DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES

2013-14

McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

http://www.mcmaster.ca/economics/

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

Steven F. Koch† and Jeffrey S. Racine‡

September 6, 2013

Abstract

We apply parametric and nonparametric regression discontinuity methodology within a multinomial choice setting to examine the impact of public health care user fee abolition on health facility choice using data from South Africa. The nonparametric model is found to outperform the parametric model both in- and out-of-sample, while also delivering more plausible estimates of the impact of user fee abolition (i.e. the 'treatment effect'). In the parametric framework, treatment effects were relatively constant – around 7% – and that increase was drawn equally from both home care and private care groups. On the other hand, in the nonparametric framework treatment effects were largest for the least well-off (also around 7%) but fell for the most well-off. More plausibly, that increase was drawn primarily from the home care group, suggesting that the policy favoured those least well-off as more of these children received at least some minimum level of professional health care after the policy was implemented. Regarding the most well-off, despite having access to free public health care, children were still far more likely to receive health care at private facilities than at public facilities, which is also more plausible in South Africa's two-tier health sector.

*Koch would like to thank seminar participants at the University of the Free State and Emory University, as well as participants at the workshop for the Microeconometric Analysis of South African Data and Economic Research Southern Africa's Public Economics Workshop for their helpful comments and suggestions. Koch would also like to thank Dane Kennedy and the Centre for High Performance Computing (CHPC:www.chpc.ac.za) for their support. Racine would like to thank the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) for their ongoing support and to gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC:www.nserc.ca) and from the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and also acknowledge John Kealey whose feedback led to many expository improvements in the manuscript. All remaining errors are the sole responsibility of the authors.

†Department of Economics, University of Pretoria, Pretoria, Republic of South Africa, *steve.koch@up.ac.za*

‡Department of Economics & Graduate Program in Statistics, Department of Mathematics, McMaster University, Hamilton, Ontario, Canada L8S 4M4, *racinej@mcmaster.edu*

1 Introduction

Although Thistlethwaite & Campbell's (1960) regression discontinuity (RD) methodology did not, initially, receive much attention in economics, RD applications have become increasingly prevalent; see the recent reviews by van der Klaauw (2008) and Lee & Lemieux (2010) by way of illustration. RD is likely to underpin empirical assessment of policy impacts for the foreseeable future, particularly given the recent authoritative guide by Imbens & Lemieux (2008) that facilitates its implementation. As highlighted in the aforementioned reviews and guide, part of RD's appeal lies in delivering visual summaries of policy effects ('treatment effects') that are immediately accessible to the practitioner and policy analyst alike. In many cases, it is possible to instantly summarize and communicate changes in average outcomes at the RD threshold, even when that threshold is fuzzy.

RD analyses tend to focus on average treatment effects, typically through the application of linear parametric ordinary least squares. If the threshold is fuzzy, however, the construction of local average treatment effects (Imbens & Angrist 1994) on compliers via linear parametric two-stage least squares is often adopted. Linear regression models remain popular in this setting, even when the outcome data is discrete.[1] An early example of an application of the so-called linear probability model (i.e. linear regression with binary discrete outcomes) can be found in DiNardo & Lee (2002), but more recent examples abound; see by way of illustration Silles (2009), Lindeboom, Llena-Nozal & van der Klaauw (2009), Kerr, Lerner & Schoar (2010) and Arcand & Wouabe (2010). With respect to multinomial discrete outcomes, both Lalive (2008) and Schmieder, von Wachter & Bender (2012) treat discrete duration data as if it were continuous in a linear parametric regression setting. Sometimes discrete outcome data is treated as if it were continuous via the construction of cell means; see Lemieux & Milligan (2008) and Carpenter & Dobkin (2009) by way of illustration;[2] see also Zuckerman, Lee, Wutoh, Xue

[1] Regression models estimate a conditional mean, $E(Y|x)$, as opposed to a conditional probability, $Pr(Y = y|x)$. For the binary (0/1) outcome case these objects coincide since $E(Y|x) = 0 \times Pr(Y = 0|x) + 1 \times Pr(Y = 1|x) = Pr(Y = 1|x)$ (this does not hold for multinomial discrete outcomes). But with linear regression and a binary outcome, the estimated conditional mean can lie outside $[0, 1]$, thereby violating basic axioms of probability. It is often avoided for this (and other) reasons; see Aldrich & Nelson (1995) for details.

[2] In addition to parametric linear models, Lemieux & Milligan (2008) and Carpenter & Dobkin (2009)

& Stuart (2006) who apply linear least squares methods to discrete count data. Other times, multinomial discrete outcome data is re-categorized, as in Coe & Zamarro (2011), into binary outcomes such that linear probability models can be applied (although they mention considering probit models for the binary outcomes, they do not report the results). On the other hand, a nonparametric approach might proceed by treating and modeling variables according to their natural domain, i.e. 'nominal', 'ordinal', or 'numeric' ('discrete' or 'continuous'), and then modeling a conditional probability directly for nominal/ordinal/discrete outcomes rather than a conditional mean.

As underscored above, the use of linear regression with discrete, count, and multinomial outcomes remains popular.[3] Although linear models are easy to use and easy to interpret, their widespread adoption and unquestioning use is cause for concern, particularly when the outcome is binary or multinomial. One well-known problem is that it is possible for the predicted outcome for any observation, on either side of the RD threshold, to lie outside of the unit interval thereby violating basic axioms of probability. Although estimates that lie outside the unit interval may be uncommon, none of the previously mentioned linear regression-based studies discuss this shortcoming. With respect to the the impact of policy, linear probability models that at times generate predictions outside the unit interval can undermine estimates thereof. For instance, treatment effects reported in Card, Dobkin & Maestas (2004) are consistently larger when generated under the linear probability model than under the probit model.[4]

In the case of unordered categorical outcomes (the case considered in our analysis), linear and nonlinear regression models are simply inappropriate. Although it is possible to estimate separate linear regressions for every binary pair in the set of unordered categorical outcomes, such an analysis ignores the risk that predictions might fail to satisfy simple axioms of probability. It also ignores the potential for choice dependence since

consider local linear nonparametric regression; see below.

[3]Some exceptions to this include Albouy & Lequien (2009) and Ou (2010), who use non-linear probability models (i.e. non-linear regression with discrete outcomes) within an RD setting, rather than linear models in their analyses.

[4]As an aside we note that for duration data, although the use of linear probability models could produce invalid probability estimates, their use in such settings is even more problematic due to the presence of 'duration dependence', under which individuals with similar durations are likely to have common unobserved factors affecting the outcomes. Ignoring duration dependence, implicitly ignored when adopting least squares regression, could also lead to bias in the estimated treatment effect.

the relationship between such categorical outcomes is independent of other outcomes by construction.[5]

In this paper, we use an RD design to examine the effect of the 1994 abolition of public health sector user fees on health care-seeking behavior of ill South African children. The policy eliminated health care fees for children under the age of six, pregnant and nursing mothers, as well as the elderly. However, health care services in South Africa are provided by both the private and public sectors. Therefore, ill children can receive professional health care in either the public or the private sector, or not receive any professional care at all (home care). The effect of user fee abolition on the use of public health care services is analyzed while taking into account this multinomial trio of unordered health care-seeking options (home care, private, and public). We avoid the methods outlined above, and instead estimate a parametric linear index multinomial logit model, a specification that has been used in similar settings. We then estimate a nonparametric multinomial outcome model that constructs the conditional probability directly; both of these empirical specifications guarantee that basic axioms of probability are satisfied. The nonparametric model is found to fit the data better than the popular multinomial logit model in both in-sample and out-of-sample assessment. These results suggest that the linear index multinomial logit model (which is not altogether different from the linear probability model employed in the majority of previous RD studies) is inappropriate in our setting.

We also construct estimates of average treatment effects across the sub-population most likely to be affected by the policy, i.e. the least well-off from a socio-economic perspective. The measured impacts, summarized in a series of figures, indicate that non-constant treatment effects are at work in the data. The robust nonparametric results therefore call into question the commonly maintained assumption that treatment effects are constant. They also raise questions about the dominant focus in the literature on scalar estimates computed from pooled linear probability models with an RD indicator.

We are certainly not the first to adopt nonparametric methods within an RD context. Both Hahn, Todd & van der Klaauw (2001) and Imbens & Lemieux (2008) outline non-

[5]This problem is qualitatively similar to violating the Independence of Irrelevant Alternatives (IIA) assumption in the multinomial logit regression setting.

parametric local linear regression methods and discuss practical problems associated with nonparametric regression at a boundary point, which is important when using certain nonparametric methods in an RD setting.[6] Although the use of local linear regression mitigates boundary-bias problems, the local linear regression estimator (a non-linear probability model) is subject to the same critique as linear probability models in multinomial choice settings. That is, in multinomial choice settings, the estimated probabilities can lie outside the unit interval and violate basic axioms of probability. Furthermore, applying this nonparametric method to each binary pair that can be defined in a multinomial outcome setting, especially if those outcomes cannot be ranked, would implicitly assume independence of irrelevant alternatives (IIA), and thus applying the method within such a setting may not be valid. The IIA assumption is not presumed in the nonparametric method outlined by Hall, Racine & Li (2004) that is applied below. Therefore the method can be applied in all categorical outcome models, binary or multinomial, ranked or unranked. In addition, we model conditional probabilities directly rather than conditional means, as is done by the local linear approach which otherwise mirrors the linear probability model.[7] While we are critical of the dominant linear parametric paradigm, we intend this paper to be constructive and instructive in nature. Not only are nonparametric methods capable of revealing features of the data that are masked by rigid parametric specifications, but they also offer practitioners a feasible alternative to such approaches as we hope to demonstrate below. All code for the analysis undertaken in this paper is available upon request from the authors.

2 Methodology

The user fee policy change announced in 1994 consisted of a number of components, including free public health care for ill children under the age of six, the elderly, as well as pregnant and nursing mothers. However, our analysis focuses only on the demand

[6] Their methods rely on Cheng, Fan & Marron's (1997) triangular kernel, and have been applied by Carpenter & Dobkin (2009) and McCrary & Royer (2011). Optimal bandwidths for these estimators are outlined in Imbens & Kalyanaraman (2012).

[7] As an aside, the RD methodology outlined below can be applied in a discrete duration data setting, including the case where there is duration dependence, so the generality of this approach ought to be appealing to practitioners.

for curative care services for children under the age of six, given that data limitations preclude consideration of preventative care, antenatal care or effects related to nursing mothers. Furthermore, a number of other changes related to retirement pensions were enacted within a similar time frame. Thus, it was not possible to consider the effect of the policy on the elderly.[8] The demand for curative care services is analyzed within the context of health care facility choice. Gupta & Dasgupta (2002), among others, note that provider choice decisions are primarily related to curative care.

The component of the South African user fee abolition policy considered here was based on an age threshold, and so the analysis will be based on the application of RD. The age data described below is generally only available in years, although it is possible to merge exact birth dates 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 do not make use of exact birth dates in this analysis.[9]

As noted previously, the policy was designed to improve access to health care within the public sector, although other health care-seeking options are available for ill children. These options, such as care within the private sector and home care, are potential substitutes for public care. Therefore, the analysis is placed within a three-outcome model of health care facility choice. A parametric analysis of multinomial outcomes could be built on a multinomial logit or probit framework, which is where we shall begin our analysis (we report results for the logit model only, as both link functions deliver similar results). However, in addition to the multinomial logit framework, we also undertake nonparametric analysis based on direct estimation of conditional probabilities for the reasons outlined earlier. Each is described, in turn, below.

[8] For further information about the policy and previous analyses of the policy impact, the interested reader is pointed to Koch (2012) and the citations therein.

[9] Analysis with a continuous running variable is available, although the main results and conclusions presented here remain unaffected.

2.1 Parametric Multinomial Logit Analysis

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

$$Y_i = \begin{cases} 0, & \text{No professional medical treatment sought (home care)} \\ 1, & \text{Treatment sought at a public facility} \\ 2, & \text{Treatment sought at a private facility.} \end{cases} \tag{1}$$

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 socio-economic and demographic characteristics of the ill child, including a function of the child's age. Given its central role in the analysis, we will make the age function explicit a little later on in our discussion. Following 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 definition, $\sum_j p_{ij} = 1$, such that parameters in the parametric model can only be identified relative to a base category. Without loss of generality, $j = 0$ (home care) will be the base category.

Finally, assuming that the stochastic error 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 coefficient vectors, β_1 and β_2, are for outcome choices 1 and 2, respectively, and they are relative to home care (outcome 0). That is,

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

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

$$p_{i2} = e^{x_i' \beta_2} \left(1 + \sum_{j=1}^{2} e^{x_i' \beta_j}\right)^{-1}. \tag{4}$$

The multinomial logit model can be estimated via maximum likelihood, where, 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, $\mathbf{1}(y_i = j)$, assumes a value of 1 if health care choice j is chosen for child i, and 0 otherwise. The model is estimated using the 'multinom' function in the R (R Core Team 2013) package 'nnet' (Venables & Ripley 2002, Version 7.3-7).

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

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

is completely independent of the base choice, and would remain so for any other choices that could be added to the set of outcomes. Although IIA is a testable assumption (see e.g. Small & Hsiao (1985)), it will not be formally tested here, given the dominant performance of the robust nonparametric approach. Instead, the predictive performance of the multinomial logit model 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 – for instance, through the nesting of alternatives, the allowance of random parameters, or assuming normally distributed, but correlated, stochastic error terms. We leave such analysis to the interested reader.

2.2 Nonparametric Conditional Probability Analysis

Although IIA can be relaxed in a number of different ways, most of the options remain restrictive and are, at least to some degree, ad hoc. For example, nesting requires the practitioner 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 at a health care facility, and once that decision is made, the treatment location might be selected. However, there is no reason to believe that the presumed nesting structure is necessarily valid. Meanwhile, assuming multivariate normality imposes a

distribution on the error structure that may not be correct. Therefore, we also consider a consistent nonparametric estimator of the outcome probabilities rather than relying unquestioningly on the parametric multinomial logit model to obtain estimates of the respective probabilities.

Begin by defining $f(\cdot)$ and $m(\cdot)$ as the joint and marginal densities of (X, Y) and X, respectively, where Y represents the unordered categorical outcomes associated with health facility choice outlined in (1), while X can include continuous, ordered and unordered categorical variables. The conditional probability density function of $Y = y$, given $X = x$, is defined by

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

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

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

Given the mix of continuous variables, ordered discrete variables and unordered discrete 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

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

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

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

Within the structure of equations (9) and (10), there are three different X data types

along with the unordered outcome Y, requiring different kernel specifications. In the analysis, we use: a second-order Gaussian kernel for continuous predictors ('$W(\cdot)$'), the Li & Racine (2007) kernel for both unordered categorical predictors ('$\ell^u(\cdot)$') and ordered categorical predictors ('$\ell^o(\cdot)$'), and Aitchison & Aitken's (1976) unordered kernel for the outcome ('$\ell_y^u(\cdot)$'). For positive bandwidth $h_k > 0$,

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

and, for $\lambda_k \in [0,1]$ and $\lambda_0 \in [0, 0.5]$,

$$\ell^u(X_{ik}^u, x_k^u) = \begin{cases} 1 & \text{if } x_k^u = X_{ik}^u \\ \lambda_k & \text{if } x_k^u \neq X_{ik}^u \end{cases}, \tag{12}$$

$$\ell^o(X_{ik}^o, x_k^o) = \begin{cases} 1 & \text{if } x_k^o = X_{ik}^o \\ \lambda_k^{|x_k^o - X_{ik}^o|} & \text{if } x_k^o \neq X_{ik}^o, \end{cases}, \tag{13}$$

$$\ell_y^u(Y_i^u, y^u) = \begin{cases} 1 - \lambda_0 & \text{if } y^u = Y_i^u \\ \lambda_0 & \text{if } y^u \neq Y_i^u \end{cases}. \tag{14}$$

Although other kernels can be used, the estimates are relatively insensitive to the choice of the kernel (see Li & Racine (2007) for details). Instead, it is the choice of bandwidth vector $\gamma = (h, \lambda)$ that is paramount, and we choose delete-one likelihood cross-validation for this purpose (Duin 1976). In addition to being computationally tractable, this method has strong intuitive appeal for those familiar with the likelihood principle. Furthermore, selecting γ to maximize the delete-one log-likelihood function given by

$$\ln \mathcal{L}(\gamma) = n^{-1} \sum_{i=1}^{n} \ln \hat{g}_{-i}(y_i | x_i) \tag{15}$$

yields a density estimate which is close to the true density in terms of Kullback-Leibler information distance, where $\hat{g}_{-i}(y_i|x_i)$ is the conditional density estimate constructed from all the data points except the ith. As an added bonus, it possesses the ability

to remove irrelevant predictors from the analysis along the lines of Hall et al.'s (2004) more computationally intensive least-squares cross-validation method. Estimation is undertaken using the 'npcdens' and 'npconmode' functions in the R (R Core Team 2013) package 'np' (Hayfield & Racine 2008, Version 0.50-1).

2.3 Model Comparison

The preceding discussion outlined two different estimation methodologies, the parametric linear index multinomial logit model 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 for the interested reader. Intuitively, apparent error is derived from in-sample measures of fit, such as R^2 in linear regression, while true error is derived from out-of-sample 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) applied to multinomial outcomes (Racine & Parmeter forthcoming).

The outcomes Y_i are mapped to a $k \times 1$ vector Υ_i, one value for each of the three health care facility options:

$$\Upsilon_{ik} = \begin{cases} 1 & \text{if } Y_i = k \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The predictions from the empirical model define a similar prediction vector, $\hat{\Upsilon}_i$, which is based on the predicted probabilities from the model:

$$\hat{\Upsilon}_{ik} = \begin{cases} 1 & \text{if } \hat{p}_{ik} = \max_j\{\hat{p}_{ij}\} \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Given these predictions, we define a loss function penalizing incorrect predictions given

by

$$Q_i(\Upsilon, \hat{\Upsilon}, n) = \begin{cases} 1 & \text{if } \Upsilon_i = \hat{\Upsilon}_i \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where Υ and $\hat{\Upsilon}$ are $n \times k$ matrices whose ith columns are Υ_i and $\hat{\Upsilon}_i$, respectively. The loss function can then be used to define the correct classification ratio (CCR):

$$CCR = n^{-1} \sum_{i=1}^{n} Q_i(\Upsilon, \hat{\Upsilon}, n). \tag{19}$$

In addition to the loss function and CCR, the underlying $k \times k$ confusion matrix (CM) provides useful information regarding a model's ability to properly predict actual choices. The CM tabulates the counts of actual outcomes against predicted outcomes defined as

$$CM = \Upsilon' \hat{\Upsilon}. \tag{20}$$

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 (19) is an in-sample estimate of the expected loss, or apparent 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) $\hat{F}$. The training sample would yield an estimate of apparent error, $E_{n_1,\hat{F}}[Q(\Upsilon, \hat{\Upsilon}, n_1)]$, which is not of interest here; 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 an estimate of true error, $E_{n_2,F}[Q(\Upsilon, \hat{\Upsilon}, n_2)]$. The expected true error is the expectation of the estimator of true error, $E\{E_{n_2,F}[Q(\Upsilon, \hat{\Upsilon}, n_2)]\}$. This can be constructed as the sample average of repeated estimates of true error based on repeated shuffles of the full data set which are then split into the training and evaluation samples of sizes n_1 and n_2, respectively.

The preceding discussion hints at the resampling procedure used to assess model performance outlined by Racine & Parmeter (forthcoming).

1. Shuffle 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.
3. Use estimates from the full sample, i.e. in the case of the nonparametric model hold smoothing fixed, and in the case of the multinomial logit model, hold functional form 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, $T = 10,000$ – which results in T draws of CCR for both models.

The draws from the resampling procedure are used to construct and contrast the underlying empirical distribution functions of expected true error for the multinomial logit and nonparametric models, respectively. We report boxplots along with the median and mean values from the empirical distribution of CCRs for each model, and tests for 'equal performance' are based on these statistics (P-values from these tests are reported in the captions for figures 1 and 2).

2.4 Policy Impacts

Having estimated and compared the parametric and nonparametric approaches, we then proceed to examine the impact of user fee abolition on health care facility choices, based on the difference between predicted facility choice probabilities across the RD threshold. Given the policy focus on public care, this is central to our analysis. Rather than assuming fixed treatment effects, we analyze and present differences across quantiles of the explanatory variables (X_q below). We denote quantiles with $q \in [0, 1]$, and they encompass relative living standards (socio-economic 'well-being') that increase with q.

Essentially, the average difference (across the age threshold defined by the policy) in the predicted probability of a child receiving professional health care in either a private or a public facility, or not receiving professional care (home care), is calculated at various levels of q. Recalling that j represents health facility choice, the treatment effect for each facility option at each quantile, denoted τ_{jq}, is as follows:[10]

$$\hat{\tau}_{jq} = n^{-1} \sum_{i=1}^{n} \left[\hat{p}_{ij}(X_q, \text{age} = 5, \text{rd.age}{=}0) - \hat{p}_{ij}(X_q, \text{age} = 6, \text{rd.age}{=}1) \right], \quad j = 0, 1, 2 \quad (21)$$

[10] Given $\sum_j \hat{p}_{ij} = 1$, by construction, $\sum_j \hat{\tau}_{jq} = 0 \ \forall q$.

(the construction of X_q is described in detail in Footnote 15 where we discuss evaluation of the policy change, 'rd.age' is a binary factor taking value 0 for children under the age of 6, 1 for those 6 or older). For the multinomial logit model, the $\hat{p}_{ij}$s are estimated via (5) in which the age function is a linear function of age (i.e. piece-wise linear in age – this 'linear index' specification is common in the literature). Meanwhile, for the nonparametric model, the $\hat{p}_{ij}$s are estimated via (8), and rd.age and age are simply part of the information set.

In order for (21) 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. The difference would also be consistently estimated, as long as any bias in the predicted probabilities across the age threshold was constant and could be differenced-out. Potential confounders 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. However, this would be a mistake at the facility, as opposed to something that a caregiver could guarantee, and thus, it is not expected to be a significant source of bias. Similarly, it is possible that children under six could be more likely to be reported as ill in the data, since they could receive free health care at a public facility. If such an anticipation effect was in the data, but not properly addressed, estimated policy impacts would likely be overstated. Unfortunately, there are no exclusion restrictions lying dormant in the data that could be used to identify either facility-level mistakes or anticipation effects.[11] Due to the lack of instruments, the policy impacts estimated below are to be interpreted with such limitations in mind. This caveat notwithstanding, the estimated effects are most plausibly understood as lower bounds on the actual impacts.

Finally, confidence intervals for the average policy impact within a data quantile are calculated via bootstrap methods. Following Li, Racine & Wooldridge (2008), samples of the data are drawn, with replacement, from the original data on which the sample treatment effect was constructed. The average treatment effect at a given quantile is calculated for each resample, and the process is repeated $B = 100$ times. This yields a

[11] 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. However, Koch (2012) highlights a reduction in reported illnesses among those eligible for free public health care, suggesting that anticipation effects are not a cause for concern.

series of resampled estimates of the policy impact at a given quantile, which are then used to construct a 90% confidence interval around the sample treatment effect.

3 Data

3.1 Data Source

Data for the analysis is sourced from the South African October Household Survey (OHS) of 1995.[12] The main purpose of the OHS conducted by Statistics South Africa (1995), was to collect information on households and individuals across the nine provinces of South Africa. The survey includes questions related to dwellings/dwelling services, perceived quality of life, socio-demographics, employment/unemployment, informal and formal labour markets, as well as births and deaths in the household. In addition to this information, there is a short series of questions related to illness, injury, health care-seeking behavior and access to medical aid or health insurance.

The survey follows a stratified random sampling method, explicitly stratified by province, magisterial district, urban/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. From this sampling process, responses are available for $121,538$ individuals living in $29,700$ households. However, we further restrict the sample to recently (within the last 30 days) ill or injured children under the age of 14. Therefore, the sample only includes children potentially affected by the policy who are reasonably close to the age threshold, which is in keeping with the RD. The resulting sample contains $2,716$ such children, which nearly constitutes 12% of all children in that age range in the survey. Although post-stratification weights are available, they are not used in the analysis because the weights are not calibrated for a subsample of this nature.

A series of sections in the survey cover a variety of different topics, although one can

[12] Although data from 1994 is also available, 1994 was deemed too soon after policy implementation for purposes of analysis. Furthermore, data was not similarly collected in 1994 and 1995, especially with regard to health facility choice.

merge all relevant information to create child-level data. For this analysis, data for each child is taken from the individual questionnaire, and this data includes information on the mother and father. Information about the child's mother and/or father is merged into the child dataset, as is the household-related data.

For the analysis, only health care-seeking behavior is considered. Therefore, we consider the multinomial outcomes a) care for the ill or injured child was sought in a public facility, b) care for the ill or injured child was sought at a private facility or c) care was not sought not at all. In addition to the outcome variable, we create dummy variables for the child's sex, population group and the age eligibility. Note that the latter age dummy is used for the parametric model only as the nonparametric model treats this simply as a multinomial predictor. Age serves as the running variable in the analysis, and it is available for all children in the survey. Household-level controls include the size of the household, the distance (measured in time) the dwelling sits from the health facility usually attended when a household member seeks medical care, and categorical variables for urban/provincial locales. For mothers and fathers, we first determine whether or not either parent is alive, and then we capture their education and whether or not they have health insurance.[13]

3.2 Data Description

The variables included in the analysis are described in Tables A.1 and A.2. Table A.1 describes the data for all ill or injured children under the age of 6, while Table A.2 does the same for children over the age of 6. As expected, the mean age and the mean value of the RD indicator are higher in Table A.2 than in Table A.1. Otherwise, the means presented in these tables suggest only minor differences between the two subsets.

4 Empirical Model Comparison

Before examining the impacts of policy, we examine the empirical fit of the multinomial logit and nonparametric models. The comparison begins on a subset of the chosen

[13]Very few children in the sample do not have at least one parent alive. For that reason, a secondary analysis, dropping the indicators for either parent being alive, was undertaken. The results did not differ markedly from what is presented below.

variables, and within this framework, both models perform comparably. However, when the analysis is extended to include additional variables, model performance diverges rather starkly.

4.1 A Baseline with Similar Predictive Performance

In addition to the outcome (health care facility choice 'hlthsk'), variables included in the initial analysis are limited to controls for household earnings ('inc'), a binary indicator of access to health insurance ('insure'), the running variable ('age'), a binary indicator of policy eligibility ('rd.age') and, in the case of the multinomial logit, the age-eligibility interaction.[14] As parameter estimates in and of themselves are not the focus of the analysis, the multinomial logit estimates have been relegated to Appendix Table B.1, while the nonparametric bandwidths and their scale factors have been relegated to Appendix Table B.2. Although the estimates are not the primary focus, the results suggest that the control variables are statistically significant determinants of health facility choice.

Rather than focusing on parameter estimates, we focus on model performance primarily for purposes of benchmarking. A secondary secondary reason for this focus is to see if the parametric multinomial logit model assumption appears reasonable. Specifically, the empirical results and the data are used to calculate in-sample performance, which is presented in Table 1 and Table 2, and out-of-sample performance, which is discussed below. One of the striking results in the two tables is the inability of either model to predict home care 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 1) and the nonparametric model (see Table 2) is similar in this limited setting. For the full sample of data, the nonparametric overall CCR is 0.554, while the parametric overall CCR is 0.553. For the full sample of data, the nonparametric log-likelihood was -2648.65, while the parametric log-likelihood was -2657.64.

Although a larger CCR is indicative of better predictive power, it is important to note that the preceding CCRs are all in-sample, and represent apparent error. However,

[14]The nonparametric model includes rd.age and age separately as predictors as issues of 'functional form' are modelled nonparametrically.

Table 1: Parametric Confusion Matrix for Model 1

Actual Facility Choices	Predicted Facility Choices		
	No Treatment	Public	Private
No Treatment	0	472	124
Public	0	1136	208
Private	0	410	366

Source: Authors' calculations, see equation (20). Correct predictions observed down the diagonal, where actual choice corresponds to predicted choice.

Table 2: Nonparametric Confusion Matrix for Model 1

Actual Facility Choices	Predicted Facility Choices		
	No Treatment	Public	Private
No Treatment	0	466	130
Public	0	1126	218
Private	0	397	379

Source: Authors' calculations, see equation (20). Correct predictions observed down the diagonal, where actual choice corresponds to predicted choice.

one concern that arises in the case of nonparametric models is that the model could be overfit, thereby inflating the in-sample CCR. Even though the data-driven bandwidth selection process is theoretically optimal, it is not guaranteed to deliver sound results for every possible sample and could be misleading (i.e. not 'externally valid'). Hence, we conduct the out-of-sample performance evaluation exercise described in Section 2.3 as a robustness check on the nonparametric results. As an extra precaution, duplicate observations are removed from the training 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 and we note that results were insensitive to the removal of duplicate observations in the training resamples.

Despite the potential for overfitting in the nonparametric model, the out-of-sample fits, or expected true errors, illustrated in Figure 1 do not reveal any such issues being present here. As suggested by the in-sample comparison, there is very little difference in performance between the multinomial and nonparametric models when a limited subset of control variables are included in the initial model. Furthermore, at least for the

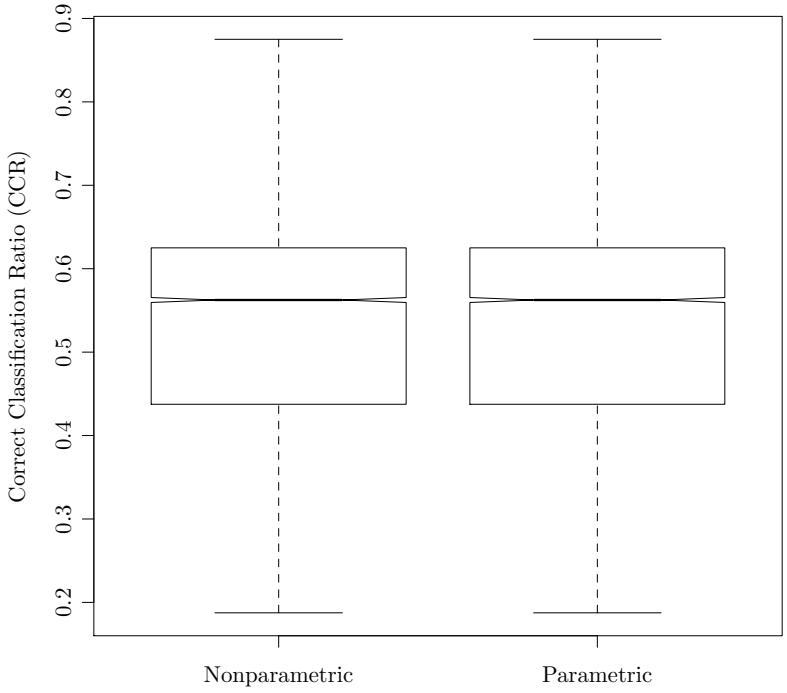

Figure 1: Boxplots for out-of-sample performance assessment of the benchmark model (mean nonparametric CCR: 0.5504, mean parametric CCR: 0.5492, median nonparametric CCR: 0.5625, median parametric CCR: 0.5625, 10^4 splits of the data, training data size $n_1 = 2,700$, evaluation data size $n_2 = 16$, higher CCR is better). The test for revealed performance under the null of equal performance delivers a P-value of 0.2469 indicating that there is no significant difference in the predictive abilities of either model.

benchmark case, the out-of-sample performance comparison is similar to the in-sample performance comparison.

4.2 Beyond the Baseline: No Longer Similar

With just a few explanatory variables, model performance does not differ, and neither model manages to predict home care outcomes successfully. However, the initial model did not include many of the determinants of health care facility choice previously identified in the literature. Therefore, in what follows, a number of explanatory variables are added to the model. These include: population group ('pop.group', one of "black", "colour", "asian", "white"), region ('prov', one of "Northern Cape", "Western Cape", "Eastern Cape", "Free State", "KwaZulu-Natal", "Northwest", "Gauteng",

"Mpumalanga", "Limpopo"), household size ('hh.size'), distance to nearest medical facility ('time.med'), an urban-rural dummy ('urban'), an indicator for whether either parent is alive ('mom.live', 'dad.live'), parental education ('mom.ed', 'dad.ed', one of "none", "some", "primary", "matric"), and parental health insurance status ('mom.ins', 'dad.ins'). As before, the focus of the analysis is not on parameters, so we relegate the multinomial logit estimates to Appendix Table B.3 and report nonparametric bandwidths and scale factors in Appendix Table B.4. For the parametric model, we see that the child's age, access to medical aid and eligibility for free public health care remain significant determinants. In addition to those variables, there are significant differences across population groups and regions. Household size and parental effects also appear to be of some importance. For the nonparametric model, bandwidths and scale factors have limited direct interpretation. However, age has been smoothed out of the estimated conditional probability (the size of the age bandwidth far exceeds the age range included in the data) which means it is deemed an irrelevant predictor by the cross-validation method. On the other hand, the indicator for free public care eligibility, rd.age, is deemed to have predictive power. Below, we see that there is a difference in health care facility choice that can be attributed to policy eligibility.

Given the fact that many of the included variables are statistically significant in the parametric model, one would expect the predictive performance of the multinomial logit model to improve. In-sample, however, this expectation does not appear to materialize. With only a few explanatory variables, just over half of the outcomes were predicted correctly, in-sample. Even though many of the additional variables are statistically significant, including them only increased the parametric model's in-sample performance to 0.581, which is a rather small improvement. Meanwhile, the nonparametric model's CCR increases rather substantially; the nonparametric overall CCR is 0.911. For the full sample of data, the nonparametric log-likelihood was -1119.65, while the parametric log-likelihood was -2523.06. As can be seen in the confusion matrices - see Tables 3 and 4 - the multinomial logit model still has very limited success in predicting home care. This is probably due to the fact that we do not have any information on the severity of the child's illness. Despite not having that information, the nonparametric

model appears to be more successful in-sample, possibly because some of the variables included in the model are correlated with the unobserved depth of child illness.

Table 3: Parametric Confusion Matrix for Model 2

Actual Facility Choices	Predicted Facility Choices		
	No Treatment	Public	Private
No Treatment	60	422	114
Public	37	1141	166
Private	23	375	378

Source: Authors' calculations, see equation (20). Correct predictions observed down the diagonal, where actual choice corresponds to predicted choice.

Table 4: Nonparametric Confusion Matrix for Model 2

Actual Facility Choices	Predicted Facility Choices		
	No Treatment	Public	Private
No Treatment	462	120	14
Public	9	1325	10
Private	9	81	686

Source: Authors' calculations, see equation (20). Correct predictions observed down the diagonal, where actual choice corresponds to predicted choice.

Although the nonparametric model exhibits superior in-sample performance, it is possible that this simply reflects overfitting the data. Therefore, the same performance comparison that was outlined above is also undertaken here. The results of the training exercise are illustrated in Figure 2. The illustration, which agrees with the in-sample performance, shows that the multinomial logit model's predictive performance leaves much to be desired relative to the nonparametric model. The out-of-sample performance of the nonparametric model (mean nonparametric CCR: 0.631) over the parametric model (mean parametric CCR: 0.566) is statistically significant at any conventional level ($P-$value$=6.0566 \times 10^{-300}$, recall that duplicate observations were removed for this exercise). In other words, even though many of the explanatory variables in the multinomial logit model are statistically significant, they do not appear to provide much by way of additional explanatory power, at least in this analysis. The nonparametric model, on the other hand, appears to be able to exploit their presence both in- and out-of-sample.

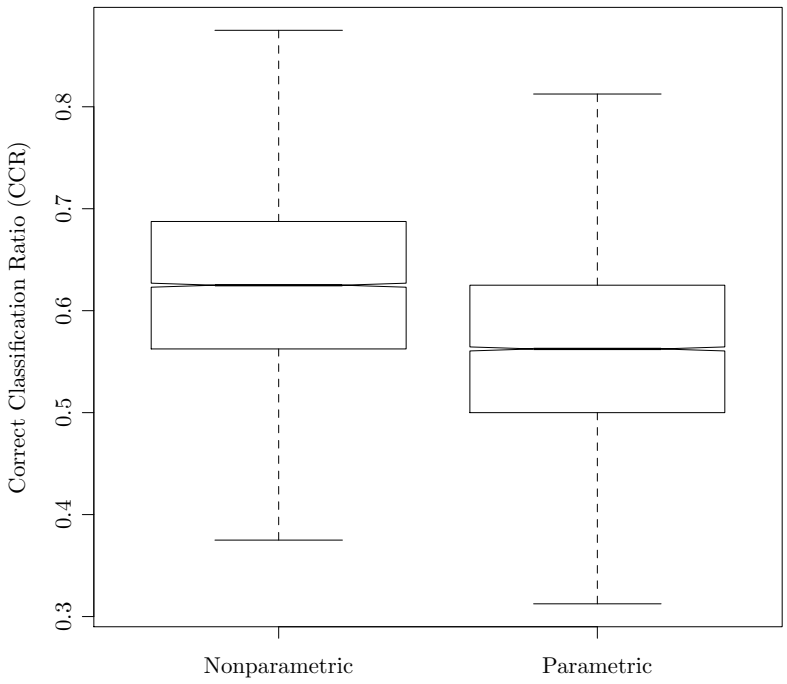

Figure 2: Boxplots for out-of-sample performance assessment of Model 2 (mean nonparametric CCR: 0.6309, mean parametric CCR: 0.5659, median nonparametric CCR: 0.625, median parametric CCR: 0.5625, 10^4 splits of the data, training data size $n_1 = 2,700$, evaluation data size $n_2 = 16$, higher CCR is better). The test for revealed performance under the null of equal performance delivers a P-value of 6.0566×10^{-300} indicating that there is a highly significant improvement in the predictive abilities of the nonparametric model over the parametric model.

5 Evaluation of Policy

Having assessed the improvement in performance of the nonparametric model relative to the parametric model, we now turn to the evaluation of the policy. As Berk & Rauma (1983) noted in their non-linear RD setting, there are many treatment effects, and some attention should be paid to extending the analysis beyond the mean. Therefore, we estimate and present average treatment effects at different quantiles of the distribution of the socio-economic predictors (i.e. levels of 'well-being'), for a fixed population group and region. In this analysis, the construction of quantiles is not limited only to income.[15]

[15]Instead, with the exception of province, race and health insurance access, which we set to KwaZulu-Natal, black and no insurance access, respectively, all other independent variables are allowed to vary. Specifically, we define all continuous variables, like income, and ordered factors, including binary factors,

The primary results from the analysis are illustrated in Figure 3 and presented across quantiles. Since the user fees were eliminated in the public sector, that is the main focus here. More detailed illustrations of quantile-specific treatment effects for both empirical models and all three health care facility choice options are illustrated in Figures 4 and 5.

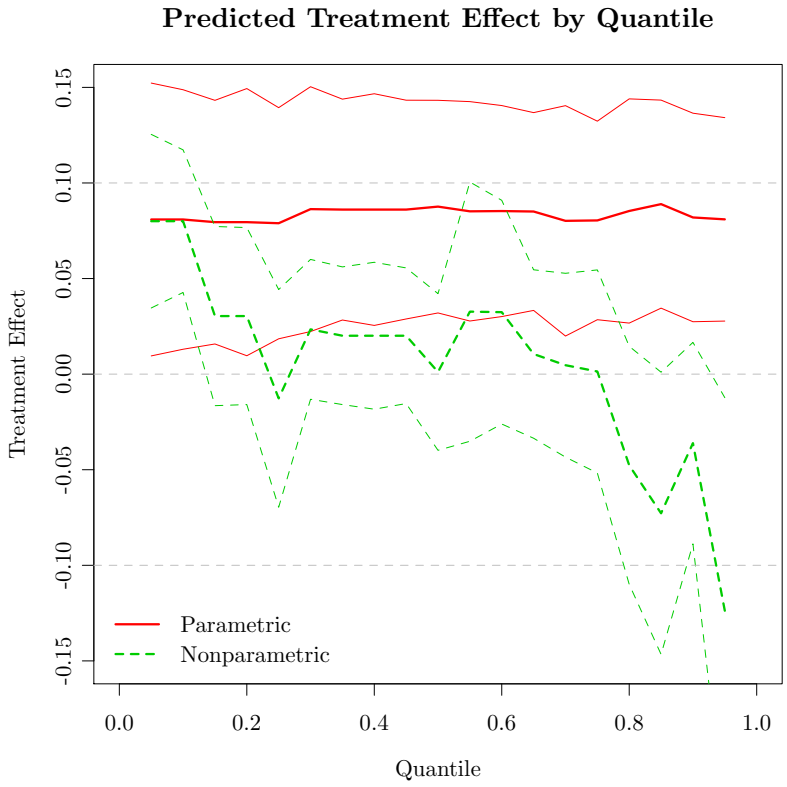

Figure 3: Estimated Treatment Effects and Bootstrapped 90% Confidence Intervals for the Effect of User Fee Abolition on the Use of Public Health Facilities Amongst Ill Children. Analysis undertaken across quantiles (0.05-0.95) of the data for black children living in KwaZulu-Natal without health insurance. Treatment effects calculated from both multinomial logit and nonparametric conditional probability models. Moving from lower to higher quantiles implies an improvement in living standards. See Footnote 15 for further clarification.

Figure 3 illustrates the parametric and nonparametric treatment effects across data quantiles, along with 90% confidence bands. The parametric treatment effects are fairly

to be "better" as the quantile value rises. For example, having an alive father or mother is better than either or both parents being dead; higher levels of parental education and more household income is also better. On the other hand, living closer to a medical facility is better than living farther away, and living in a smaller household is better than living in a larger household. Finally, in order to calculate the treatment effect as the difference across the policy (RD) threshold, the age of young children is set to 5, while the age in the older sample is set to 6.

constant as well-being (i.e. the quantile q) increases, averaging roughly 7%. Curiously, according to the parametric model, children living in the best of circumstances receive the same benefit as do less well-off children. The parametric results agree with Koch's (2012) analysis, which was based on linear probability models applied to each binary pair of health care facility choice options. On the other hand, the nonparametric results paint a more reasonable albeit somewhat noisier picture with respect to equity considerations. According to the bootstrapped nonparametric treatment effects, user fee abolition increased the use of public health care facilities among the least well-off young children by up to 7%, whereas this effect is entirely eliminated for ill children living in the best of circumstances.

The primary reason for considering the outcome data in its entirety, i.e. as an unordered categorical outcome variable, is the potential for substitution across health care facility choices. Therefore, in addition to estimating the effect of the policy on access to public health care as previously discussed, we estimated treatment effects for all three health care facility options, including no medical care. These estimates are illustrated separately for the parametric model (Figure 4) and for the nonparametric model (Figure 5).

Given the constancy in the public care treatment effects obtained from the multinomial logit model, it is not surprising that the private care and home care treatment effects are rather constant for children, regardless of their living conditions. It is also not surprising that, for children living in the worst of circumstances, as least some of the increase in public care use is driven by a decrease in the proportion of ill children not accessing any health care. There is also evidence of substitution away from the more expensive private care to the freely available public health care for those in the worst living conditions. Interestingly, this last result is also true for children living in the best of circumstances, although there is also some evidence of substitution away from home care and towards public health care for children in better living conditions. Despite the goal of the policy – which was to improve access to health care for the poorest, primarily within the public health sector, through the elimination of user fees – we observe substitution away from privately provided care towards freely provided public health

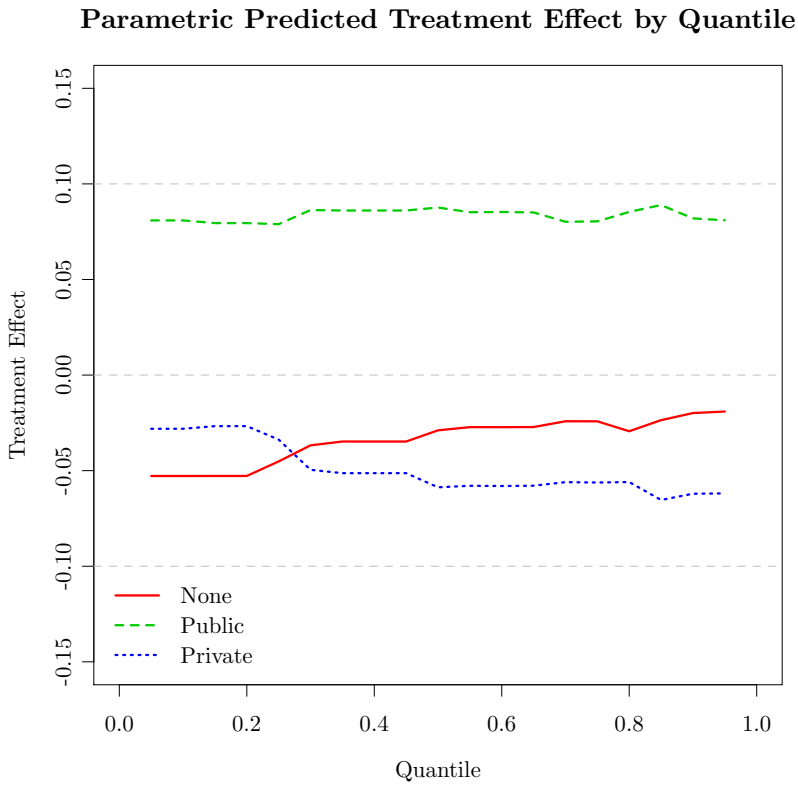

Figure 4: Estimated Treatment Effects of User Fee Abolition on all Health Facility Choices made for Ill Children. Analysis undertaken across quantiles (0.05-0.95) of the data for black children living in KwaZulu-Natal without health insurance. Treatment effects calculated from the multinomial logit model. Moving from lower to higher quantiles implies an improvement in living standards. See Footnote 15 for further clarification.

care amongst children living in the best of circumstances. In other words, according to the multinomial logit model results, the policy change benefited all children, including those it was not designed to benefit.

Within the nonparametric setting, the substitution patterns are more interesting and somewhat more plausible from the policy perspective. As was the case for the parametric model treatment effects, the least well-off children were more likely to access public health care facilities after the policy was implemented; however, the nonparametric treatment effects uncover a different substitution pattern. Rather than seeing a similar draw from both home care and private care, the increase is drawn almost entirely from home care. Very few of the poorest are able to access private health facilities in the first place, and, therefore, little substitution would be expected. As living conditions improve, up to the median, the effect of user fee abolition on the private sector remains

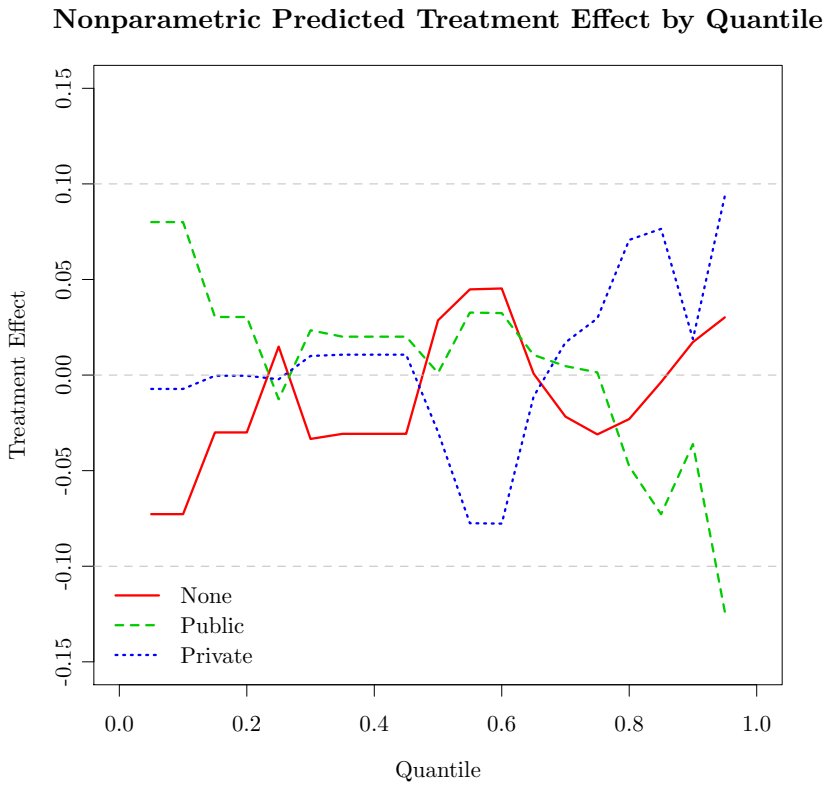

Figure 5: Estimated Treatment Effects of User Fee Abolition on all Health Facility Choices made for Ill Children. Analysis undertaken across quantiles (0.05-0.95) of the data for black children living in KwaZulu-Natal without health insurance. Treatment effects calculated from the nonparametric conditional probability model. Moving from lower to higher quantiles implies an improvement in living standards. See Footnote 15 for further clarification.

limited; the change in fees continues to be associated with a small increase in access to public health care drawn from those most likely to not have received any health care.

Within the third quartile, however, the observed small increase in receipt of public health care is drawn entirely from the private health care sector. One could make the case that children living in households capable of accessing the private health sector were affected by the policy change. The reduction in public health care user fees could have led to both simple substitution effects and broader changes in behavior. For example, Koch (2012) argues that user fee abolition reduced the proportion of children covered by health insurance. Although not investigated in that study, it is reasonable to believe that health insurance coverage would have been reduced amongst households in this particular range of well-being, not so poor that they could not originally afford health

insurance and not so well-off that health insurance could be viewed as a necessity.[16] If the impact of user fee abolition on health facility choice within this range can be attributed to the reduction in health insurance coverage, the observed increase in home care within this group can be similarly attributed.

In the best of living conditions, i.e. children living in households in the top quartile, the estimated treatment effects are the complete opposite to those experienced by children in the worst of living conditions. Despite being eligible (by age) to receive free health care in the public sector, ill children living in the best of circumstances are more likely to receive health care from the private sector and less likely to receive health care from the public sector, as would be expected in a two-tier health setting like South Africa's. In terms of treatment effects, the implied effects of user fee abolition amongst children at the top of the distribution are an increase in health care-seeking in the private sector, a reduction in the public sector, and a limited effect on home care. Although one might have expected the policy impact to have been minimal, since well-off children received their health care primarily from the private sector, the influx in patients at public facilities could have lowered the quality of care (or increased queue length), driving those with resources away from the public health care sector.

6 Conclusion

This research examines the effect of user fee abolition on health care facility choice. The analysis focuses on young children because the policy was developed, at least in part, to improve health outcomes for poor young children. The effects of that policy are modeled both parametrically and nonparametrically, under the usual RD assumption that the policy is independent of any unobserved factors that differ across children aged five and six. The results suggest that the parametric model does not fare as well as the nonparametric model in terms of prediction of outcomes, both in- and out-of-sample. In

[16]The policy was not meant to apply to children with access to health insurance, who are likely to be amongst those in the best of circumstances; health insurance coverage is only about 15% in the country, (Koch & Alaba 2010). For that reason, the health insurance indicator was switched off for the calculations illustrated in Figures 3, 4 and 5. Although the direct effect of health insurance is not included in the treatment effects, since the medical aid indicator is switched-off, indirect treatment effects arising from correlation between insurance coverage and other variables remain.

addition to the difference in model performance, the estimated treatment effects differ markedly across the two models.

Within the nonparametric setting, treatment effects are estimated to be smaller than they are in the parametric setting. Although the multinomial logit treatment effects are in the neighborhood of 7%, and remain so regardless of the child's living standards, the nonparametric treatment effects are generally smaller and disappear entirely for children living in the best of circumstances. The varied impact uncovered by the nonparametric specification suggests that the policy was pro-poor. This reinforces the views held by the nurses interviewed by Walker & Gilson (2004), even though these nurses' beliefs had not been empirically verified by any previous research. The degree to which the policy was pro-poor, however, is left for future research.

The observed differences in both model performance and treatment effects suggest that future RD research in multinomial settings should pay more attention to model choice. As was observed here, multinomial logit models may perform relatively poorly, even when predictors are found to be significant. The fact that multinomial linear probability models implicitly assume IIA further suggests that the results based on such specifications should not be accepted unquestioningly.

References

Aitchison, J. & Aitken, C. G. G. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* **63**(3), 413–420.

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

Aldrich, J. H. & Nelson, F. D. (1995), *Linear Probability, Logit, and Probit Models*, SAGE Publications, Inc.

Arcand, J. & Wouabe, E. D. (2010), 'Teacher training and HIV/AIDS prevention in west Africa: regression discontinuity design evidence from the Cameroon', *Health Economics* **19**, 36–54.

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.

Card, D., Dobkin, C. & Maestas, N. (2004), 'The impact of nearly universal insurance coverage on health care: evidence from Medicare', NBER Working Paper no. 10365.

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.

Cheng, M., Fan, J. & Marron, J. S. (1997), 'On automatic boundary corrections', *Annals of Statistics* **25**(4), 1691–1708.

Coe, N. B. & Zamarro, G. (2011), 'Retirement effects on health in Europe', *Journal of Health Economics* **30**, 77–86.

DiNardo, J. & Lee, D. S. (2002), 'The impact of unionization on establishment closure: a regression discontinuity analysis of representation elections', NBER Working Paper no. 8993.

Duin, R. (1976), 'On the choice of smoothing parameters for Parzen estimators of probability density functions', *I.E.E.E. Transactions in Computing* **C-25**, 1175–1179.

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

Gupta, I. & Dasgupta, P. (2002), 'Demand for curative care in rural India: choosing between private, public and no care', National Council of Applied Economic Research Working Paper Series no. 82.

Hahn, J., Todd, P. & van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression disontinuity design', *Econometrica* **69**(1), 201–209.

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. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**, 467–476.

Imbens, G. W. & Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *Review of Economic Studies* **79**, 933–959.

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

Kerr, W. R., Lerner, J. & Schoar, A. (2010), 'The consequences of entrepreneurial finance: A regression discontinuity analysis', NBER Working Paper no. 15831.

Koch, S. F. (2012), 'The abolition of user fees and the demand for health care: Re-evaluating the impact', *Economic Research Southern Africa, WP 307* .

Koch, S. F. & Alaba, O. A. (2010), 'On health insurance and household decisions: a treatment effects analysis', *Social Science & Medicine* **70**(2), 175–182.

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

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**, 266–292.

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

Li, Q., Racine, J. S. & Wooldridge, J. M. (2008), 'Estimating average treatment effects with continuous and discrete covariates: The case of Swan-Ganz Catheterization', *American Economic Review* **98**(2), 357–362.

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.

R Core Team (2013), *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.

Schmieder, J. F., von Wachter, T. & Bender, S. (2012), 'The effects of extended unemployment insurance over the business cycle: evidence from regression discontinuity etimates over 20 years', *The Quarterly Journal of Economics* **127**, 701–752.

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*

Walker, L. & Gilson, L. (2004), '"We are bitter, but we are satisfied": nurses as street-level bureaucrats in South Africa', *Social Science & Medicine* **59**(6), 1251–1261.

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 Descriptive Statistics

Table A.1: Observed Mean of Data for Ill Children Under 6 Years Old (and eligible for free public care) by Health Facility Choice

	Private Care	Public Care	No Treatment
Insured Child	0.434	0.174	0.222
Black Child	0.538	0.758	0.731
Coloured Child	0.113	0.137	0.133
Asian Child	0.082	0.035	0.025
White Child	0.268	0.070	0.111
Northern Cape Province	0.117	0.084	0.079
Western Cape Province	0.150	0.206	0.208
Eastern Cape Province	0.026	0.055	0.043
Free State Province	0.106	0.069	0.133
KwaZulu-Natal Province	0.230	0.302	0.254
Northwest Province	0.073	0.100	0.043
Gauteng Province	0.176	0.099	0.115
Mpumalanga Province	0.087	0.059	0.097
Limpopo Province	0.035	0.026	0.029
Urban Locale	0.709	0.539	0.563
Med Center > 60min	0.073	0.158	0.161
30min < Med Center < 60min	0.096	0.193	0.168
15min < Med Center < 30min	0.345	0.304	0.308
Med Center < 15 min away from house	0.486	0.345	0.362
>10 in Household (HH)	0.056	0.130	0.136
8-9 in HH	0.103	0.154	0.122
7 in HH	0.070	0.126	0.136
6 in HH	0.117	0.131	0.136
5 in HH	0.214	0.165	0.183
4 in HH	0.268	0.197	0.168
<4 in HH	0.171	0.097	0.118
Mother: No Education	0.178	0.203	0.233
Mom: Some Education	0.167	0.273	0.237
Mom: Primary Education	0.254	0.350	0.301
Mom: Matric Completed	0.401	0.174	0.229
Mother: Alive	0.993	0.979	0.978
Mother: Insured	0.411	0.140	0.215
Father: No Education	0.453	0.574	0.584
Dad: Some Education	0.089	0.166	0.143
Dad: Primary Education	0.157	0.152	0.147
Dad: Matric Completed	0.300	0.108	0.125
Father: Alive	0.948	0.890	0.914
Father: Insured	0.369	0.122	0.190
HH Income	2235.2	1117.4	1369.0
Child Age	2.2	2.3	2.6

Table A.2: Observed Mean of Data for Ill Children 6 Years Old or Older (and not eligible for free public care) by Health Facility Choice

	Private Care	Public Care	No Treatment
Insured Child	0.554	0.144	0.215
Black Child	0.420	0.763	0.685
Coloured Child	0.157	0.142	0.158
Asian Child	0.114	0.028	0.038
White Child	0.309	0.066	0.120
Northern Cape Province	0.171	0.087	0.104
Western Cape Province	0.117	0.197	0.208
Eastern Cape Province	0.043	0.047	0.050
Free State Province	0.063	0.044	0.107
KwaZulu-Natal Province	0.266	0.288	0.243
Northwest Province	0.103	0.125	0.076
Gauteng Province	0.151	0.106	0.085
Mpumalanga Province	0.060	0.080	0.091
Limpopo Province	0.026	0.025	0.035
Urban Locale	0.777	0.526	0.511
Med Center > 60min	0.046	0.144	0.183
30min < Med Center < 60min	0.083	0.190	0.189
15min < Med Center < 30min	0.308	0.345	0.287
Med Center < 15 min away from house	0.362	0.321	0.341
>10 in Household (HH)	0.063	0.102	0.132
8-9 in HH	0.029	0.157	0.199
7 in HH	0.074	0.106	0.114
6 in HH	0.120	0.169	0.183
5 in HH	0.297	0.180	0.161
4 in HH	0.306	0.197	0.148
<4 in HH	0.111	0.087	0.063
Mother: No Education	0.231	0.292	0.353
Mom: Some Education	0.149	0.252	0.252
Mom: Primary Education	0.271	0.315	0.262
Mom: Matric Completed	0.349	0.140	0.132
Mother: Alive	0.971	0.960	0.975
Mother: Insured	0.509	0.133	0.221
Father: No Education	0.391	0.588	0.571
Dad: Some Education	0.103	0.154	0.155
Dad: Primary Education	0.223	0.169	0.174
Dad: Matric Completed	0.283	0.089	0.101
Father: Alive	0.917	0.861	0.890
Father: Insured	0.440	0.116	0.215
HH Income	2502.1	1013.5	1262.2
Child Age	8.4	8.4	8.4

B Estimation Results

Table B.1: Parametric Multinomial Logit Model Parameter Summary for Model 1

	Public Facility		Private Facility	
Variable	Coefficient	Stand. Err.	Coefficient	Stand. Err.
(Intercept)	1.4349	0.031	0.4154	0.030
rd.age6+	-0.7866	0.032	-0.7812	0.031
age	-0.1116	0.018	-0.1921	0.020
insure1	-0.2630	0.059	1.1316	0.061
inc	-0.0000	0.000	0.0000	0.000
rd.age6+:age	0.1056	0.015	0.1890	0.017

Source: Authors' calculations.

Table B.2: Nonparametric Bandwidth Summary for Model 1

Variable	Bandwidth	Scale Factor
hlthsk	0.00	0.00
rd.age	0.33	4.66
age	1.84	1.97
insure	0.02	0.23
inc	4229.83	11.22

Source: Authors' calculations.

Table B.3: Parametric Multinomial Logit Model Parameter Summary for Model 2

	Public Facility		Private Facility	
Variable	Coefficient	Stand. Err.	Coefficient	Stand. Err.
(Intercept)	1.0944	0.019	0.1624	0.018
rd.age6+	-0.8785	0.019	-0.9040	0.018
age	-0.1096	0.019	-0.1983	0.021
insure1	0.6094	0.034	1.2019	0.033
inc	-0.0000	0.000	0.0000	0.000
pop.groupcolour	-0.1901	0.041	-0.0041	0.036
pop.groupasian	-0.3367	0.044	0.7416	0.040
pop.groupwhite	-0.3745	0.023	0.3730	0.026
prov2	-0.1041	0.022	-0.4316	0.020
prov3	0.2240	0.062	-0.4830	0.043
prov4	-0.8232	0.018	-0.5475	0.014
prov5	0.2431	0.023	-0.0246	0.019
prov6	0.6792	0.052	0.4503	0.041
prov7	0.0773	0.030	-0.0204	0.025
prov8	-0.2988	0.033	-0.2255	0.031
prov9	-0.1417	0.024	-0.0539	0.019
urban1	0.0549	0.050	0.1793	0.058
hh.size.L	0.4851	0.016	0.8804	0.013
hh.size.Q	-0.0478	0.025	-0.0197	0.022
hh.size.C	-0.1766	0.039	-0.3093	0.034
hh.size^4	-0.2055	0.047	-0.0096	0.040
hh.size^5	-0.0924	0.037	-0.0442	0.035
hh.size^6	-0.0618	0.044	0.1425	0.037
time.med.L	0.1776	0.040	0.6040	0.031
time.med.Q	-0.1165	0.044	-0.0951	0.034
time.med.C	0.0608	0.058	-0.1935	0.039
mom.ed.L	0.0970	0.052	0.1718	0.044
mom.ed.Q	-0.2393	0.053	0.0997	0.039
mom.ed.C	-0.0357	0.054	0.1429	0.040
mom.live.L	-0.1937	0.016	-0.1559	0.015
mom.ins.L	-0.4468	0.021	-0.1062	0.021
dad.ed.L	0.2119	0.036	0.2963	0.035
dad.ed.Q	0.0234	0.040	0.1382	0.036
dad.ed.C	0.1179	0.058	0.0895	0.049
dad.live.L	-0.1899	0.038	-0.0684	0.028
dad.ins.L	-0.4788	0.020	-0.5380	0.023
rd.age6+:age	0.1145	0.015	0.2063	0.017

Source: Authors' calculations. Ordered factors are estimated with orthogonal polynomial coding – ".L" for linear, ".Q" for quadratic, ".C" for cubic, etc. – to capture potential nonlinear relationships between the ordered factor and the dependent variable. With a binary ordered factor, such as 'mom.live', interpretation is the same as for an unordered factor, or dummy variable.

Table B.4: Nonparametric Bandwidth Summary for Model 2

Variable	Bandwidth	Scale Factor
hlthsk	0.08	1.15
rd.age	0.61	8.51
age	11742314.95	12595186.26
insure	0.07	0.94
inc	1079.78	2.86
pop.group	0.16	2.26
prov	0.07	0.95
urban	0.49	6.78
hh.size	0.17	2.42
time.med	0.16	2.29
mom.ed	0.28	3.91
mom.live	0.04	0.60
mom.ins	0.26	3.70
dad.ed	0.43	6.05
dad.live	0.18	2.54
dad.ins	0.91	12.69

Source: Authors' calculations.