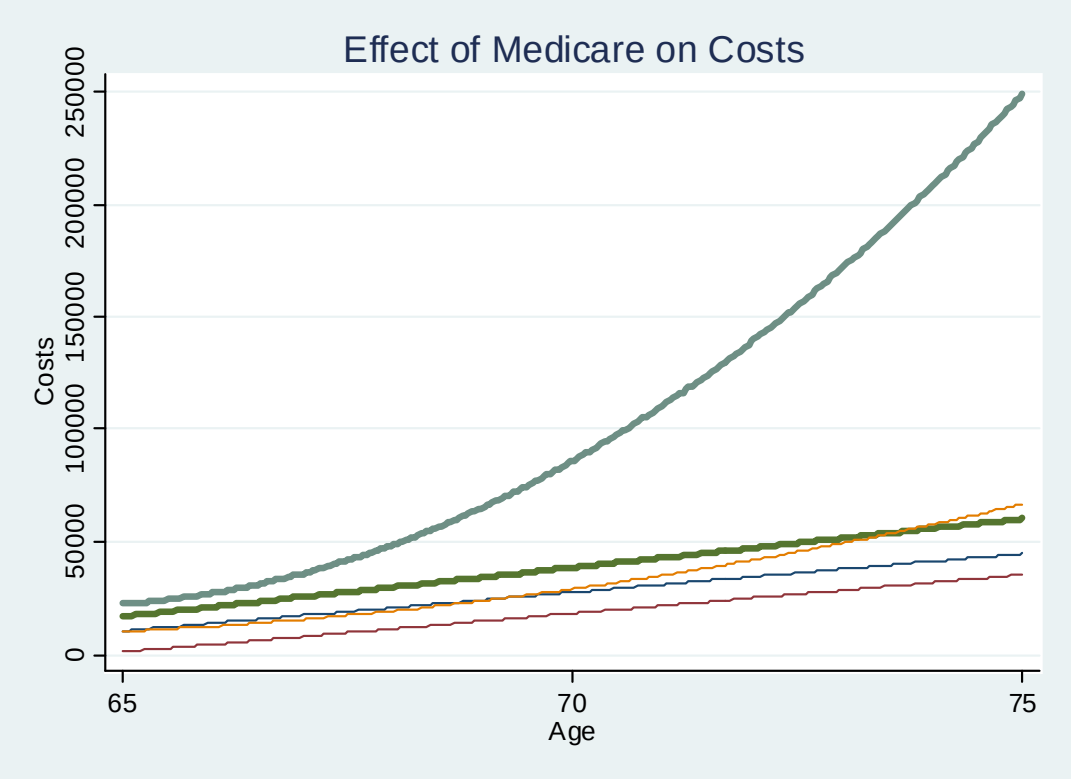
Effect of Medicare on Costs
Costs
0
50000
100000
150000
200000
250000
65
70
75
Age

Figure 2:

yields better estimates than the existing procedures. We also discuss the consequence of misspecifying the treatment effect function, and find that our estimators still provide useful information about the treatment effect even if our parameterization is false.

Our estimation method could be applied to any empirical model where the researcher is interested in the treatment effect not only at the jump point, but beyond. In health economics, this would apply to most program evaluations. We are typically interested not only in the immediate effect but most programs have long-run and cumulative effects. The full story may abe of far greater interest than the simple story that traditional RD can tell.

References

Ai C. 2005a. A Nonparametric Maximum Likelihood Estimation of Conditional Moment Restrictions Models. University of Florida working paper.

Ai C. 2005b. An Alternative Root-N Consistent Estimator for Panel Data Binary Choice Models. University of Florida working paper.

Ai C. 2008. A Root-N Consistent Estimation of Regression Discontinuity Models. University of Florida working paper.

Ai C, Chen X. 2003. Efficient Estimation of Conditional Moment Restrictions Models Containing Unknown Functions. *Econometrica* 71: 1795–1843.

Ai C, Chen X. 2004. Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditional Variables. *Journal of Econometrics* 141(1): 5–43.

Card D, Dobkin C, Maestas N. 2008. The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. *American Economic Review* 98: 2242–2258.

Card D, Dobkin C, Maestas N. 2009. Does Medicare Save Lives? *The Quarterly Journal of Economics* 124: 597–636.

Gallant AR, Nychka DW. 1987. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica* 55: 363–390.

Hahn J, Todd P, Van der Klaauw W. 2000. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica* 69: 201–209.

Manning WG, Newhouse JP, Duan N, Keeler EB, Leibowitz A, Marquis MS. 1987. Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *American Economic Review* 77: 251–277.

Newey WK. 1997. Convergence rates and Asymptotic Normality for Series Estimators. *Journal of Econometrics* 79: 147–168.

Polsky D, Doshi JA, Escarce J, Manning WG, Paddock SM, Cen L, Rogowski J. 2009. The Health Effects of Medicare for the Near-Elderly Uninsured. *Health Services Research* 44: 926–945.

Porter J. 1998. Estimation of Regression Discontinuity. Manuscript.

Yang Z, Gilleskie DB, Norton EC. 2009. Health Insurance, Medical Care, and Health Outcomes: A Model of Elderly Health Dynamics. *Health Economics* 44: 47–114.

Table 1: Summary statistics of the NIS Sample by age and state: 2001–2003

	Mean	Std. Dev.
Dependent variables		
Admission rate per 10,000 capita	285	190
Hospital cost per 1,000 capita	$204,249	$136,265
Independent variables		
Age	64.5	11.5
Unemployment rate	5,18	1.03
Income per capita	$30,992	$4,455
High school graduation rate	71.3%	9.61%
# of MDs per 1,000 capita	2.61	0.62
# of Hospital beds per 1,000 capita	3.55	0.88

Table 2: Results of initial regression for admissions per 10,000 capita, by age

Admissions	If Age ≤ 65	If Age ≥ 65
Age	12.524 **	13.68 **
	(0.829)	(1.92)
Age^2	.2778 **	0.3192 **
	(0.0383)	(0.0976)
Unemployment rate	1.86	–1.99
	(1.18)	(3.00)
Income per capita	–.000941 *	–0.00290 **
	(0.000414)	(0.00105)
HS graduation rate	–1.467 **	–3.081 **
	(0.129)	(0.328)
MDs per capita	–2.37	3.47
	(2.76)	7.03
Beds per capita	13.19 **	36.93 **
	(1.41)	(3.58)
Constant	323.6 **	432.5 **
	(16.3)	(41.1)
N	1,840	1,840

Table 3: Results of initial regression for cost per 1,000 capita, by age

Hospital Costs	If Age ≤ 65	If Age ≥ 65
Age	10,498 **	13,600 **
	(711)	(1559)
Age^2	211.7 **	–119.1
	(32.9)	(79.2)
Unemployment rate	–231	–382
	(1009)	(2,431)
Income per capita	1.142 **	3.811 **
	(.355)	(.855)
HS graduation rate	–1,076 **	–2,214 **
	(111)	(266)
MDs per capita	–13,681 **	–19781 **
	(2,367)	(5,700)
Beds per capita	2,992 *	7,972 **
	(1,207)	(2,905)
Constant	253,863 **	252,964 **
	(13,976)	(33,353)
N	1,840	1,840

Table 4: Results of final model, for admissions and for cost

Model		Admissions	Cost
21	Age	1.118 *	3,432 **
		(.540)	(415)
	Constant	11.40	10,511 *
		(6.00)	(4,610)
31	Age	1.118 *	3,432 **
		(.540)	(415)
	Constant	6.00	1,103
		(6.00)	(4,610)
41	Age	7.484 **	4,352 **
		(.540)	(415)
	Constant	19.17 **	16,872 **
		(6.00)	(4,611)
32	Age	1.04	1,888
		(2.07)	(1,589)
	Age^2	.447 **	376.6 **
		(.105)	(80.7)
	Constant	17.81 *	10,311
		(8.48)	(6,514)
42	Age	–2.97	2,767
		(2.07)	(1,589)
	Age^2	1.499 **	1,987.5 **
		(.105)	(80.7)
	Constant	23.48 **	22,507 **
		(8.48)	(6,514)

Table 5: Cross Validation

	Admissions	Hospital Cost
Age	43,537,696	2.389e+13
Age^2	**38,052,316**	2.306e+13
Age^3	38,071,676	2.287e+13
Age^4	38,093,160	**2.286e+13**
Age^5	38,116,688	2.287e+13
Age^6	38,140,280	2.288e+13