Healthcare Facility Choice and User Fee Abolition: Regression Discontinuity in a Multinomial Setting*

Steven F. Koch †

May 2, 2013

Abstract

We consider a multinomial logit, as well as a nonparametric conditional mode analysis.

*The author would like to thank Jeffrey S. Racine for his help with coding, as well as seminar participants at the University of the Free State and Emory University, as well as participants at the workshop the Microeconometric Analysis of South African Data for their comments on this work. All remaining errors are the sole responsibility of the author.

†Professor and Head, Department of Economics, University of Pretoria, Pretoria, Republic of South Africa; (O) 27-12-420-5285, (F) 27-86-691-2749. *steve.koch@up.ac.za*

1 Introduction

Although Thistlethwaite & Campbell's (1960) idea did not, initially, receive much attention in economics, regression discontinuity (RD) applications have become increasingly common. The past fifty years of applications were recently reviewed in van der Klaauw (2008) and Lee & Lemieux (2010), and, if having a guide to application is any indication - see Imbens & Lemieux (2008) - RD is likely to underpin empirical examinations of policy impacts for the forseeable future. Part of RD's appeal lies in visualization. In many cases, at least in published analyses, it is possible to illustrate large changes in average outcomes at the RD threshold, even when that threshold is fuzzy.

Regardless of whether the threshold is fuzzy or the outcome variable is continous, RD analyses tend to focus on average treatment effects, possibly local, through the application of ordinary least squares or two-stage least squares. Although Imbens & Lemieux's (2008) guide suggests local linear, or nonparametric, regression, as well, local linear regression is limited to estimation in the dimension of the running variable.

Examples of this approach abound in the applied economics literature. Within education and labour, Angrist & Lavy's (1999) seminal analysis, as well as, Lemieux & Milligan (2008), Calcagno & Long (2008) and Lalive (2008) use linear models in their analyses. Although Angrist & Lavy (1999) have continuous outcome variables, the remaining papers analyze discrete outcome data. The outcome variable in Lalive (2008), for example, is duration data.

Within the health economics literature, where our research is most logically placed, the approaches are similar. Palangkaraya & Yong (2007) follow a linear probability model, while Coe & Zamarro (2011) use two-stage least squares, even though, for the most part, outcomes in their analyses are either binary or categorical. The same can be said for other, recent, RD applications in health economics; see Zuckerman, Lee, Wutoh, Xue & Stuart (2006), Card, Dobkin & Maestas (2008), Carpenter & Dobkin (2009), Lindeboom, Llena-Nozal & van der Klaauw (2009), Silles (2009) and McCrary & Royer (2011). Although the early RD literature, see, for example, Berk & Rauma (1983), did not model categorical outcomes within linear settings, the same cannot be said for the more recent literature. We were only able to locate Albouy & Lequien (2009)

and Ou (2010) making use of nonlinear probability models within an RD setting.

However, forcing linear models on discrete outcomes can be problematic, even when those discrete outcomes are ordered. In the case of binary outcomes, it is possible that the predicted outcome for any individual, on either side of the RD threshold, lies outside of the unit interval. Although unlikely, none of the previously mentioned studies either discusses or tests for the possibility. With respect to the policy impact, more importantly, predictions outside the unit inverval would tend to yield overestimates of the policy impact. For duration data, although incorrect probability estimates are a problem, the incorrect predictions are more likely due to duration dependence, under which, individuals with similar durations are likely to have common unobserved factors affecting the outcomes. Ignoring duration dependence, as would be the case under least squares estimation, could also lead to bias in the estimated treatment effect. Finally, in the case of unordered categorical outcomes, linear models are simply inappropriate. Although it is feasible to estimate separate linear regressions for every binary pair in the set of categorical outcomes, such an analysis ignores the potential that predictions fail the simple rules of probability. Furthermore, such an analysis ignores the potential for unobserved categorical dependence, as the relationship between categorical outcomes is forced to be independent of other outcomes.[1]

In this research, we use a regression discontinuity design to examine the effect of public health sector user fee abolition on healthcare-seeking behavior for ill children in South Africa. The policy explicitly removed healthcare fees for children under the age of six, as well as pregnant and nursing mothers, and the elderly. However, healthcare services in South Africa are provided by both the private and public sectors. Therefore, ill children can receive treatment in either the public or the private sector, or not receive any treatment. The effect of user fee abolition on the use of public healthcare services is analyzed taking account of the trivariate set of unordered healthcare-seeking options. In addition to the linear index multinomial logit model, which ignores unobserved heterogeneity, we also estimate a nonparametric multinomial outcome model that is not

[1]This problem, within the ordinary least squares setting, is qualitatively similar to violating the Independence of Irrelevant Alternatives (IIA) assumption in the multinomial logit regression setting. We will discuss this in more detail, below.

similarly restricted. The nonparametric model is found to fit the data better than the standard multinomial logit model. The results suggest that the linear index assumption, which is not much different from the linearity assumption employed in the majority of the previous RD studies, the unobserved heterogeneity assumption or both assumptions are inappropriate, at least in this setting.

Furthermore, from the models, the distrubution of average treatment effects for the sub-population most likely to be affected by the policy is predicted. The predictions, illustrated in a series of figures, describe non-constant treatment effects. The results raised further questions regarding the empirical focus in the literuture, which is currently placed on pooled linear regression models with an RD indicator. In other words, the results further question the typically held assumption that treatment effects are constant.

2 Methodology

Although the user fee policy change announced in 1994 had a number of components, this analysis focuses only on the demand for curaticve care services for children under the age of six, as the data does not make it possible to consider preventative care, antenatal care or effects related to nursing mothers. The demand for curative care services is analyzed within the context of healthcare facility choice. guptadasgupta2002, amongst others, note that provider choice decisions are primarily related to curative care.

The user fee abolition policy is based on an age threshold, and, therefore, the analysis will be based on the application of regression discontinuity (RD).[2] The age data, described below, is generally only available in years, although it is possible to merge exact birthdates from the survey, allowing for a more general analysis. However, that data is available only for children living with their mothers. For that reason, we undertake two separate analyses: one in which only an indicator of eligibility is used as the policy variable, and another, in which age is treated as a continuous variable, to see if the results are dependent upon this choice of specification.

As noted previously, the policy was designed to improve access to healthcare within

[2]In a companion paper, the 1994 and 1996 policy changes are examined within a differences-in-differences context.

the public sector, although other healthcare-seeking options are available for ill children. These options, such as care within the private sector and non-treatment, are potential substitutes for public care. Therefore, the analysis is placed within a three-outcome model of healthcare facility choice. Typically, an analysis of this sort would follow a multinomial logit or probit framework. However, in this analysis, in addition to the multinomial logit setting, which is common in the literature, we also undertake nonparametric analysis.

2.1 Must fit this in, somehow

Berk & Rauma (1983) use only a dummy variable for the RD, but include a number of intuitive discussions related to some much older papers Goldberger (1972*a*), Goldberger (1972*b*) and Rubin (1977) that should probably be considered. However, the intuition in the paper is decent. The intuition:

- the model includes the running variable **and** the treatment dummy
- assignment only due to running variable, and no fuzziness
- to run a pooled regression, the effect has to be a parallel shift; this is reaonably obvious in the linear regression model
- with interactions, there are multiple treatment effects, and an approach to reporting ought to be considered; although the mean is an option, it is not too obvious that is the best. we apply something rather more interesting, I think.
- the high correlation between the running variable and the treatment dummy leads to some variance inflation of the treatment effect.
- additional controls can be included (and should be to increase power)

Lee & Card (2008) tries to examine specification error in the functional form. Interestingly, they, like most others, (except, Imbens & Lemieux (2008)?) suggest that the empirical analysis should straddle the discontinuity. In that setup, there is a version of the curse of dimensionality complaint. Anyway, the solution revolves around a clustering procedure, which they tie to a Bayesian approach.

2.2 Mulitnomial logit

Denote, with Y_i, with realizations y_i, a categorical indicator of health facility choice, which takes on the values $j \in \{0, 1, 2\}$.

$$Y_i = \begin{cases} 0, & \text{No treatment sought} \\ 1, & \text{Treatment at a public facility} \\ 2, & \text{Treatment at a private facility} \end{cases} \tag{1}$$

Furthermore, assume that there is a vector of explanatory variables, denoted by X_i, which have realizations x_i in the data. These are assumed to represent socioeconomic and demographic characteristics of the ill child, including a function of the child's age; in further discussions, below, we will make the age function explicit, given its central role in the analysis. Continuing with standard convention, we define p_{ij} to be the probability that ill child i receives treatment j, i.e., $p_{ij} = \text{prob}(y_i = j | X_i = x_i)$. By assumption, $\sum_j p_{ij} = 1$, such that parameters in the model can only be identified, relative to a base category. Without loss of generality, $j = 0$, or nontreatment, will be that base category.

Finally, assuming that the stochastic terms are *iid* and follow an extreme value distribution, while the explanatory variables follow a linear index formulation, the underlying probabilities take on the familiar multinomial logit structure. The coefficients, β_1 and β_2, are the coefficients for outcome choices 1 and 2, respectively, and they are relative to nontreatment, outcome 0.

$$p_{i0} = \left(1 + \sum_{k=1}^{2} e^{x_i \beta_k}\right)^{-1} \tag{2}$$

$$p_{i1} = e^{x_i \beta_1} \left(1 + \sum_{k=1}^{2} e^{x_i \beta_k}\right)^{-1} \tag{3}$$

$$p_{i2} = e^{x_i \beta_2} \left(1 + \sum_{k=1}^{2} e^{x_i \beta_k}\right)^{-1} \tag{4}$$

The multinomial logit model can be estimated via maximum likelihood. For any ill

child, the contribution to the log-likelihood is

$$\ln \mathcal{L}_i(\beta) = \sum_{j=0}^{2} \mathbf{1}[y_i = j] \ln p_{ij} \tag{5}$$

In (5), the indicator function, $\mathbf{1}[y_i = j]$, kicks in for the alternative chosen for the ill child. The model is estimated using the nnet package, Venables & Ripley (2002), in R, R Core Team (2012).

Underlying this structure is the independence of irrelevant alternatives (IIA) assumption, wherein the odds ratios derived in the model do not depend on the number of choices available. For example

$$\frac{p_{i1}}{p_{i2}} = \frac{e^{x_i\beta_1}}{1 + \sum_{k=1}^{2} e^{x_i\beta_k}} \Bigg/ \frac{e^{x_i\beta_2}}{1 + \sum_{k=1}^{2} e^{x_i\beta_k}} = e^{x_i(\beta_1 - \beta_2)} \tag{6}$$

is completely independent of the base choice, and for any other choices that could be added to the set of outcomes. Although IIA is a testable assumption, Small & Hsiao (1985), it will not be formally tested. Instead, the predictive performance of the multinomial logit will be compared to the predictive performance of the nonparametric model; the comparison is outlined, below. It is also true that IIA can be relaxed in a number of different ways, through the nesting of alternatives, the allowance of random parameters, or assuming normally distributed, but correlated, stochastic terms.

2.3 Nonparametric conditional mode estimation

Although IIA can be relaxed in a number of different ways, most of the options remain restrictive, and are, at least to some degree, arbitrary. For example, nesting requires the econometrician to assume that decisions are made in groups. An analyst might be willing to assume that a caregiver first decides whether or not an ill child should be treated, and once that decision is made, a decision on where to treat might be made. However, there is no reason to believe that nesting structure is necessarily correct. On the other hand, assuming normality imposes a distribution on the error structure that may not be correct. Therefore, we adopt a nonparametric analysis of the outcomes.

Begin by defining f and m as the joint and marginal densities of (X, Y) and X,

respectively, where Y represents the unordered categorical outcomes associated with health facility choice, while X can include continuous, ordered and unordered categorical variables. Mathematically, the conditional density of $Y = y$, given $X = x$ is

$$g(y|x) = \frac{f(x,y)}{m(x)} \tag{7}$$

An estimate of the conditional density can be formulated from the kernel estimates of the underlying joint and marginal densities, $\widehat{f}$ and $\widehat{m}$. Replacing the unknown densities in (7) with their estimates, yields an estimate of the conditional density of $Y = y$, given $X = x$.

$$\widehat{g}(y|x) = \frac{\widehat{f}(x,y)}{\widehat{m}(x)} \tag{8}$$

Given the mix of continuous and discrete, which could be either ordered or unordered, variables, Li & Racine's (2003) generalized product kernel is used in the estimation. Following Li & Racine (2003), let $X = (X^c, X^u, X^o)$ denote a split of X into s continuous, t discrete unordered and r discrete ordered variables. The marginal density m for realizations x is given by

$$\begin{aligned} \widehat{m}(x) &= \widehat{m}\,(x^c, x^u, x^o) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\prod_{k=1}^{s} W(X_{ik}^c, x_k^c)\prod_{k=1}^{t}\ell^u(X_{ik}^u, x_k^u)\prod_{k=1}^{r}\ell^o(X_{ik}^o, x_k^o)\right]. \end{aligned} \tag{9}$$

Similarly, the joint density f for realizations (x, y) is given by

$$\begin{aligned} \widehat{f}(x,y) &= \widehat{f}\,(x^c, x^u, x^o, y^u) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\prod_{k=1}^{s} W(X_{ik}^c, x_k^c)\prod_{k=1}^{t}\ell^u(X_{ik}^u, x_k^u)\prod_{k=1}^{r}\ell^o(X_{ik}^o, x_k^o)\times\ell^u(Y_i^u, y^u)\right]. \end{aligned} \tag{10}$$

Within the structure of equations (9) and (10), there are three different data types, and, therefore, three different kernels are used for estimationof these densities: a second-order Gaussian kernel for continuous variables, Wang & van Ryzin's (1981) kernel for ordered categorical variables, and the Li & Racine (2007) kernel for unordered categorical

variables.[3] For positive bandwidth $h_k > 0$,

$$W(X_{ik}^c, x_k^c) = \frac{1}{h_k} K\left(\frac{X_{ik}^c - x_k^c}{h_k}\right)$$
$$K(\cdot) = e^{-z^2/2} \Big/ \sqrt{2\pi}\,, \qquad z = \frac{|X_{ik}^c - x_k^c|}{h_k} \tag{11}$$

For values of $\lambda_k \in [0, 1]$,

$$\ell^u(Z_{ik}^u, z_k^u) = \begin{cases} 1 - \lambda_k & \text{if } |z_k^u - Z_{ik}^u| = 0 \\ \frac{1-\lambda_k}{2} \lambda_k^{|z_i^u - Z_{ik}^u|} & \text{if } |z_k^u - Z_{ik}^u| \geq 1 \end{cases} \tag{12}$$

and, for $\gamma_k \in [0, 1]$,

$$\ell^o(Z_{ik}^u, z_k^u) = \begin{cases} 1 & \text{if } z_k^u = Z_{ik}^u \\ \gamma_k & \text{if } z_k^u \neq Z_{ik}^u \end{cases} \tag{13}$$

Although other kernels can be used, the estimates are not too sensitive to the choice of the kernel, Li & Racine (2007). Instead, it is the choice of bandwidths (h, λ, γ) that is paramount. If the bandwidth is too small, undersmoothing results, in which the bias has been reduced at the cost of increased variance. On the other hand, if the bandwidth is too large, oversmoothing arises, wherein variance is reduced at the cost of increased bias. Intuitively, choosing bandwidths involves the optimal balance between bias and variance. Hall, Racine & Li (2004) outline that inuition, finding that appropriate bandwidths can be estimated via least-squares cross-validation (LSCV), which we employ. Their result follows from the minimization of the weighted integrated square error. Although computationally intensive, LSCV has the added advantage of effectively removing irrelevant variables from the analysis. Estimation is undertaken via the np package, Hayfield & Racine (2008), in R, R Core Team (2012).

2.4 Policy impacts

In order to determine the impact of policy, we consider the general relationship between age and all health facility choices, even though the focus of the policy is on public

[3] In the analysis, binary variables are described as ordered. This does not impact the results, but also simplifies the policy evaluation, described below.

health care. The analysis is further subsetted to consider quantiles of the explanatory variables, denoted by $q \in [0,1]$.[4] The impact of the policy within a data quantile can be estimated, for each health facility choice, as the average difference between the predicted probability of using facility j above and below the age threshold for that data quantile, where $\phi(a)$ is the function of age included in the empirical model.

$$\widehat{\tau}_{jq} = n^{-1} \sum_{i=1}^{n} [\widehat{p}_{ij}(X, \phi(a)|X_q, a < a_0) - \widehat{p}_{ij}(X, \phi(a)|X_q, a > a_0)] \tag{14}$$

Given $\sum_j \widehat{p}_{ij} = 1$, by construction, $\sum_j \widehat{\tau}_{jq} = 0 \ \forall q$. In the initial analysis, ϕ is a simple binary indicator of eligibility for free health care, such that $\phi(a) = \mathbf{1}[a < a_0]$. In the follow-up analysis, $\phi(a)$ is more general. For the multinomial logit, $\phi(a)$ is a linear function of age, while the exact functional form of $\phi(a)$ is not specified *a priori* in the nonparametric model.

In order for (14) to represent the true policy impact for choice j in quantile q, the predicted probabilities must be consistently estimated on either side of the age threshold, although the difference would be consistently estimated, as long as any inconsistency in the predicted probabilities across the age threshold were constant, allowing for the inconsistencies to be differenced out. Potential worries include, for example, the possibility that children just slightly above the age of six could be passed off at the public facility as being under the age of six, although this would be a mistake at the facility, rather than something that a caregiver could guarantee, and, thus, is not likely to be a serious problem. Similarly, it is possible that children under six could be more likely to be reported as ill in the data, since they, in fact, could be received at a public facility for free. Such a selection problem, would tend to overstate the policy impact. Unfortunately, there are no useful exclusion restrictions available in the data that could be used to identify either anticipation effects or selection effects.[5] Due to the lack of instruments, the policy impacts estimated, below, must be understood in that context. Although that is an obvious caveat, it should be noted that previous research related to

[4] Ordering the binary explanatory variables allows the values of these indicators to switch at some quantile within the distribution.

[5] Data from 1993, before user fees were abolished, is available. However, that data does not allow for the separation of public care from private care, and, therefore, it is not possible to provide a pre-policy falsification test.

provider choice in South Africa suggests that private health care is used even amongst the poor, Burger & Grobler (2007) and Grobler & Stuart (2007).

Finally, confidence intervals for the average policy impact within a data quantile are calculated via bootstrap methods. Samples of the data are drawn, with replacement; the average treatment effects within a quantile are calculated, and the process is repeated, yielding a series of estimates of the policy impact within a quantile. From that series, the 5^{th} and 95^{th} percentile are drawn. These values constitute the lower and upper bound, respectively, of the estimated average. **Really could use a citation or two for the approach that I am following...**

2.5 Model comparison

The preceding discussion outlined two different estimation models, the multinomial logit and the nonparametric conditional probability model, which are not nested. In order to compare the two models, we consider out-of-sample performance, borrowing terminology from discriminant analysis. Rather than assuming that one of the models is the true model, we assume that both models are approximations, and, thus, we are interested in the model with the lowest expected true error. Efron (1982) outlines apparent versus true error estimation in greater detail than will be described here. Intuitively, apparent error is derived from in-sample measures of fit, such as R^2 in linear regression, while true error is derived out-of-sample from attempts to fit the model to new data drawn from the underlying data generating process. We apply this intuition, through the examination of the Correct Classification Ratio (CCR), Efron (1978), applied to multinomial outcomes.

The observed outcomes $Y_i = k$ are mapped to a $j \times 1$ vector Υ_i.

$$\Upsilon_{ik} = \begin{cases} 1 & \text{if } Y_i = k \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The predictions from the empirical model define a similar prediction vector, which is

based on the predicted probabilities from the model. $\hat{\Upsilon}_i$.

$$\hat{\Upsilon}_{ik} = \begin{cases} 1 & \text{if } \widehat{p}_{ik} = \max_j\{\widehat{p}_{ij}\} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Given these predictions, it is straightforward to define a loss function penalizing incorrect predictions.

$$Q_i(\Upsilon, \hat{\Upsilon}, n) = \begin{cases} 0 & \text{if } \Upsilon_i = \hat{\Upsilon}_i \\ 1 & \text{otherwise} \end{cases} \tag{17}$$

The loss function can then be used to define the correct classification ratio (CCR).

$$CCR = 1 - n^{-1} \sum_{i=1}^{n} Q_i(\Upsilon, \hat{\Upsilon}, n) \tag{18}$$

In addition to the loss function and CCR, the underlying confusion matrix (CM) provides useful information regarding a model's ability to properly predict one set of outcomes relative to another set. The CM presents the relative counts of actual outcomes against predicted outcomes.

$$CM = \Upsilon' \hat{\Upsilon} \tag{19}$$

Based on insights from Efron (1982), Racine & Parmeter (forthcoming) suggest a revealed performance test related to the CCR and its associated loss function. The sample moment in (18) is an in-sample estimate of the expected loss, or true error, as it uses all of the observations from the original sample. Instead of using the full sample, define an *iid* training sample, $Z^{n_1} = \{Y_i, X_i\}_{i=1}^{n_1}$, distributed with cumulative distribution function (CDF) $\widehat{F}$. The training sample would yield the true error $E_{n_1,\widehat{F}}[Q(\Upsilon, \hat{\Upsilon}, n_1)]$; see Efron (1982). In addition to the training sample, consider an *iid* evaluation sample, $Z^{n_2} = \{Y_i, X_i\}_{i=n_1+1}^{n}$, that is also independent of the training sample. The evaluation sample, is assumed to be distributed with CDF F, and yields true error $E_{n_2,F}[Q(\Upsilon, \hat{\Upsilon}, n_2)]$. The expected true error is the expectation of the true error, $E\{E_{n_2,F}[Q(\Upsilon, \hat{\Upsilon}, n_2)]\}$.

The preceding discussion hints at a bootstrap procedure used to estimate model

performance along the lines of Racine & Parmeter (forthcoming).

1. Resample the original data $Z = \{X, Y\}$, without replacement. Refer to this new data as Z_*.

2. Define $Z_*^{n_1}$ and $Z_*^{n_2}$ as above, although all duplicate observations are removed from $Z_*^{n_2}$ to maintain *iid* assumption.

3. Use estimates from the full sample, i.e., hold smoothing, in the case of the nonparametric model, and parameters, in the case of the multinomial logit model, fixed. Fit each model on $Z_*^{n_1}$, and then obtain predicted values for $Z_*^{n_2}$.

4. Compute CCR for each model.

5. Repeat T times; in our example, $T = 10000$, which results in T draws of CCR for both models.

The draws from the bootstrap procedure are used to illustrate and compare the underlying empirical distribution functions of expected true error in the multinomial logit and nonparametric models.

3 Analysis Data

3.1 Data Source

Data for the analysis was sourced from the South African October Household Survey (OHS) of 1995.[6] The main purpose of the OHS, Statistics South Africa (1995), was to collect information on households and individuals across the nine provinces of South Africa, and the survey included questions related to dwellings and dwelling services, perceived quality of life, socio-demographic information, employment and unemployment, the informal and formal labour markets, as well as births and deaths in the household. Along with this information, there is a short series of questions related to illness, injury, healthcare-seeking behavior and access to medical aid or health insurance.

[6] Although data from both 1994 and 1996 is also available, 1994 was deemed too soon after policy implementation, while 1996 was assumed to be confounded by a further change in the policy that extended free primary care within the public sector to all.

The survey included responses for 121 538 individuals, living in 29 700 households; however responses are only available from one respondent in each household. The survey follows a stratified random sampling method, being explicitly stratified by province, magisterial district, urban or rural locale and population group. These enumeration areas were selected systematically based on probabilities proportional to their size, where the size was estimated from the 1991 population census. Within a selected enumeration area, ten households were drawn for interview. Post-stratified weights are available, but are not used in the analysis, due to the fact that the analysis sample is limited to all children aged 14 and under that have been reported ill or injured in the last 30 days. The weights, though, are not calibrated for a subsample of this nature, and, therefore, the weights are not used in the analysis.

A series of different sections in the survey cover a variety of different topics; however, it is possible to merge the relevant information to create data at the child level. For this analysis, data for each child was taken from the individual questionnaire, including information on the mother and father, if they are members of the household. Information related to the child's mother and/or father was merged into the child dataset, as was data related to the household.[7]

For the analysis, only healthcare-seeking behavior is considered. Therefore, we consider the multinomial outcome: (1) whether care for the ill or injured child was sought in a public facility, (2) private facility or (3) not at all. In addition to the outcome variable, we create dummy variables for the sex, the population group and the age of the child. Age serves as the running variable in the analysis, which is availabe for all children in the survey in years, but only available in days, weeks or months, for children matched to their mothers, given the structure of the survey.[8]

Household level controls include the size of the household, as well as categorical variables for urban and provincial locales, and the distance (measured in time) the dwelling sits from the health facility usually attended, if a household member seeks medical care. For mothers and fathers, we capture their education, whether or not they

[7] It is important to note that children, who could not be matched to mothers, are not included in this data.

[8] Probably need to think a bit more about dealing with this differentiation in matchability of the children.

have health insurance, and whether or not either the parent is alive.

3.2 Data Descriptives

The variables included in the analysis are described in Tables 1 and 2.[9] Table 1 describes the data for all ill or injured children under the age of 6, while Table 2 does the same for children over the age of 6. As expected, the mean age and the mean value of the regression discontinuity indicator are higher in Table 2 than in Table 2. Otherwise, at the level of the mean, which is presented in these tables, there are relatively minor differences between the two subsets. Given the difference in the results that will be discussed below, it is most likely necessary to try to present a wider illustration, and maybe pull out distribution issues in the underlying data?

Table 1: Descriptive Statistics below the Threshold

	n	mean	sd	se
hlthsk*	1522	2.097	0.674	0.017
age	1522	2.305	1.668	0.043
insure*	1522	1.256	0.436	0.011
inc	1522	1476.396	2448.769	62.768
inc.sq	1522	8172273.122	27942246.453	716232.449
pop.group*	1522	1.621	1.061	0.027
prov*	1522	4.543	2.228	0.057
urban*	1522	1.591	0.492	0.013
hh.size*	1522	4.258	1.924	0.049
time.med*	1522	2.957	1.043	0.027
rd.age*	1522	1.000	0.000	0.000
mom.ed*	1522	2.608	1.067	0.027
mom.live*	1522	1.983	0.130	0.003
mom.ins*	1522	1.229	0.421	0.011
dad.ed*	1522	1.940	1.163	0.030
dad.live*	1522	1.911	0.285	0.007
dad.ins*	1522	1.204	0.403	0.010

4 Empirical Model Comparison

Before examining the impacts of policy, the empirical fit of the multinomial logit and nonparamatric models are examined. The comparison begins on a subset of the analysis

[9] At this stage, only simple means are described, but it would be rather easy to include a bit more of a distribution discussion.

Table 2: Descriptive Statistics above the Threshold

	n	mean	sd	se
hlthsk*	1194	2.028	0.747	0.022
age	1194	8.422	1.750	0.051
insure*	1194	1.283	0.451	0.013
inc	1194	1515.894	2631.678	76.161
inc.sq	1194	9217865.572	30930784.099	895135.469
pop.group*	1194	1.718	1.107	0.032
prov*	1194	4.519	2.275	0.066
urban*	1194	1.595	0.491	0.014
hh.size*	1194	4.249	1.821	0.053
time.med*	1194	2.994	1.032	0.030
rd.age*	1194	2.000	0.000	0.000
mom.ed*	1194	2.396	1.105	0.032
mom.live*	1194	1.967	0.178	0.005
mom.ins*	1194	1.266	0.442	0.013
dad.ed*	1194	1.958	1.143	0.033
dad.live*	1194	1.885	0.319	0.009
dad.ins*	1194	1.237	0.425	0.012

variables, in which both models perform identically. However, when the analysis is extended to include additional variables, however, model performance differs starkly.

4.1 A baseline with similar predictive performance

The variables included in the initial analysis are limited to controls for earnings in the household (income), a binary indicator of access to health insurance (insure) and a binary indicator of policy eligibility (rd.age6+). As the parameter estimates are not the most important component of the analysis, the multinomial logit estimates have been relegated to Appendix Table A.1, while the nonparametric bandwidths and scale factors have been relegated to Appendix Table A.2. Although the estimates are not the primary focus at this stage, the results suggest that the control variables are statistically significant determinants of health facility choice.

Rather than focusing on parameter estimations, the empirical results and the data were used to calculate in-sample performance, which is presented within two tables, Table 3 and Table 4. One of the striking results within the two tables is the inability of either multinomial model to predict nontreatment outcomes. Another striking result, and the primary reason for choosing this set of explanatory variables, is that the in-

sample predictive performance for both the multinomial logit - see Table 3 - and the nonparametric model - see Table 4 - yield similar in-sample performance. In fact, the multinomial logit model performs better than the nonparametric model, in this limited setting. For the full sample of data, the nonparametric overall CCR is 0.5541, while the parametric overall CCR is 0.553.

Table 3: Multinomial Logit Confusion Matrix for Model 1

	notreat	public	private
notreat	0	472	124
public	0	1136	208
private	0	410	366

Table 4: Nonparametric Confusion Matrix for Model 1

	notreat	public	private
notreat	0	466	130
public	0	1126	218
private	0	397	379

Although a larger CCR is indicative of better predictive power, it is important to note that the preceding CCRs are all in-sample. One concern that arises when using nonparametric models is the potential for overfitting. Even though the bandwidth selection process is theoretically optimal, bandwidth selection is still a data-driven process. In other words, the similarity in results could be specific to the data. Hence, we conduct an out-of-sample performance evaluation exercise, as described in Section 2.5. Importantly, duplicate observations are removed from the data, before evaluation, and, therefore, the evaluation data set is not the same as the initial data set, over which the bandwidth selection was made. If the nonparametric model has placed too much weight on duplicate observations, the aforementioned correction will uncover the problem. The out-of-sample fit comparison is illustrated in Figure 1.

Despite the potential for overfitting in the nonparametric model, the out-of-sample fits illustrated in Figure 1 do not provide evidence of it. As was suggested by the in-sample comparison, there is very little difference in performance between the multinomial and nonparametric models, given the limited subset of control variables included in the initial model. The out-of-sample performance comparison is rather similar to the in-

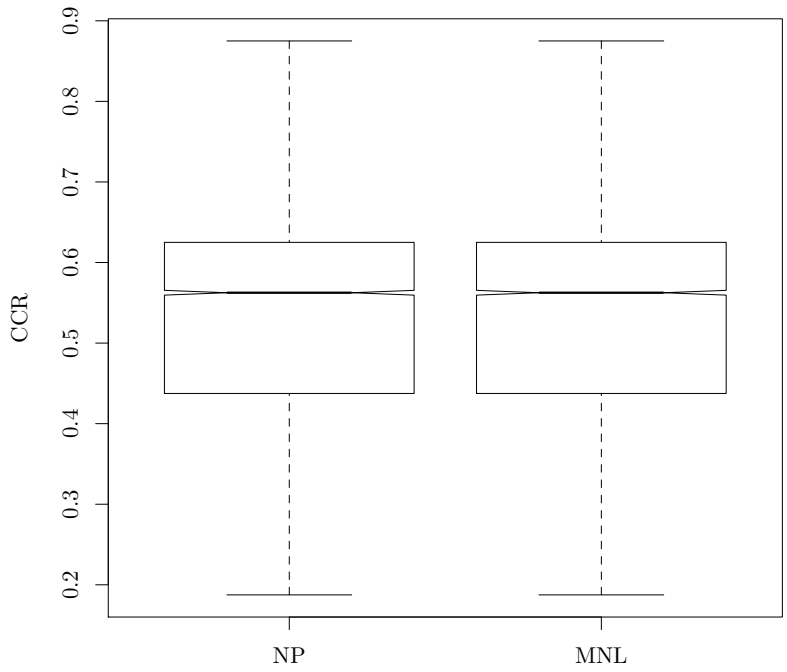

Figure 1: Boxplots for Model 1 (mean nonparametric CCR: 0.5504, mean parametric CCR: 0.5492, median nonparametric CCR: 0.5625, median parametric CCR: 0.5625, 10^4 splits of the data, training data size $n_1 = 2700$, evaluation data size $n_2 = 16$, higher CCR is better).

sample performance comparison.

4.2 Beyond the baseline: no longer similar

With just a few explanatory variables, model performance does not differ, and, more poignantly, neither model manages to predict outcomes, especially the no-healthcare-seeking outcome, particularly well. However, the initial model did not include many of the determinants of healthcare facility choice previously identified in the literature. Therefore, in what follows, additional explanatory variables are included, such as: population group, region, household size, distance to nearest medical facility, parental education (if available), parental health insurance status, and indicators for whether or not either parent is alive.

As before, the focus of the analysis is not on the estimates, and, therefore, the

multinomial logit estimates are presented in Table A.3, while nonparametric bandwidths and scale factors are presented in Table A.4. As before, the child's access to a medical aid and eligibility for free public health care, remain significant determinants in the parametric model. In addition to those variables, there are significant differences across population groups and regions, as well as household size and parental effects. Given the fact that many of the included variables are statistically significant, one would expect the predictive performance of the multinomial logit model to improve.

In-sample, however, that does not appear to be the case. With only a few explanatory variables, approximately 56% of the outcomes were predicted correctly, in-sample. Including additional explanatory variables increased the parametric model's in-sample predictive performance to only 0.5796, a rather small improvement, even though many of the included variables are statistically significant. Meanwhile, the nonparametric model is predicting much better; the nonparametric overall CCR is 0.9111. As can be seen in the confusion matrixes - see Tables 5 and 6 - the multinomial logit model still has very limited succes in predicting nontreatment, probably due to the fact that we do not have any information on the depth of the child's illness. Depite not having that information, the nonparametric model appears to be more successful, possibly because some of the variables included in the model are correlated with the unoberved depth of child illness.

Table 5: Multinomial Logit Confusion Matrix for Model 2

	notreat	public	private
notreat	59	419	115
public	37	1134	166
private	23	375	372

Table 6: Nonparametric Confusion Matrix for Model 2

	notreat	public	private
notreat	458	120	15
public	9	1320	8
private	6	82	682

Although in-sample performance is strongly in favour of the nonparametric model, it is possible that the nonparametric model is overfitting the data, and, therefore, the same performance comparison, see above, is undertaken here, as well. The results of

that training exercise are illustrated in Figure 2. As can be seen, and in agreement with the in-sample performance, the multinomial logit model's predictive performance leaves much to be desired, relative to the nonparametric model. In other words, even though many of the explanatory variables in the multinomial logit model are statistically significant, they do not appear to provide any explanatory power, at least in this analysis. Furthermore, there is strong evidence that the nonparametric model was overfitting the data, in-sample; however, nonparametric out-of-sample performance is still much better than either in-sample or out-of-sample paramatric model performance.

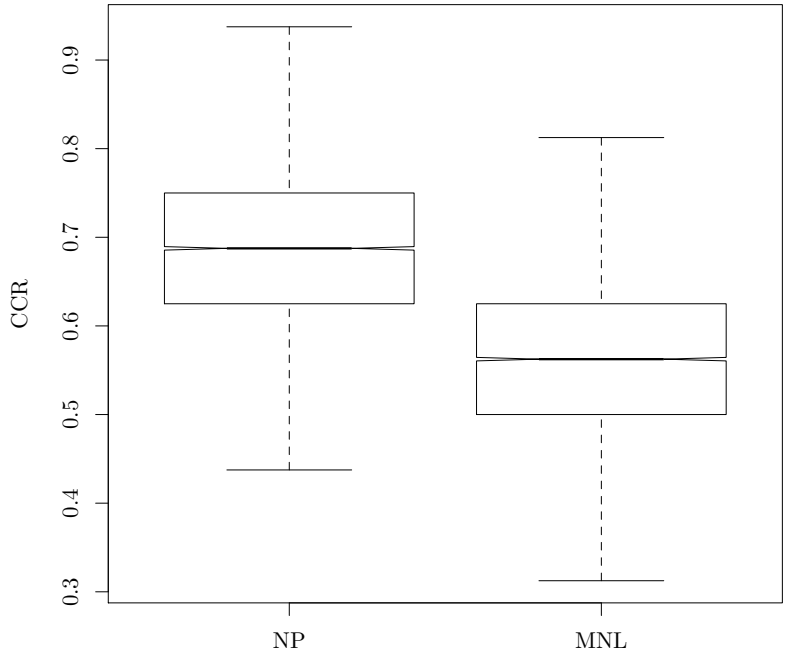

Figure 2: Boxplots for Model 2 (mean nonparametric CCR: 0.6988, mean parametric CCR: 0.5633, median nonparametric CCR:0.6875, median parametric CCR: 0.5625, 10^4 splits of the data, training data size n_1 = 2684, evaluation data size n_2 = 16, higher CCR is better).

5 Evaluation of policy

Now that we have seen the relative performance improvement associated with the nonparametric model over the parametric model, we turn to the evaluation of the policy.

The results of the analysis are presented in two Figures, Figure 3 and Figure 6. Since the user fees were eliminated in the public sector, that is the focus of the figures.

5.1 Discrete running variable

In the initial analysis, age is demarcated in years, the data is pooled, and a regression discontinuity indicator is included. In other words, the policy is assumed to have a constant impact, once the rest of the relationships in the data have been conditioned-out. Importantly, that does not require the policy impact to be constant across the entire sample, as the conditioning effects, especially in the nonparametric model, are not necessarily constant.

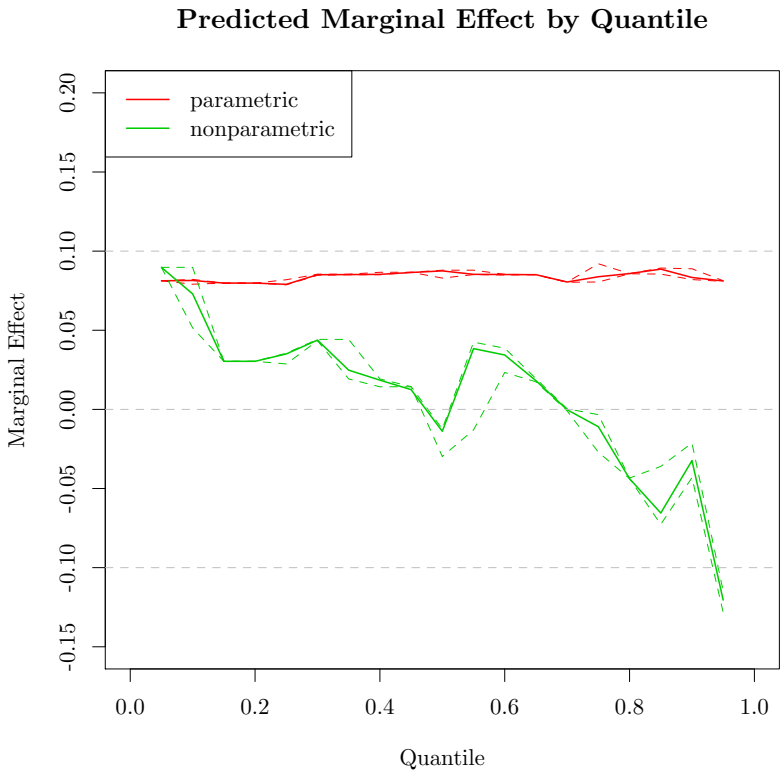

Figure 3: Public Care Marginal Effects Across Data Quantiles with Discrete Running Variable

In Figure 3, the paramatric and nonparametric marginal effects, across data quantiles, along with 95% confidence bands, are illustrated. The parametric marginal effects are fairly constant across the data, and in the range of 7%. Disappointingly, children in

households in the top of the distribution receive the same benefit, as children in households in the bottom of the distribution. The nonparametric results, on the other hand, paint a more preferable picture, at least with respect to equity considerations, even if those results are noisier. According to the bootstrapped nonparametric results, user fee abolition increased the use of public healthcare facilities amongst young children by up to about 7% at the bottom of the distribution, although that effect is entirely eliminated for ill children in the top 70^{th} percentile of the distribution.

5.2 Substitution

The primary reason for considering the outcome data in its entirety, i.e., as an unordered categorical outcome variable, is substitution across healthcare facility choices. Estimated marginal effects for all three healthcare facility options, including no treatment, are illustrated separately for the parametric model, in Figure 4, and the nonparametric model, in Figure 5.

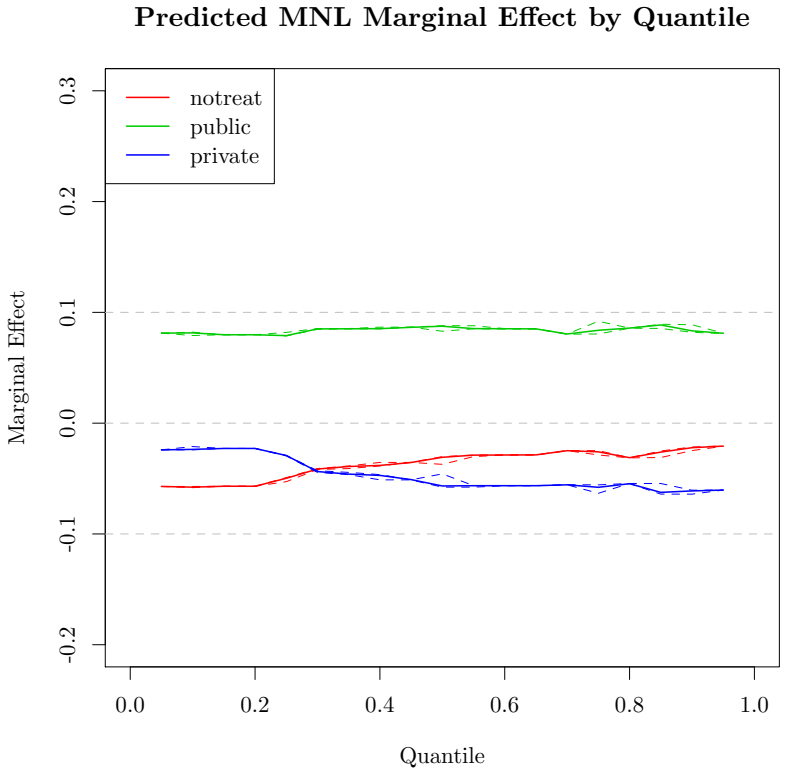

Figure 4: Multinomial Logit Marginal Effects for all Outcomes with Discrete Running Variable

Given the constancy in the public care marginal effects estimated for the multinomial logit model, it is not surprising that the private care and nontreatment marginal effects are not too varied over the distribution. Similarly, it is not surprising that, at the bottom of the distribution, the increase in public care use is driven by a decrease in the proportion of nontreated ill children. On the other hand, at the top of the distribution, the substitution is relatively more intense for private treatment.

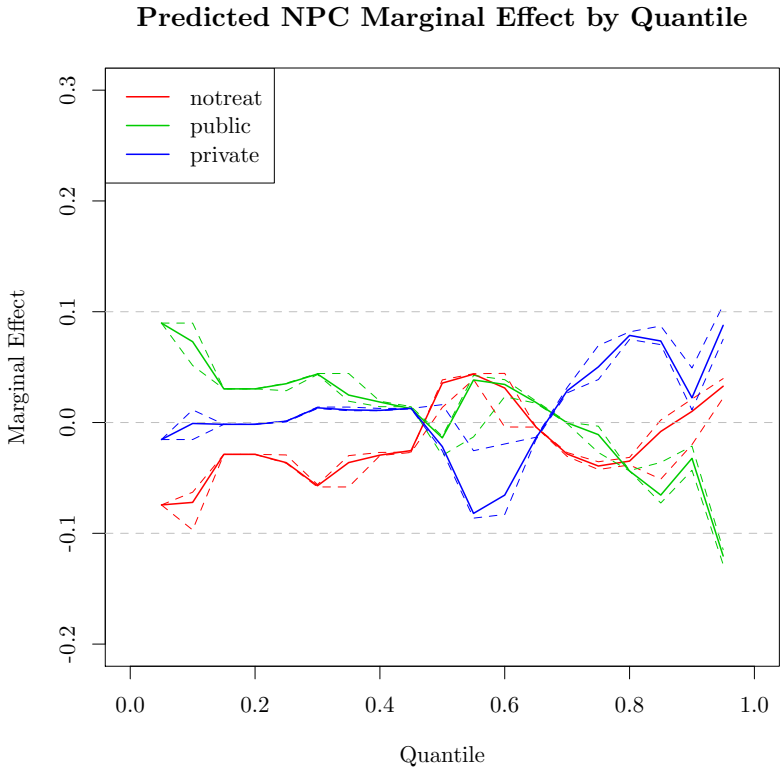

Figure 5: Nonparametric Marginal Effects for all Outcomes with Discrete Running Variable

Although there is some evidence of substitution patterns, within the parametric setting, that are expected, they are not particularly pronounced or varied. Within the nonparametric setting, the substition patterns are more interesting, and more believable. At the bottom end of the distribution, user fees increased the use of public facilities, simultaneously decreasing the proportion of ill children that do not receive healthcare; virtually no effect is uncovered in the private sector. Within the middle of the distribution, on the other hand, the etimated marginal effecs are rather different. A reduced

proportion of ill children receive public healthcare, near the middle of the distribution, while both public healthcare use and non-use are increased.

5.3 Continuous running variable

At this stage, I am not going to say too much here. I am not sure I know what to think about the big drop and the big increase around the 40% range. As noted before, it is necessary to describe the data *distribution* a bit better. I realize that I have not really described how any of it was put together for the illustration. Also, I should note that the regressions are done separately on either side of the threshold, and that is probably not the best idea.

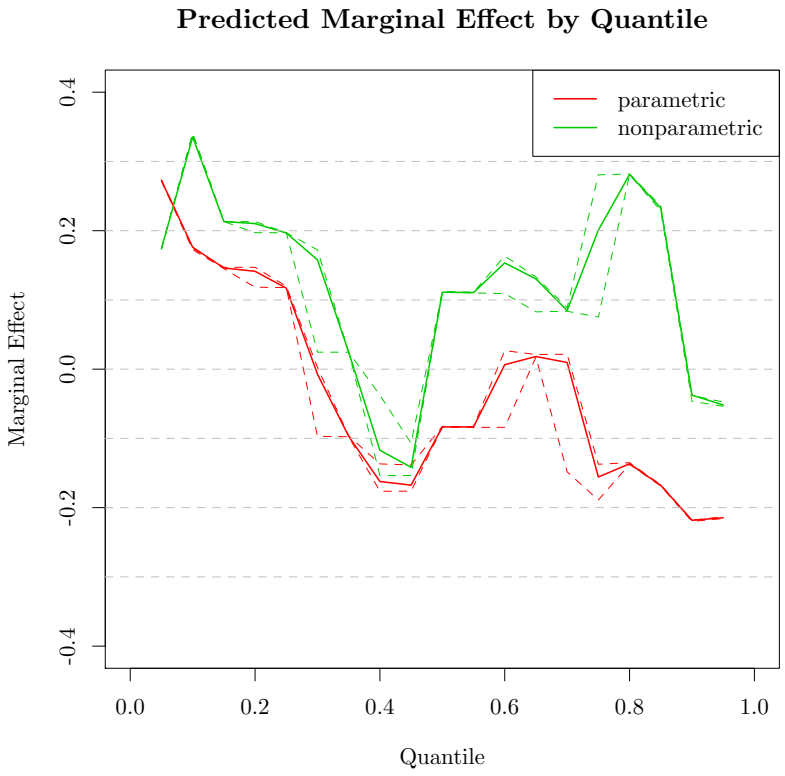

Figure 6: Public Care Marginal Effects Across Data Quantiles with Continuous Running Variable

6 Conclusion

This research focuses on the effect of user fee abolition on healthcare facility choice. The analysis focuses on young children, because that was the focus of the policy. Parametric and nonparametric models of multinomial outcome models were estimated. The results suggest that parametric models poorly predict the outcomes, relative to the nonparametric models. In addition to the difference in model performance, the estimated marginal effects are rather different across the two models.

Within the nonparametric setting, marginal effects are estimated to be smaller than they are in the parametric setting. Although multinomial marginal effects are in the neighborhood of 7%, and are so across the entire distribution, the nonparametric marginal effects are generally smaller and disappear entirely, once the 70^{th} percentile of the data is reached.

References

Albouy, V. & Lequien, L. (2009), 'Does compulsory education lower mortality', *Journal of Health Economics* **28**, 155–168.

Angrist, J. D. & Lavy, V. (1999), 'Using Maimonides' Rule to estimate the effect of class size on scholastic achievement', *Quarterly Journal of Economics* pp. 533–575.

Berk, R. A. & Rauma, D. (1983), 'Capitalizing on nonrandom assigment to treatments: A regression discontinuity evaluation of a crime-control program', *Journal of the American Statistical Association* **78**(381), 21–27.

Burger, R. & Grobler, C. (2007), 'Have pro-poor health policies improved the targeting of spending and effective delivery of health care in South Africa', Development Policy Research Unit Working Paper 07/122.

Calcagno, J. C. & Long, B. T. (2008), 'The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance', NBER Working Paper no. 14194.

Card, D., Dobkin, C. & Maestas, N. (2008), 'The impact of nearly universal insurance coverage on health care: evidence from Medicare', *American Economic Review* **98**(5), 2242–2258.

Carpenter, C. & Dobkin, C. (2009), 'The effect of alcohol consumption on mortality: Regression discontinuity evidence from the minimum drinking age', *American Economic Journal: Applied Economics* **1**(1), 164–182.

Coe, N. B. & Zamarro, G. (2011), 'Retirement effects on health in Europe', *Journal of Health Economics* **30**, 77.

Efron, B. (1978), 'Regression and anove with zero-one data: Measures of residual variation', *Journal of the American Statistical Association* **73**, 113–121.

Efron, B. (1982), *The Jacknife, the Bootstrap and Other Resampling Plans*, Society for Industrial Mathematics.

Goldberger, A. S. (1972*a*), 'Selection bias in evaluating treatment effects: Some formal illustrations', Discussion Paper No. 123, University of Wisconsin, Madison, Institute for Research on Poverty.

Goldberger, A. S. (1972*b*), 'Selection bias in evaluation treatment effects: The case of interaction', Discussion Paper No. 129, University of Wisconsin, Madison, Institute for Research on Poverty.

Grobler, C. & Stuart, I. C. (2007), 'Health care provider choice', *South African Journal of Economics* **75**(2), 327–350.

Hall, P., Racine, J. S. & Li, Q. (2004), 'Cross-validation and the estimation of conditional probability densities', *Journal of the American Statistical Association* **468**, 1015–1026.

Hayfield, T. & Racine, J. S. (2008), 'Nonparametric econometrics: The np package', *Journal of Statistical Software* **27**(5), 1–32.

Imbens, G. W. & Lemieux, T. (2008), 'Regression discontinuity designs: a guide to practice', *Journal of Econometrics* **142**(2), 615–635.

Lalive, R. (2008), 'How do extended benefits affect unemployment duration? A regression discontinuity approach', *Journal of Econometrics* **142**, 785–806.

Lee, D. S. & Card, D. (2008), 'Regression discontinuity inference with specification error', *Journal of Econometrics* **142**, 655–674.

Lee, D. S. & Lemieux, T. (2010), 'Regression discontinuity designs in Economics', *Journal of Economic Literature* **48**(2), 271–355.

Lemieux, T. & Milligan, K. (2008), 'Incentive effects of social assistance: A regression discontinuity approach', *Journal of Econometrics* **142**, 807–828.

Li, Q. & Racine, J. S. (2003), 'Nonparametric estimation of distributions with categorial and continuous data', *Journal of Multivariate Analysis* **86**, 266.

Li, Q. & Racine, J. S. (2007), *Nonparametric Econometrics*, Princetone University Press, Princeton.

Lindeboom, M., Llena-Nozal, A. & van der Klaauw, B. (2009), 'Parental education and child health: Evidence from a schooling reform', *Journal of Health Economics* **28**, 109–131.

McCrary, J. & Royer, H. (2011), 'The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth', *American Economic Review* **101**(1), 158–195.

Ou, D. (2010), 'To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam', *Economics of Education Review* **29**, 171–186.

Palangkaraya, A. & Yong, J. (2007), 'How effective is "lifetime health cover" in raising private health insurance coverage in Australia? An assessment using regression discontinuity', *Applied Economics* **39**(11), 1361–1374.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Racine, J. S. & Parmeter, C. F. (forthcoming), Data-driven model evaluation: A test for revealed performance, *in* A. Ullah, J. S. Racine & L. Su, eds, 'Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics', Oxford University Press.

Rubin, D. B. (1977), 'Assignment to treatment group on the basis of a covariate', *Journal of Educational Statistics* **2**, 1–26.

Silles, M. A. (2009), 'The causal effect of education on health: Evidence from the United Kingdom', *Economics of Education Review* **28**, 122–128.

Small, K. & Hsiao, C. (1985), 'Multinomial logit specification tests', *International Economic Review* **26**(619-627).

Statistics South Africa (1995), *October Household Survey (South Africa), 1995*, Statistics South Africa (producer) and South African Data Archive (distributor), Pretoria.

Thistlethwaite, D. & Campbell, D. (1960), 'Regression-discontinuity analysis: An alternative to the *ex post facto* experiment', *Journal of Educational Psychology* **51**, 309–317.

van der Klaauw, W. (2008), 'Regression-discontinuity analysis A survey of recent developments in economics', *Labour* **22**(2), 219–245.

Venables, W. N. & Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.
URL: *http://www.stats.ox.ac.uk/pub/MASS4*

Wang, M. C. & van Ryzin, J. (1981), 'A class of smooth estimators for discrete distributions', *Biometrika* **68**, 301–309.

Zuckerman, I. H., Lee, E., Wutoh, A. K., Xue, Z. & Stuart, B. (2006), 'Application of regression-discontinuity analysis in pharmaceutical health services research', *Health Services Research* **41**(2), 550–563.

A Discrete Running Variable

A.1 Model 1 Estimation Results

In this subsection, model one estimates and bandwidth summaries are presented.

A.2 Model 2 Estimation Results

In this subsection, model two estimates and bandwidth summaries are presented.

B Continuous Running Variable

B.1 Estimation Results

Currently, these are estimated separately on either side of the regression, and I am not absolutely sure that is the best idea.

B.2 Continuous Running Variable Marginal Effects Illustrations

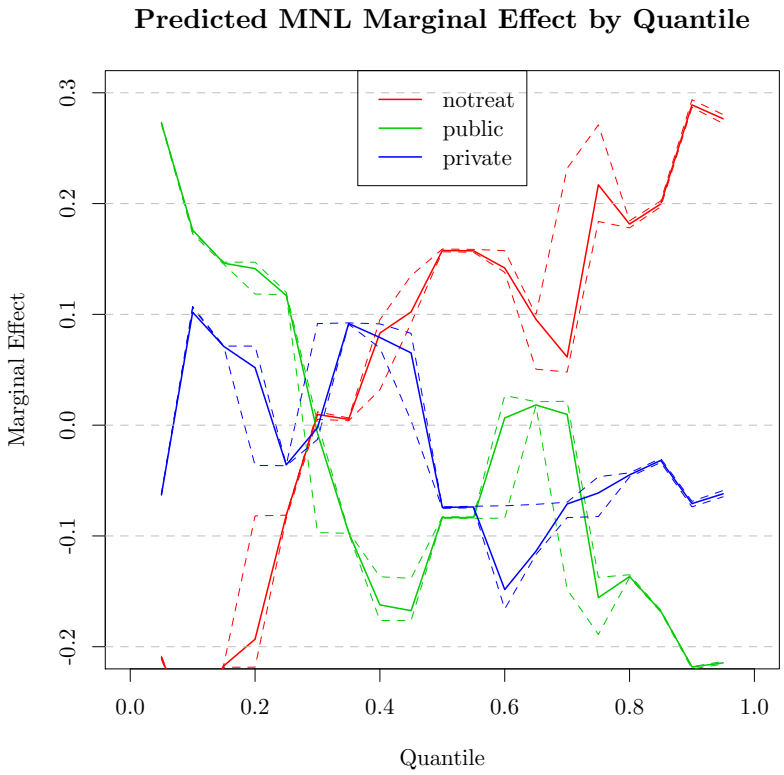

Figure B.1: Multinomial Logit Marginal Effects for all Outcomes with Continuous Running Variable

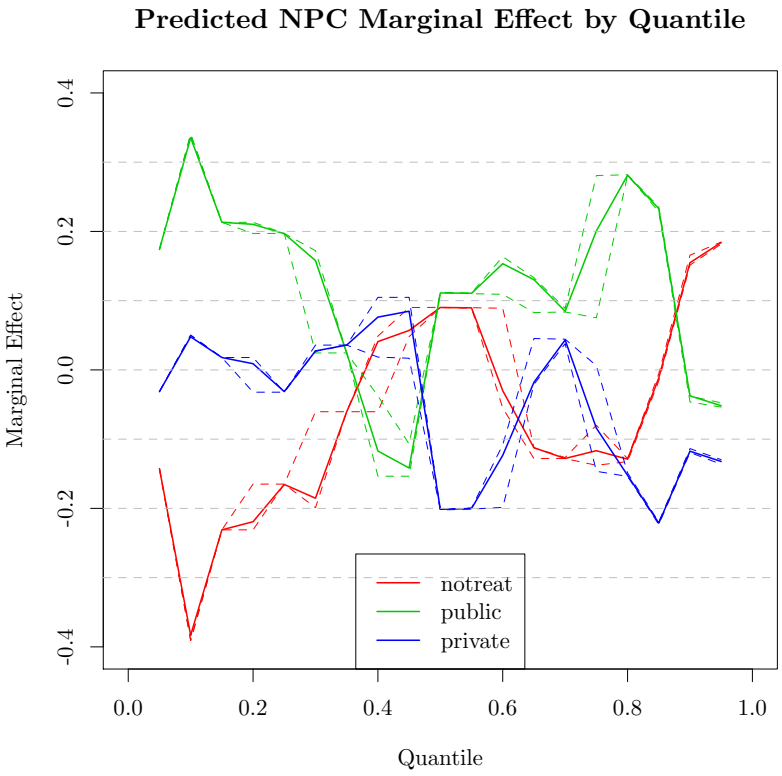

Figure B.2: Nonparametric Marginal Effects for all Outomes with Continuous Running Variable

Table A.1: MNL Parameter Summary for Model 1

variable	pub:coef	pub:se	prv:coef	prv:se
(Intercept)	1.0979	0.0271	0.2188	0.0261
rd.age6+	-0.8958	0.0058	-0.9103	0.0061
age	-0.1083	0.0324	-0.1949	0.0360
insure1	0.6111	0.0270	1.2078	0.0253
inc	-0.0000	0.0000	0.0000	0.0000
pop.groupcolour	-0.1800	0.0187	-0.0008	0.0184
pop.groupasian	-0.3364	0.0097	0.7048	0.0083
pop.groupwhite	-0.3694	0.0164	0.3567	0.0148
prov2	-0.0750	0.0563	-0.4049	0.0451
prov3	0.2294	0.0071	-0.4748	0.0065
prov4	-0.8126	0.0139	-0.5394	0.0111
prov5	0.2479	0.0495	-0.0289	0.0400
prov6	0.6921	0.0238	0.4558	0.0193
prov7	0.0873	0.0225	-0.0077	0.0181
prov8	-0.2861	0.0142	-0.2185	0.0109
prov9	-0.1759	0.0042	-0.0879	0.0033
urban.L	0.0418	0.0581	0.1227	0.0592
hh.size.L	0.4867	0.0194	0.8819	0.0168
hh.size.Q	-0.0479	0.0283	-0.0191	0.0230
hh.size.C	-0.1716	0.0394	-0.3026	0.0318
hh.size^4	-0.2104	0.0517	-0.0131	0.0408
hh.size^5	-0.1031	0.0415	-0.0528	0.0324
hh.size^6	-0.0762	0.0500	0.1240	0.0391
time.med.L	0.1853	0.0340	0.6051	0.0375
time.med.Q	-0.1219	0.0499	-0.0849	0.0398
time.med.C	0.0646	0.0559	-0.2018	0.0447
mom.ed.L	0.0999	0.0537	0.1766	0.0444
mom.ed.Q	-0.2458	0.0504	0.0880	0.0415
mom.ed.C	-0.0308	0.0535	0.1529	0.0440
mom.live.L	-0.1911	0.0219	-0.1543	0.0212
mom.ins.L	-0.4523	0.0359	-0.1017	0.0377
dad.ed.L	0.2128	0.0391	0.3017	0.0338
dad.ed.Q	0.0288	0.0397	0.1530	0.0310
dad.ed.C	0.1165	0.0612	0.0788	0.0485
dad.live.L	-0.1894	0.0415	-0.0602	0.0374
dad.ins.L	-0.4766	0.0337	-0.5587	0.0355
rd.age6+:age	0.1162	0.0278	0.2053	0.0314

Table A.2: Nonparametric Bandwidth Summary for Model 1

Variable	Bandwidth	Scale Factor
hlthsk	0.07	1.00
rd.age	0.59	8.18
age	53265988.06	57085641.52
insure	0.07	1.04
inc	1119.85	2.97
pop.group	0.16	2.27
prov	0.07	0.95
urban	0.94	13.02
hh.size	0.34	4.77
time.med	0.28	3.96
mom.ed	0.52	7.25
mom.live	0.08	1.13
mom.ins	0.80	11.15
dad.ed	0.69	9.63
dad.live	0.38	5.30
dad.ins	1.00	13.92

Table A.3: MNL Parameter Summary for Model 2

variable	pub:coef	pub:se	prv:coef	prv:se
(Intercept)	1.0979	0.0271	0.2188	0.0261
rd.age6+	-0.8958	0.0058	-0.9103	0.0061
age	-0.1083	0.0324	-0.1949	0.0360
insure1	0.6111	0.0270	1.2078	0.0253
inc	-0.0000	0.0000	0.0000	0.0000
pop.groupcolour	-0.1800	0.0187	-0.0008	0.0184
pop.groupasian	-0.3364	0.0097	0.7048	0.0083
pop.groupwhite	-0.3694	0.0164	0.3567	0.0148
prov2	-0.0750	0.0563	-0.4049	0.0451
prov3	0.2294	0.0071	-0.4748	0.0065
prov4	-0.8126	0.0139	-0.5394	0.0111
prov5	0.2479	0.0495	-0.0289	0.0400
prov6	0.6921	0.0238	0.4558	0.0193
prov7	0.0873	0.0225	-0.0077	0.0181
prov8	-0.2861	0.0142	-0.2185	0.0109
prov9	-0.1759	0.0042	-0.0879	0.0033
urban.L	0.0418	0.0581	0.1227	0.0592
hh.size.L	0.4867	0.0194	0.8819	0.0168
hh.size.Q	-0.0479	0.0283	-0.0191	0.0230
hh.size.C	-0.1716	0.0394	-0.3026	0.0318
hh.size^4	-0.2104	0.0517	-0.0131	0.0408
hh.size^5	-0.1031	0.0415	-0.0528	0.0324
hh.size^6	-0.0762	0.0500	0.1240	0.0391
time.med.L	0.1853	0.0340	0.6051	0.0375
time.med.Q	-0.1219	0.0499	-0.0849	0.0398
time.med.C	0.0646	0.0559	-0.2018	0.0447
mom.ed.L	0.0999	0.0537	0.1766	0.0444
mom.ed.Q	-0.2458	0.0504	0.0880	0.0415
mom.ed.C	-0.0308	0.0535	0.1529	0.0440
mom.live.L	-0.1911	0.0219	-0.1543	0.0212
mom.ins.L	-0.4523	0.0359	-0.1017	0.0377
dad.ed.L	0.2128	0.0391	0.3017	0.0338
dad.ed.Q	0.0288	0.0397	0.1530	0.0310
dad.ed.C	0.1165	0.0612	0.0788	0.0485
dad.live.L	-0.1894	0.0415	-0.0602	0.0374
dad.ins.L	-0.4766	0.0337	-0.5587	0.0355
rd.age6+:age	0.1162	0.0278	0.2053	0.0314

Table A.4: Nonparametric Bandwidth Summary for Model 2

Variable	Bandwidth	Scale Factor
hlthsk	0.07	1.00
rd.age	0.59	8.18
age	53265988.06	57085641.52
insure	0.07	1.04
inc	1119.85	2.97
pop.group	0.16	2.27
prov	0.07	0.95
urban	0.94	13.02
hh.size	0.34	4.77
time.med	0.28	3.96
mom.ed	0.52	7.25
mom.live	0.08	1.13
mom.ins	0.80	11.15
dad.ed	0.69	9.63
dad.live	0.38	5.30
dad.ins	1.00	13.92

Table B.1: MNL Parameter Summary below Threshold

variable	pub:coef	pub:se	prv:coef	prv:se
(Intercept)	0.0572	0.0520	-0.9826	0.0486
weeks	-0.0024	0.0007	-0.0036	0.0007
insure1	1.1949	0.0463	1.1869	0.0394
inc	-0.0000	0.0000	-0.0000	0.0000
pop.groupcolour	-0.1952	0.0229	-0.0945	0.0192
pop.groupasian	0.3762	0.0138	1.2640	0.0140
pop.groupwhite	-0.1676	0.0224	0.5440	0.0241
prov2	-0.0027	0.0783	-0.1487	0.0608
prov3	0.7559	0.0079	-0.2261	0.0059
prov4	-0.5638	0.0293	-0.0665	0.0248
prov5	0.1362	0.0661	0.0301	0.0558
prov6	0.5722	0.0217	0.6112	0.0192
prov7	-0.0769	0.0422	0.1857	0.0411
prov8	-0.5389	0.0137	0.0498	0.0120
prov9	-0.1368	0.0058	0.4062	0.0054
urban.L	-0.0055	0.0617	0.1266	0.0548
hh.size.L	0.0592	0.0337	0.5413	0.0264
hh.size.Q	0.0323	0.0450	0.0382	0.0409
hh.size.C	-0.1614	0.0538	-0.0319	0.0500
hh.size^4	-0.2354	0.0676	-0.2491	0.0608
hh.size^5	-0.0735	0.0540	0.2024	0.0490
hh.size^6	-0.1900	0.0537	-0.3005	0.0474
time.med.L	-0.0138	0.0594	0.2978	0.0479
time.med.Q	-0.0295	0.0589	-0.0756	0.0503
time.med.C	0.1461	0.0771	-0.2370	0.0587
mom.ed.L	-0.3339	0.0534	-0.2948	0.0483
mom.ed.Q	-0.1952	0.0548	0.3010	0.0480
mom.ed.C	-0.0567	0.0662	0.0957	0.0509
mom.ins.L	-0.8870	0.0537	-0.4186	0.0523
dad.ed.L	0.2271	0.0556	0.3587	0.0543
dad.ed.Q	0.0943	0.0569	0.3086	0.0518
dad.ed.C	-0.0205	0.0712	-0.1415	0.0617
dad.ins.L	-0.3439	0.0474	-0.3307	0.0534

Table B.2: MNL Parameter Summar above Threshold

variable	pub:coef	pub:se	prv:coef	prv:se
(Intercept)	-0.0166	0.0683	-0.8981	0.0661
weeks	0.0010	0.0008	0.0013	0.0008
insure1	0.3017	0.0521	1.6025	0.0439
inc	-0.0000	0.0000	-0.0000	0.0000
pop.groupcolour	-0.2617	0.0364	-0.1776	0.0406
pop.groupasian	-0.8645	0.0164	0.7089	0.0205
pop.groupwhite	-0.6966	0.0334	0.1995	0.0382
prov2	-0.2800	0.0418	-0.5915	0.0244
prov3	-0.0713	0.0098	-0.2086	0.0079
prov4	-0.4731	0.0102	-0.4345	0.0075
prov5	0.4409	0.0818	0.1270	0.0670
prov6	0.3954	0.0307	0.4466	0.0279
prov7	0.4134	0.0169	-0.1693	0.0222
prov8	-0.0058	0.0129	-1.0195	0.0081
prov9	-0.1888	0.0055	-0.2248	0.0049
urban.L	0.1328	0.1002	0.2320	0.1053
hh.size.L	1.0390	0.0323	1.3553	0.0269
hh.size.Q	-0.0434	0.0365	0.3454	0.0292
hh.size.C	-0.0826	0.0503	-0.6634	0.0435
hh.size^4	0.0994	0.0603	0.7045	0.0576
hh.size^5	-0.0188	0.0632	-0.5140	0.0648
hh.size^6	-0.0223	0.0635	0.7869	0.0554
time.med.L	0.3193	0.0542	0.5379	0.0500
time.med.Q	-0.2098	0.0579	0.1259	0.0510
time.med.C	-0.1355	0.0641	-0.1818	0.0477
mom.ed.L	0.2755	0.0741	0.1909	0.0726
mom.ed.Q	-0.1217	0.0649	0.2240	0.0596
mom.ed.C	-0.1538	0.0901	-0.0703	0.0660
mom.ins.L	-0.2260	0.0631	0.0062	0.0585
dad.ed.L	0.2420	0.0629	0.3754	0.0639
dad.ed.Q	0.1688	0.0700	0.0479	0.0608
dad.ed.C	0.1347	0.0968	0.2446	0.0813
dad.ins.L	-0.7937	0.0505	-0.8068	0.0653

Table B.3: Nonparametric Bandwidth Summary below Threshold

Variable	Bandwidth	Scale Factor
hlthsk	0.06	0.67
weeks	579.79	21.80
insure	0.30	3.26
inc	2977.29	6.63
pop.group	0.03	0.30
prov	0.07	0.77
urban	0.13	1.42
hh.size	0.69	7.54
time.med	0.40	4.32
mom.ed	0.58	6.32
mom.ins	1.00	10.90
dad.ed	1.00	10.90
dad.ins	1.00	10.90

Table B.4: Nonparametric Bandwidth Summary above Threshold

Variable	Bandwidth	Scale Factor
hlthsk	0.05	0.52
weeks	56311760.86	1906861.74
insure	0.07	0.65
inc	16146142182.19	28404281.31
pop.group	0.80	7.71
prov	0.19	1.83
urban	1.00	9.69
hh.size	0.39	3.77
time.med	0.42	4.02
mom.ed	0.67	6.48
mom.ins	1.00	9.69
dad.ed	0.35	3.40
dad.ins	0.28	2.72