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

Steven F. Koch† and Jeffrey S. Racine‡

September 21, 2015

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 10% – and that increase was drawn equally from home care and private care. On the other hand, in the nonparametric framework treatment effects were largest for large (and poor) families located farther from health facilities – approximately 5%. More plausibly, the positive treatment effect was drawn primarily from home care, suggesting that the policy favoured children living in poorer conditions, as those children received at least some minimum level of professional health care after the policy was implemented.

*The authors would like to thank the editor, the associate editor and two anonymous reviewers for their insightful comments and suggestions. 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 Economic Research Southern Africa for its financial support of this research, as well thanking 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 and Graduate Program in Statistics, McMaster University, *racinej@mcmaster.ca*; Department of Economics and Finance, La Trobe University; Info-Metrics Institute, American University; Rimini Center for Economic Analysis

1 Introduction

Applications of Thistlethwaite & Campbell's (1960) 'regression discontinuity design' (RD) methodology have become increasingly prevalent in economics and political science; 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 guides by Imbens & Lemieux (2008) and Eggers, Fowler, Hainmuller, Hall & Snyder, Jr. (2015) that facilitate its implementation. As highlighted in the aforementioned papers, 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, although the effects at various quantiles can also be considered; see Frandsen, Frölich & Melly (2012). However, there is little in these guides for researchers applying RD to multinomial, and, therefore, potentially interrelated, outcomes. In this paper, we provide one possible avenue for such analysis, focusing on an application within an unordered multinomial setting.

Briefly, RD is a pseudo-experimental design used to quantify the causal effect of an intervention in environments for which randomization is not feasible. In it simplest form, the design requires the policy intervention to be assigned, based on the crossing of a threshold, such that the intervention only affects the population on one side of the threshold. The variable that underscores the threshold is referred to as the running variable. A wide range of running variables have been used in the literature, including: class size limits (Angrist & Lavy 1999); election vote shares in the public sphere (Lee (2008), Caughey & Sekhon (2011), Eggers et al. (2015)); election vote shares in the private sector (Flammer forthcoming); student performance (van der Klaauw (2002), Ou (2010)); the duration of benefits (Caliendo, Tatsiramos & Uhlendorff 2013); geographic location (Grout, Jaeger & Plantinga 2011) and age (Carpenter & Dobkin (2009), Deza (2015)). In this paper, the intervention is the abolishment of health care user fees, and the intervention is in place only for children below the age threshold of six years, and only if those children seek treatment at public health care facilities in South Africa; the change was introduced by Nelson Mandela, when he assumed the Presidency in 1994.

Because the intervention is determined by the running variable, failure to include the running variable in the empirical analysis would result in omitted variable bias. Thus, early RD applications included both a policy indicator and separate polynomials of the running variable (above and below the policy threshold) to estimate the effect of the policy change on an outcome of interest (at the threshold). However, RD effects in the literature are implicitly one-dimensional. Is there an incumbency advantage in democratic elections (Lee (2008), Caughey & Sekhon (2011), Eggers et al. (2015))? Does alcohol consumption affect drug use (Deza 2015)? In some policy settings, like the one considered below, there exists a set of mutually exclusive (discrete and unordered) outcomes that can be affected by the policy. Health care services in South Africa are provided by both the private and public sectors, and, 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). Importantly, a policy intervention affecting access to the public sector will influence the relative price of access to the private sector and home care. In other words, any public sector impact must be offset across the remaining outcomes, and this feature (constraint) ought to be incorporated into the subsequent analysis.

RD effects on unordered multinomial data, as opposed to continuous, ordered or binary data, cannot be estimated via standard methods, although Coe & Zamarro (2011) create binary categories from their set of multinomial outcomes, and estimate separate linear probability models. If unordered multinomial data is re-categorized in this fashion, the researcher is implicitly accepting the Independence of Irrelevant Alternative (IIA) assumption, i.e., the assumption upon which the multinomial logit model is founded, without testing its validity. Instead, we argue that a natural extension to the typical RD structure, in this setting, is to model the probabilities of the categorical outcomes. One candidate probability model is the linear index multinomial logit (MNL) model, although such a model assumes IIA. If that assumption is invalid, MNL treatment effects estimates could be biased. Therefore, we also consider a nonparametric probability model that constructs the conditional probability directly. 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, whether binary or multinomial, ranked or unranked. Other parametric approaches could also be modelled, for example, an error components model would also relax IIA; however, an error components model cannot be estimated in our context, because we do not have access to panel data. Although we are not the first to adopt nonparametric methods within an RD context – see Hahn, Todd & van der Klaauw (2001), Imbens & Lemieux (2008), Carpenter & Dobkin (2009) and McCrary & Royer (2011) for examples, as well as Imbens & Kalyanaraman (2012) and Gelman & Imbens (2014) for practical guides to bandwidth selection – we are the first to extend the nonparametric analysis to multiple outcome models, to the best of our knowledge.

In this paper, we analyze the effect of user fee abolition on the use of health care services, and model the three (unordered) health care-seeking options (home care, private, and public) using parametric and nonparametric approaches. The nonparametric model is found to fit the data better than the MNL model both in- and out-of-sample in terms of its classification ability (i.e., in terms of matching actual with predicted choices). These results suggest that the linear index multinomial logit model is inappropriate in our setting, i.e., there is unobserved correlation across the health care-seeking options. 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. With the nonparametric model, we uncover treatment effects that are not constant. The least well-off have benefited from the policy, while those in better socio-economic circumstances have not. Furthermore, the least well-off have benefited because they are more likely to use public services than undertake home care, post user fee abolition. These results differ markedly from those recovered from the MNL model, which suggests instead that the effects were fairly constant and that the increase in public facility use was due to a reduction in home care and a reduction in private facility use.

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 revision in 1994 contained a number of components, including free public health care for ill children under the age of six, the elderly, pregnant women and nursing mothers. For further information about the policy, see Koch (2012) and Brink & Koch (2015). Given data limitations, our analysis focuses only on the effect of the policy on the demand for curative care services for young children. 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 policy change was designed to improve access to health care within the public sector, even though other health care-seeking options are available for ill children. These other options, such as care within the private sector and home care, are potential substitutes for public care. Therefore, the RD analysis is also 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 begin our analysis (we report results for the logit only, as both link functions deliver similar results). However, we also undertake nonparametric analysis based on direct estimation of conditional probabilities for the reasons outlined earlier. Each is described, in turn, below.

2.1 Parametric Multinomial Logit Analysis

Denote by Y_i, with realizations y_i, a categorical indicator of health facility choice for child i, 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. The X_i represent socio-economic and demographic characteristics of the ill child, including the child's age. Given the central role played by age in our analysis, we will postpone detailing our approach to modeling age for the moment until we first establish some notation. Following convention, we define p_{ij} to be the probability that ill child i receives treatment j, i.e., $p_{ij} = P(Y_i = j|X_i = x_i)$. 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 explanatory variables follow a linear index formulation within the logistic function, 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 2015) package 'nnet' (Venables & Ripley 2002, Version 7.3-9).

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

As an alternative model, we consider a consistent nonparametric estimator of the outcome 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, ordered, and unordered variables, Li & Racine's (2003) generalized product kernel is adopted in the estimation. Following Li & Racine (2003), let $X = (X^c, X^u, X^o)$ denote a split of X into t continuous, r unordered and s ordered variables. The marginal density m for realizations x is given by

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

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

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

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 (‘$\ell^u_y(\cdot)$’). For positive bandwidth $h_k > 0$,

$$W(X^c_{ik}, x^c_k) = \frac{1}{h_k} K\left(\frac{X^c_{ik} - x^c_k}{h_k}\right)$$
$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad z = \frac{X^c_{ik} - x^c_k}{h_k} \tag{11}$$

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

$$\ell^u(X^u_{ik}, x^u_k) = \begin{cases} 1 & \text{if } x^u_k = X^u_{ik} \\ \lambda_k & \text{if } x^u_k \neq X^u_{ik} \end{cases}, \tag{12}$$

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

$$\ell^u_y(Y^u_i, y^u) = \begin{cases} 1 - \lambda_0 & \text{if } y^u = Y^u_i \\ \lambda_0/2 & \text{if } y^u \neq Y^u_i \end{cases}. \tag{14}$$

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

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

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 2015) package ‘npRmpi’ (Racine & Hayfield 2014, Version 0.60-2) paired to ‘Rmpi’, (Yu 2014, Version 0.6-5); see also Hayfield & Racine (2008, Version 0.60-2) for additional information on the 'np' package.

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 2014).

The outcomes Y_i are mapped to a 3×1 vector Υ_i, one value for each of the three health care facility options:

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

The estimated model delivers a prediction vector, $\hat{\Upsilon}_i$, which is based on the predicted probabilities from the model:

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

where $\max_k\{\hat{p}_{ik}\}$ is the (conditional) mode, i.e., the most likely (highest probability) choice given the child's attributes. Given these predictions, we adopt a popular 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 $n \times 3$ 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 3×3 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

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

Observe that this approach implicitly uses the rule 'predict outcome i if the estimated probability of choice $i > 1/3$'. We apply this rule to both the parametric and nonparametric models but do not attempt to optimize the misslcassification rates separately for each model, which is a rather complex problem involving multi-class receiver operating characteristic analysis. However, we are confident that the ranking of models is not affected by this choice of cutoff given previous investigation in the much simpler binary choice setting.

Based on insights from Efron (1982), Racine & Parmeter (2014) 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 (2014).

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. Based on the estimated models (i.e., in the case of the nonparametric model hold smoothing fixed, and in the case of the multinomial logit model, hold the 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 5 and 6).

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. Because the analysis focuses on the effect of user fee abolition on young children, the relevant age threshold is 72 months (six years of age). However, we normalize that threshold to zero (subtract 72); thus, in what follows, the age threshold is zero. Given that the age of all children is net of the age of the threshold (72 months), policy eligible children will have negative ages, while ineligible children have non-negative ages. Rather than assuming fixed treatment effects, we analyze and present differences across quantiles of the explanatory variables (X_q, which is described in more detail, 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 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 $\forall q$.

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

In the preceding equation, a six month window below the threshold is used for eligible children; a similar six-month window is used above the threshold for non-eligible children. The six month window is in keeping with the RD context, wherein the policy effect should be constructed near the policy threshold. For the multinomial logit model, the $\hat{p}_{ij}$s are estimated via (5); for the nonparametric model, the $\hat{p}_{ij}$s are estimated via (8). Finally, confidence intervals for the average policy impact within a data quintile 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 $B = 1000$ times. This yields a 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

Data for the analysis is taken from the South African October Household Survey (OHS) of 1995 and the South African Income and Expenditure Survey (IES) from the same year.

The main purpose of the OHS, conducted by Statistics South Africa (1995*b*), is 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 a medical aid scheme (health insurance). The main purpose of the quinquennial IES (Statistics South Africa 1995*a*) is to collect expenditure data for use in calculating the consumer price index. As these surveys were given to the same households, they could be merged. Due to the large number of missing household earnings observations in the OHS, we select monthly household expenditure data from the IES.

Both surveys followed 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 randomly selected for interviews. After merging the two data sets, responses were available for 126,283 individuals living in 28,585 households. Our merging efforts match those of Pauw (2003), wherein 5,501 individuals were lost from 1,115 households that could not be merged across the surveys. Given our focus on health care-seeking behaviors for children, we further restrict the sample to children under the age of 14 years who experienced illness within the past month. Therefore, the sample includes only children potentially affected by the policy who are reasonably close to the age threshold. The resulting sample contains 2,556 such children, which constitutes approximately 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.

3.1 Data Description

One of the more important variables for any RD analysis is the running variable, the variable upon which the policy rule is founded. For this analysis, it is the age of the child. In the initial survey, data on age was only available in years. Fortunately, the OHS contains a separate births module, allowing us to match exact birth dates to children living with their mothers. Since there are only $2,556$ children in the estimation sample, but there are nearly 4000 potential birthdays for the 14-year age range of the children, we chose not to use exact birth dates but rather to use month of birth. There are 168 months in the 14-year age range. On average, there are approximately 15 children in any particular month. Year-of-birth data estimates were also generated, yielding similar results to those reported below.

Age can be modelled in any number of ways in this RD context; Gelman & Imbens (2014), however, suggest that a linear or quadratic polynomial with threshold breaks is most appropriate; we follow their suggestion. Thus, in the multinomial logit model, we allow for quadratic functions in age that could be completely different on either side of the threshold. In other words, after normalizing the age axis to 0 at 72 months as described above, we include a linear term, a quadratic term, a policy threshold dummy, and we interact the dummy with both the linear and quadratic terms. We also estimated a model that included only the linear term, the dummy and the interaction; the results from that model do not differ from what is reported below. Within the nonparametric models, we include only age and the threshold dummy, because the bandwidth for age (i.e., the amount of local averaging) determines the resulting relationship, while the joint distribution function (f) in (7) is general and allows for interactions between all variables in the model, hence including interactions would be inappropriate.

We focus on health care-seeking behavior and make use of the following 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 at either a public or private facility (home care). In addition to the child's age and household income (expenditure, strictly speaking), the surveys cover a variety of subjects that were also linked to each child, including information about the child's mother and father (their years of completed schooling, their access to health insurance, and whether or not the child's father is alive). For children whose father is no longer living, their father's education and access to health insurance is coded to zero. Additional information used included the race of the child (South Africa had just recently walked away from its Apartheid past in 1994 such that race is an important predictor of socio-economic status), the size of the household, the location of the household (province, urban/rural) and distance from the household to the nearest health facility, measured in minutes.

The means of the variables included in the analysis are presented in Tables B.1 and B.2 in Appendix B. Table B.1 represents the policy-eligibles, i.e., all ill or injured children under the age of 6 years, while Table B.2 contains information on the older children (6-14 years).

3.2 RD Validity

In order for (21) to represent the policy impact for choice j in quantile q, the predicted probabilities must be consistently estimated on either side of the age threshold. Consistency relies heavily on the validity of the RD design. Underlying RD is a series of assumptions primarily related to smoothness. Lee (2008) describes these assumptions in Condition 1c and Condition 2b. Condition 1c requires the expectation of the outcome, as a function of the running variable (the variable that determines policy eligibility, which is age in this analysis), to be continuous everywhere other than at the threshold. Condition 2b requires the cumulative distribution of the running variable, conditional on all unobserved determinants, to be differentiable over its support, while the density of the running variable is positive at the threshold.

With respect to Condition 1c, a discontinuity should be discernible at the threshold for the outcome variables. We provide an illustration of the potential for outcome discontinuity based on the mean use of public care (see Figure 1), private care (see Figure 2) and home care (see Figure 3). Within these figures, we include (a) the average 'attendance' by children of each age (months below/above the threshold) at that facility and (b) fitted local linear regressions estimated separately above and below the threshold. Average attendance is calculated for each age, and is the total number of children seeking public care, private care or home care divided by the total number in each age cohort. The illustrations suggest that, on average, there are discontinuities with respect to public care and private care. Below, we will see that these averages mask a more nuanced response to the policy intervention.

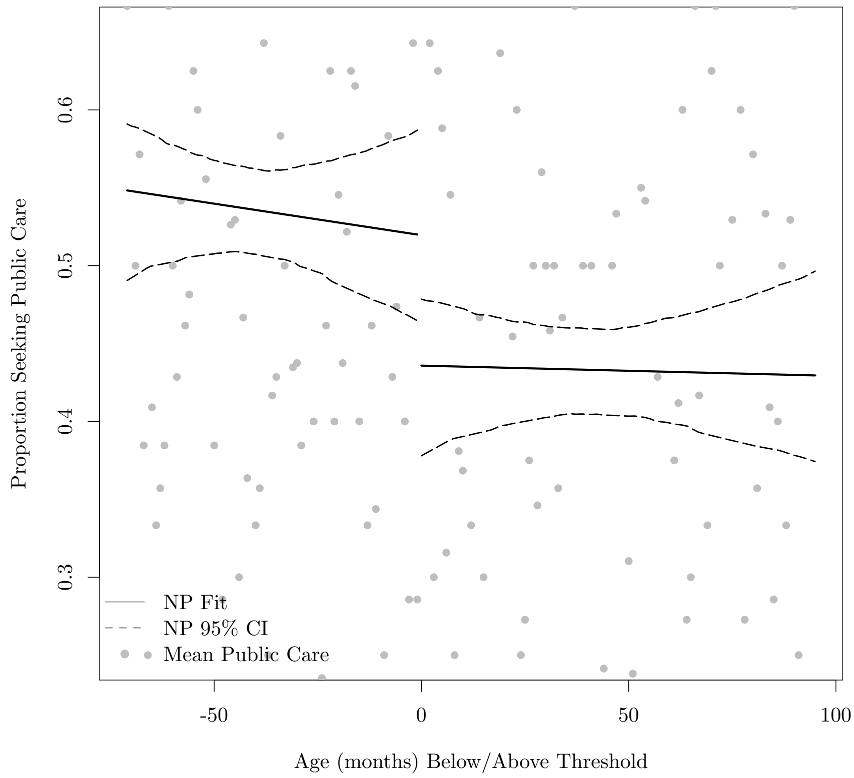

Figure 1: Is public care usage discontinuous? This figure presents fits and 95% confidence intervals from nonparametric local linear regression of children seeking public care (0/1) on age in months (threshold normalized to 0). Estimation performed separately on either side of the threshold. Optimal bandwidths derived from least-squares cross-validation: $\hat{h}_a = 6.34 \times 10^8$ is the bandwidth above the threshold and $\hat{h}_b = 4.48 \times 10^7$, below the threshold.

McCrary (2008) refers to violations of Condition 2b as 'manipulation of the running variable', and suggests a test. Manipulation could arise in this analysis if, for example, children just slightly above the age of six were passed off at the public facility as being under the age of six. However, if that were happening, it would be a mistake at the facility, as opposed to something that a caregiver could guarantee; therefore, it is not expected to be a significant source of manipulation. Relatedly, children under six could be more likely to be taken to a health facility to learn if they are ill, since they could receive free health care at a public facility. If such an anticipation effect was in the

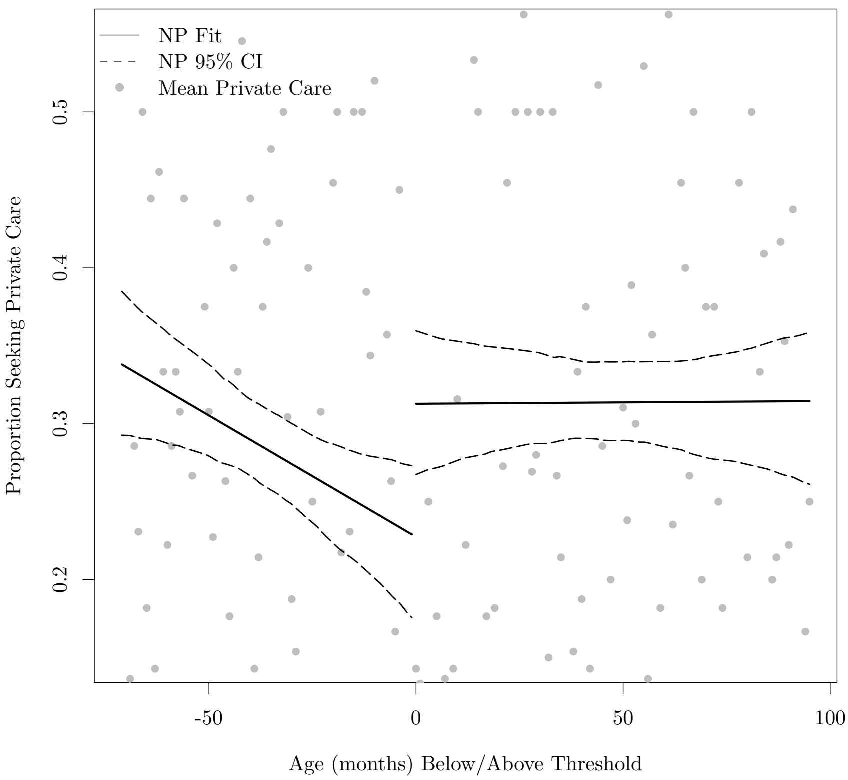

Figure 2: Is private care usage discontinuous? This figure presents fits and 95% confidence intervals from nonparametric local linear regression of children seeking private care (0/1) on age in months (threshold normalized to 0). Estimation performed separately on either side of the threshold. Optimal bandwidths derived from least-squares cross-validation: $\hat{h}_a = 2.59 \times 10^9$ is the bandwidth above the threshold and $\hat{h}_b = 2.13 \times 10^9$, below the threshold.

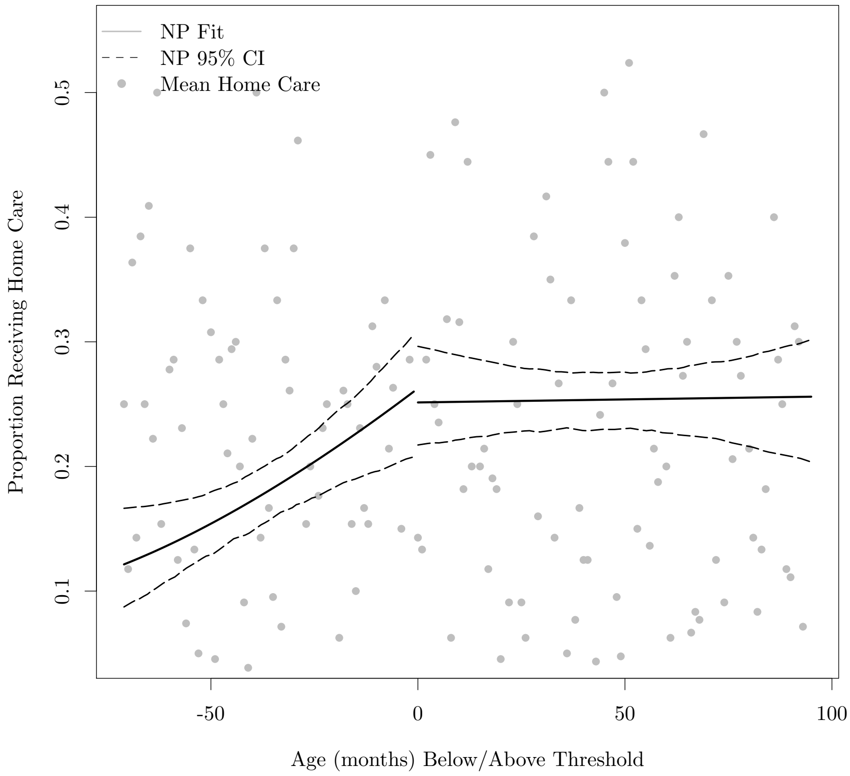

Figure 3: Is home care usage discontinuous? This figure presents fits and 95% confidence intervals from nonparametric local linear regression of children receiving home care (0/1) on age in months (threshold normalized to 0). Estimation performed separately on either side of the threshold. Optimal bandwidths derived from least-squares cross-validation: $\hat{h}_a = 3.44 \times 10^7$ is the bandwidth above the threshold and $\hat{h}_b = 36.3$, below the threshold.

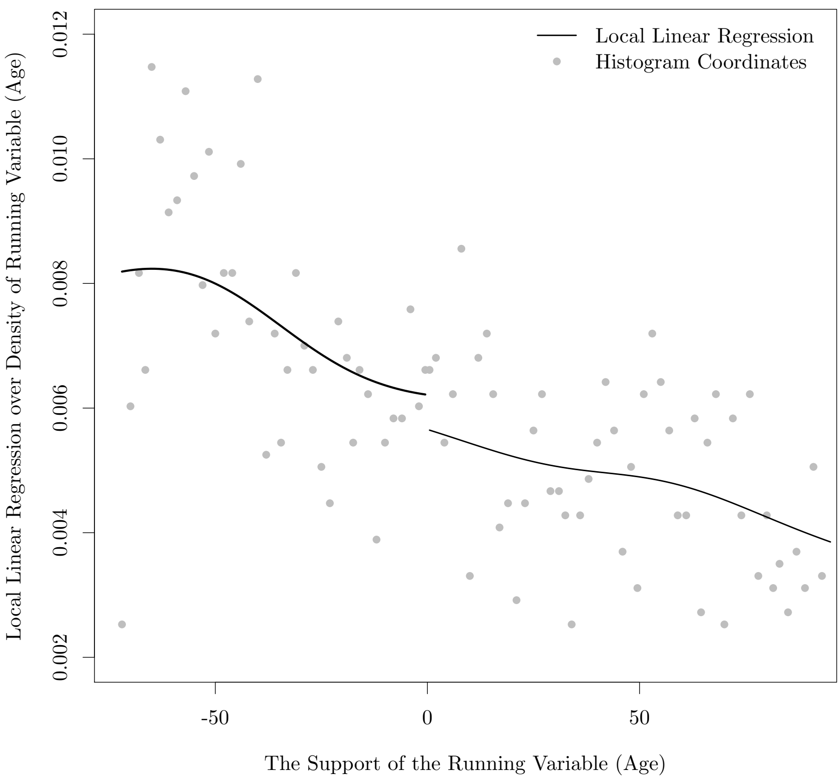

Figure 4: Is the running variable discontinuous in the analysis sample? Figure follows McCrary (2008). Included is the fitted local linear regression and coordinate points from the histogram, where the $x-$coordinate is the histogram midpoint, and the $y-$coordinate is the histogram height. The estimated discontinuity is $\hat{\theta} = -0.372$, its standard error $\hat{\sigma} = 0.92$, the $p-$value$= 0.69$ from a $t-$test of discontinuity significance, histogram binsizes $\hat{b} = 1.89$ and local linear regression bandwidths $\hat{h} = 18.86$.

data, but not properly addressed, estimated policy impacts would likely be overstated. To check for possible manipulation, we apply McCrary's (2008) running variable density test (see Figure 4) and find no evidence that the running variable has been 'manipulated' ($P = 0.69$). Although we do not find direct evidence of running variable manipulation, additional illustrative 'tests' are conducted. The key to Condition 2c is that, conditional on other control variables, the density of the running variable should not be discontinuous. In other words, we should not see discontinuities in other variables. Therefore, we examine a number of variables from the model to see whether or not there is evidence of a threshold discontinuity. Conditioning variables considered include: an indicator for whether or not the child is covered by private health insurance, household expenditure, an indicator for whether or not the child's mother or father is covered by private health insurance, and an indicator for whether or not a child lives in a rural area along with the mother's and father's education. The illustrations are available in Figures A.1 – A.7. The illustrations provide further support that the running variable is not being manipulated in-sample.

However, one worry remains. Might there be manipulation out-of-sample, i.e., has our analysis sample resulted in selection, due to manipulation in the McCrary sense? For example, the policy might have impacted preventive care, reducing reports of illness among those eligible for public health care without user fees. If a prophylactic effect of this nature (even though good for public health) differentially impacted the resulting choice of health care facility, our estimated results would be biased. One might also worry that the policy could encourage medical insurance adjustments; parents might chose to lower or even eliminate health insurance coverage for eligible children. Finally, any manipulation such as the above could be related to the level of education of the parents, whether they are covered by health insurance and/or the severity of the child's illness. Unfortunately, no information was gathered during the survey that would shed light on the severity of the child's illness. However, we can use the non-filtered dataset (i.e., include all children up to the age of 14 and not just the ill and injured children) to shed some light on these concerns; see Figure A.8 and Figure A.9.

In order to test for potential prophylactic effects, we undertake a nonparametric analysis of illness in the full sample. Specifically, we examine the proportion of ill children in the full sample, to see if there is any evidence that reported illness discontinuously decreases at the threshold; our illustration does not suggest a statistically significant decrease. To see if private health insurance might have been manipulated, we apply McCrary's (2008) density test to the subsample of all children (up to the age of 14) with access to private health insurance (whether or not those children have been reported to be ill or injured in the past 30 days). See Figure A.9 for the results. There is no evidence of private health insurance manipulation in the full sample, which agrees with our finding in the analysis sample (see Figure A.1). Thus, we are fairly confident that our analysis sample is not unduly influenced by manipulation either in- or out-of-sample.

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 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 Facility'), variables included in the initial analysis are limited to controls for household expenditure ('(ln) HH Expenditure') and its square ('Squared Expenditure'), a binary indicator of access to health insurance ('Child Insured'), and our function of the running variable ('Child Age', 'Age Squared', 'Child Eligible', 'Child Age x Eligible' and 'Age Squared x Eligible'). In the nonparametric model, as noted previously, only child age and the binary indicator are included ('Child Age' and 'Child Eligible'), because the interaction effect, if there is one, is determined by the data, while the functional form for age is also determined by the data. As parameter estimates are not the focus of the analysis, summaries of the multinomial logit estimates have been relegated to Table C.1 in Appendix C, while summaries of the nonparametric bandwidths and their scale factors have been relegated to Table C.2 in Appendix C.

However, the multinomial logit parameter estimates 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 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 on the basis of a limited number of variables. 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. We find that the nonparametric overall CCR is 0.566 (with log-likelihood, −2497.54), while the parametric overall CCR is 0.569 (with log-likelihood, −2461.33).

Table 1: Parametric Confusion Matrix for Model 1

Actual Facility	Predicted Facility Choices		
Choices	Home Care	Public	Private
Home Care	0	426	121
Public	0	1076	176
Private	0	384	388

Source: 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	Predicted Facility Choices		
Choices	Home Care	Public	Private
Home Care	0	437	110
Public	0	1109	143
Private	0	426	346

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

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, if the nonparametric model is ‘overfit’ (if the data-driven bandwidths are unreasonably small, i.e., the model is undersmoothed), the nonparametric model would predict well in-sample, but perform poorly out-of-sample. 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. 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 prior to splitting the data ensuring that the resulting evaluation and training data set contain mutually exclusive records (if by chance the nonparametric model has placed too much weight on duplicate observations, the aforementioned correction will uncover the problem). The out-of-sample model comparison is illustrated in Figure 5 and we note that results were insensitive to the removal of duplicate observations in the training resamples. As seen in the figure, at least for the 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; thus, poor predictive performance might not be surprising. In what follows, a number of explanatory variables are added to the model. These include: population group (‘Black Child’, ‘Coloured Child’ and ‘Asian Child’, white children are the left out category); province (‘Northern Cape’, ‘Western Cape’, ‘Eastern Cape’, ‘Free State’, ‘KwaZulu-Natal’, ‘Northwest’, ‘Gauteng’, ‘Mpumalanga’, Limpopo is the base province); household size (‘HH Size (8-9)’, ‘7’, ‘6’, ‘5’, ‘4’, ‘<4’, more than 10 household members is the left out category); distance to nearest medical facility (‘15 min < Facility Distance’, ‘ min < Facility Distance < 30 min’, ‘Facility Distance < 60 min, more than 60 minutes is the base); an urban-rural dummy (‘Urban Household’); mother’s education (‘Mother: Some Schooling’, ‘Primary School’, ‘Secondary School’, no schooling is the base); an indicator for mother’s and father’s health insurance coverage (‘Mother: Insured’ and ‘Father: Insured’), father’s education (defined per mother’s education); and

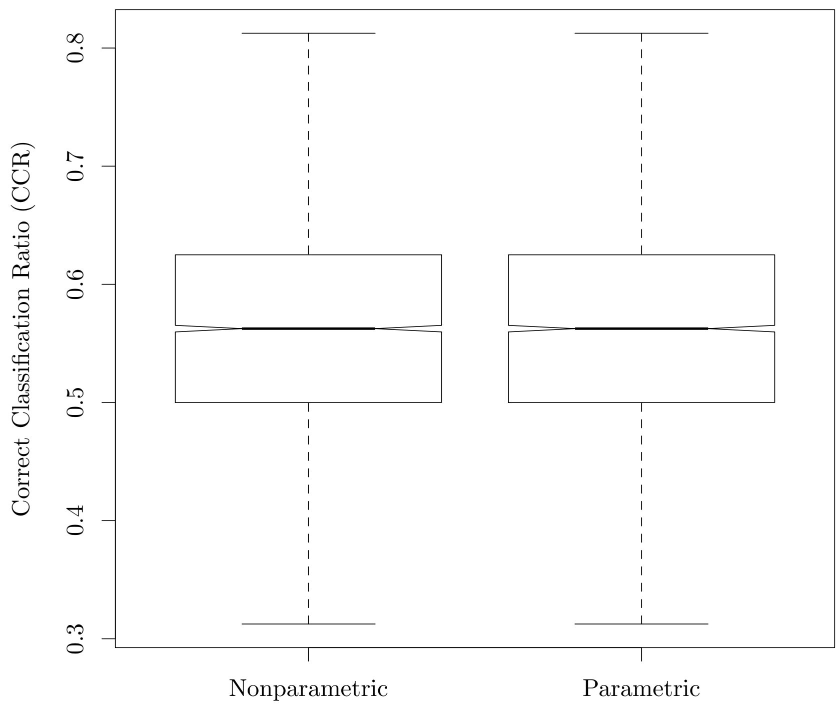

Figure 5: Boxplots for out-of-sample performance assessment of the benchmark model (mean nonparametric CCR: 0.5672, mean parametric CCR: 0.57, median nonparametric CCR: 0.562, median parametric CCR: 0.562, 5000 splits of the data, training data size $n_1 = 2,555$, evaluation data size $n_2 = 16$, higher CCR is better). The test for revealed performance under the null of equal performance delivers a P-value of 0.897 indicating that there is no significant difference in the predictive abilities of either model.

an indicator for whether the father is alive ('Father Alive');

As before, the focus of the analysis is not on parameters, so we relegate the multinomial logit estimates to Table C.3 in Appendix C and report nonparametric bandwidths and scale factors in Table C.4 in Appendix C. Briefly, a 'scale factor' is a unit free quantity that indicates the relative amount of smoothing used for each variable and can be compared within each variable type (i.e., continuous or discrete). 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 controls are also of importance in explaining facility choice. 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, 'Child Eligible', 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 additional included variables are statistically significant in the parametric model, one would expect the predictive performance of the multinomial logit model to improve as additional variables are added. 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 0.585, which is a rather small improvement. Meanwhile, the nonparametric model's CCR increases rather substantially; the nonparametric overall CCR is 0.857. For the full sample of data, the nonparametric log-likelihood was -1249.27, while the parametric log-likelihood was -2367.13. 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 severity of child illness.

Although the nonparametric model exhibits superior in-sample performance, it is possible that this simply reflects undersmoothing for this specific analysis. Therefore, the same performance comparison that was outlined above on the restricted set of variables is also undertaken here. The results of the training exercise are illustrated in Figure 6. 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

Table 3: Parametric Confusion Matrix for Model 2

Actual Facility Choices	Predicted Facility Choices Home Care	Public	Private
Home Care	46	381	120
Public	34	1055	163
Private	25	345	402

Source: 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 Home Care	Public	Private
Home Care	347	153	47
Public	7	1208	37
Private	9	114	649

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

(mean nonparametric CCR: 0.624) over the parametric model (mean parametric CCR: 0.569) is statistically significant at any conventional level ($P = 5.4 \times 10^{-112}$, 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.

5 Evaluation of Policy

Having assessed the improvement in performance of the nonparametric model relative to the parametric model, we 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 based on income;

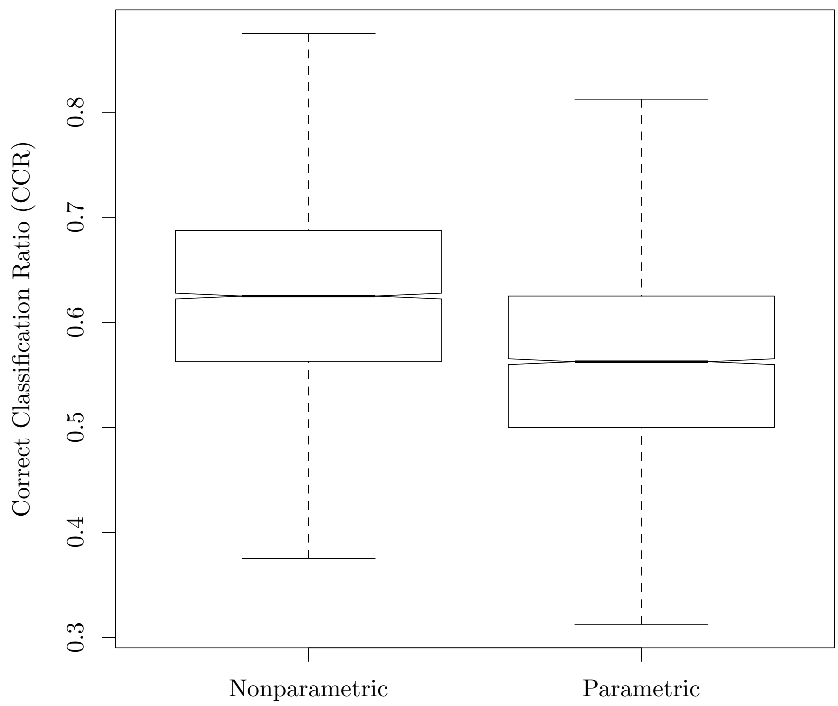

Figure 6: Boxplots for out-of-sample performance assessment of Model 2 (mean nonparametric CCR: 0.624, mean parametric CCR: 0.569, median nonparametric CCR: 0.625, median parametric CCR: 0.562, 5000 splits of the data, training data size $n_1 = 2,555$, evaluation data size $n_2 = 16$, higher CCR is better). The test for revealed performance under the null of equal performance delivers a P-value of 0 indicating that there is a highly significant improvement in the predictive abilities of the nonparametric model over the parametric model.

instead it includes household size and the dwelling's distance from the health facility. In the analysis we treat province, race and health insurance access as fixed; we set province to KwaZulu-Natal, an oversampled province in the data; we set race to black, who are poorer on average, due to South Africa's historical policies; further, we turn off the health insurance indicator. Additionally, we assume the characteristics of the median household for all of the remaining independent variables, barring household size and health facility distance. Specifically, both household size and time to facility are ordered categorical variables. We define 'poor' to be those households that are larger and are located farther away from health facilities. Thus, 'better' means households that are smaller and live closer to health care facilities. Finally, in order to calculate the treatment effect as the difference across the policy (RD) threshold, the age of young children is set below the threshold (we sample from a six month age window below the threshold and above the threshold; see (21)). The results are illustrated in Figures 7 and 8, and presented across quantiles.

The figures illustrate the parametric and nonparametric treatment effects across data quantiles, along with 90% confidence bands. The parametric public sector treatment effects are fairly constant as well-being (i.e., the quantile q) increases, averaging roughly 10%. Curiously, according to the parametric model, children living in the best of circumstances receive at least as much benefit as do less well-off children. On the other hand, the nonparametric public sector treatment effects paint a more plausible picture with respect to equity considerations; user fee abolition increased the use of public health care facilities among the least well-off young children by up to 5%, whereas this effect is entirely eliminated for ill children in the upper-half of the well-being distribution (i.e., above the median q).

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. With respect to substitution, the parametric model suggests that children from poorer households are similarly likely to switch out of either home care or private care in order to receive their health care from the public sector. At the upper end, according to the parametric model effects, ill or injured children are much more likely to have had private care substituted for public care. In other words, within the context of a parametric model, the user fee policy is found to primarily affect the 'ownership' of health care facility. Public facility usage is found to increase, at the cost of decreases in private sector facility usage. 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 care among 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 the policy was not necessarily

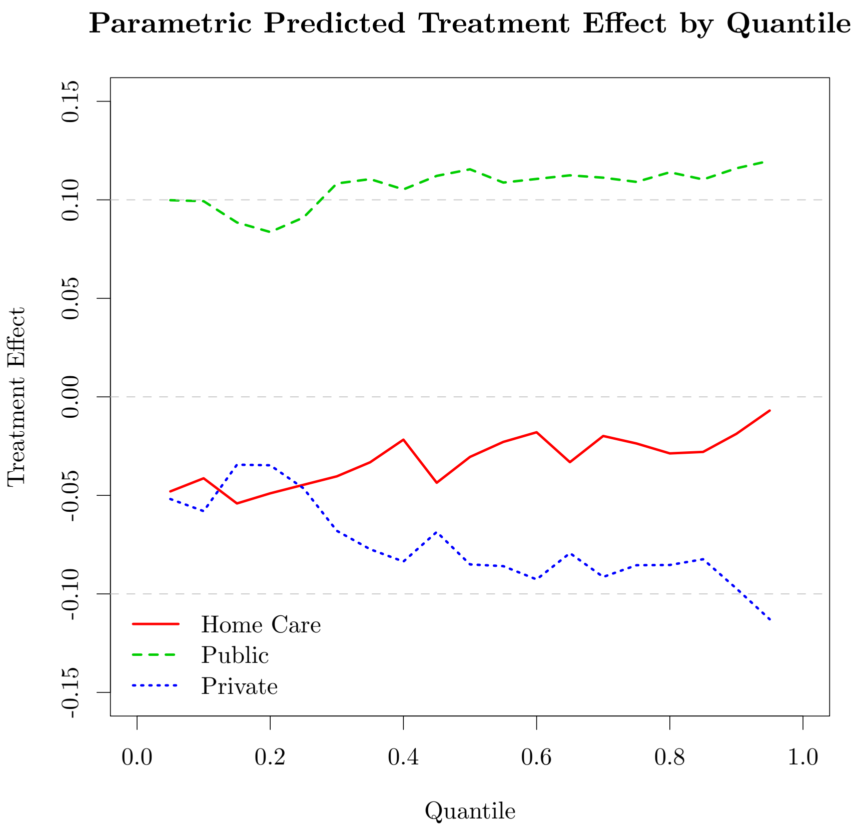

Figure 7: Estimated treatment effects of user fee abolition on all health facility choices made for ill children. Analysis undertaken across quantiles (0.05-0.95) of the data for black children living in KwaZulu-Natal without health insurance. Treatment effects calculated from the multinomial logit model. Moving from lower to higher quantiles implies an improvement in living standards.

designed to benefit.

Within the nonparametric setting, the substitution patterns are more interesting and somewhat more reasonable from a 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 in public care usage is drawn 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 limited, while the abolition of fees continues to be associated with an increase

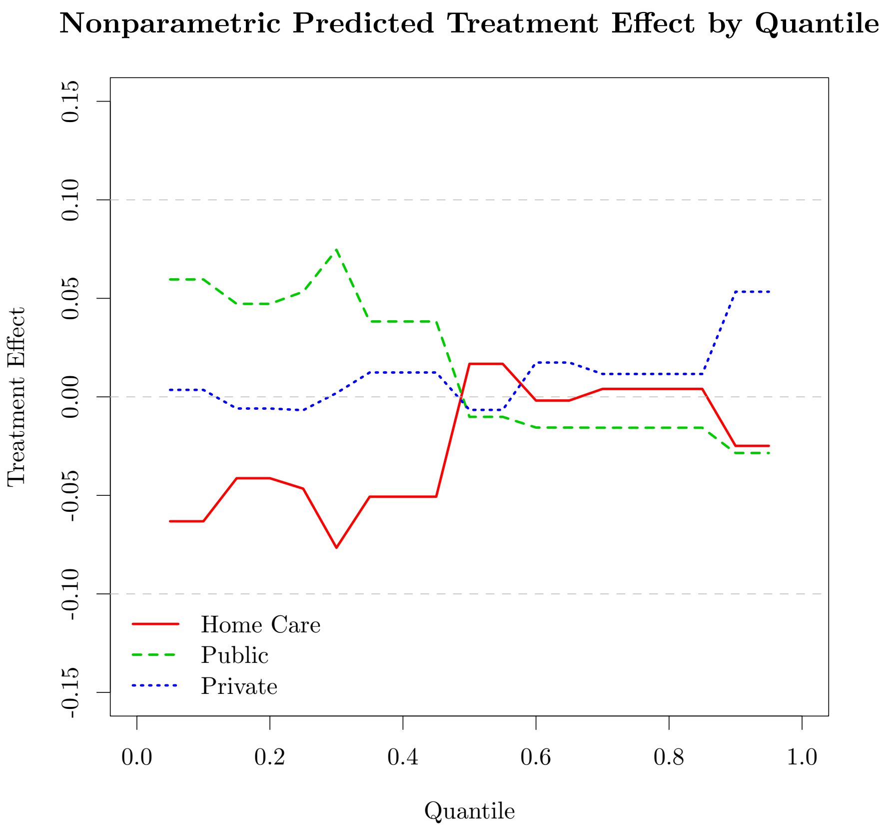

Figure 8: Estimated treatment effects of user fee abolition on all health facility choices made for ill children. Analysis undertaken across quantiles (0.05-0.95) of the data for black children living in KwaZulu-Natal without health insurance. Treatment effects calculated from the nonparametric conditional probability model. Moving from lower to higher quantiles implies an improvement in living standards.

in access to public health care drawn from those most likely to not have received any health care (i.e., home care). Moving beyond the median, little evidence of a policy impact is uncovered.

It should be noted, again, that there are a large number of treatment effects that could be calculated, based on the underlying values of the independent variables. Therefore, the results presented in the preceding figures and discussions apply to the values used. Thus, in other provinces, for example, the estimated effects could differ from what is reported here. It could also be that responses might differ by insurance status, even though the policy was not meant to impact the insured. We view possible differences in effects across regions or groups of people as important for understanding policy, rather than as a critique of the nonparametric analysis.

In addition to noting that we have reported on a limited set of results, it should be noted that our analysis focuses only on the effect of user fee abolition on curative care services for children under the age of six. This is a result of data limitations that preclude the consideration of preventative care, antenatal care or effects related to nursing mothers. Furthermore, a number of other changes related to South African pensions were enacted within a similar time frame. Thus, it was not possible to consider the effect of the policy on the elderly.

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 near the policy threshold age. Although the multinomial logit treatment effects (on public care receipt) are in the neighborhood of 10%, increasing slightly along with living standards, the nonparametric treatment effects are generally smaller and disappear entirely for children living in better circumstances. For the parametric model, we find that the increase in public care is driven primarily by reductions in private care; in other words, children appear to be substituted from private facilities into public facilities. For the nonparametric model, we find a starkly different substitution pattern: user fee abolition is found to increase access to health care, overall. Home care is less likely among the eligible children.

The differing impact uncovered by the two models suggest slightly different interpretations, even though both sets of results suggest that the policy affected welfare. In the parametric model, the welfare effect is uncovered through the reduction in more expensive private care towards less expensive public care. Within the nonparametric setting, on the other hand, the welfare effect uncovered is expected to be pro-poor; at the very least, the policy itself is found to have increased overall access to health care for the least fortunate. The nonparametric results reinforce 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.

In addition to uncovering differences in treatment effects by model structure, we found that the parametric model does not fare as well as the nonparametric model in terms of predicting outcomes, both in- and out-of-sample. Out-of-sample performance favoured the nonparametric model by a statistically significant margin. The statistical differences in model performance are expected to have arisen from the underlying differences in model assumptions. Explicitly, MNL is underpinned by IIA, which is not the case for Hall et al.'s (2004) nonparametric estimator. Thus, performance differences could be due to unobserved correlation between the outcomes. In this particular empirical problem, the MNL did not appear to be able to discriminate home care from the other care options. Possibly, that is because home care is quite different from professional care, and a model that nested the two professional care options would perform better. Undertaking a detailed examination of the exact source(s) of the observed differences between the parametric and nonparametric approaches is left for future research. However, the observed differences in both model performance and treatment effects suggest that future RD research in multinomial settings should pay particular attention to parametric model specification.

References

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

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.

Brink, A. S. & Koch, S. F. (2015), 'Did primary health care user fee abolition matter? reconsidering south africa's experience', *Development Southern Africa* **32**(2), 170–192.

Caliendo, M., Tatsiramos, K. & Uhlendorff, A. (2013), 'Benefit duration, unemployment duration and job match quality: A regression discontinuity approach', *Journal of Applied Econometrics* **28**, 604–627.

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.

Caughey, D. & Sekhon, J. S. (2011), 'Elections and the regression discontinuity design: Lessons from close U.S. house races, 1942-2008', *Political Analysis* **19**(4), 385–408.

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

Deza, M. (2015), 'The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997', *Health Economics* **24**(4), 419–438.

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.

Eggers, A. C., Fowler, A., Hainmuller, J., Hall, A. B. & Snyder, Jr., J. M. (2015), 'On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races', *American Journal of Political Science* **59**(1), 259–274.

Flammer, C. (forthcoming), 'Does corporate social responsibility lead to superior financial performance? a regression discontinuity approach', *Management Science* .
URL: *http://dx.doi.org/10.1287/mnsc.2014.2038*

Frandsen, B. R., Frölich, M. & Melly, B. (2012), 'Quantile treatment effects in the regression discontinuity design', *Journal of Econometrics* **168**(2), 382–395.

Gelman, A. & Imbens, G. W. (2014), 'Why higher-order polynomials should not be used in regression discontinuity designs', NBER Working Paper No 20405.

Grout, C. A., Jaeger, W. K. & Plantinga, A. J. (2011), 'Land-use regulations and property values in Portland, Oregon: A regression discontinuity design approach', *Regional Science and Urban Economics* **41**(2), 98–107.

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. R package version 0.60-2.

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.

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

Lee, D. S. (2008), 'Randomized experiments from non-random selection in U.S. House elections', *Journal of Econometrics* **142**(2), 675–697.

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

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.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics* **142**, 698–714.

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.

Pauw, K. (2003), Creating a 1995 OHS and combined OHS-IES data in STATA, Technical report. Technical Paper Series 15622, PROVIDE Project.
URL: *http://EconPapers.repec.org/RePEc:ags:provtp:15622*

R Core Team (2015), *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. & Hayfield, T. (2014), *npRmpi: Parallel nonparametric kernel smoothing methods for mixed data types*. R package version 0.60-2.
URL: *http://CRAN.R-project.org/package=npRmpi*

Racine, J. S. & Parmeter, C. F. (2014), 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.

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

Statistics South Africa (1995*a*), *Income and Expenditure Survey (South Africa), 1995*, Statistics South Africa (producer) and South African Data Archive (distributor), Pretoria.

Statistics South Africa (1995*b*), *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. (2002), 'Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach', *International Economic Review* **43**(4), 1249–1287.

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.

Yu, H. (2014), *Rmpi: Interface (Wrapper) to MPI (Message-Passing Interface)*. R package version 0.6-5.
URL: *http://CRAN.R-project.org/package=Rmpi*

A RD Validity Redux

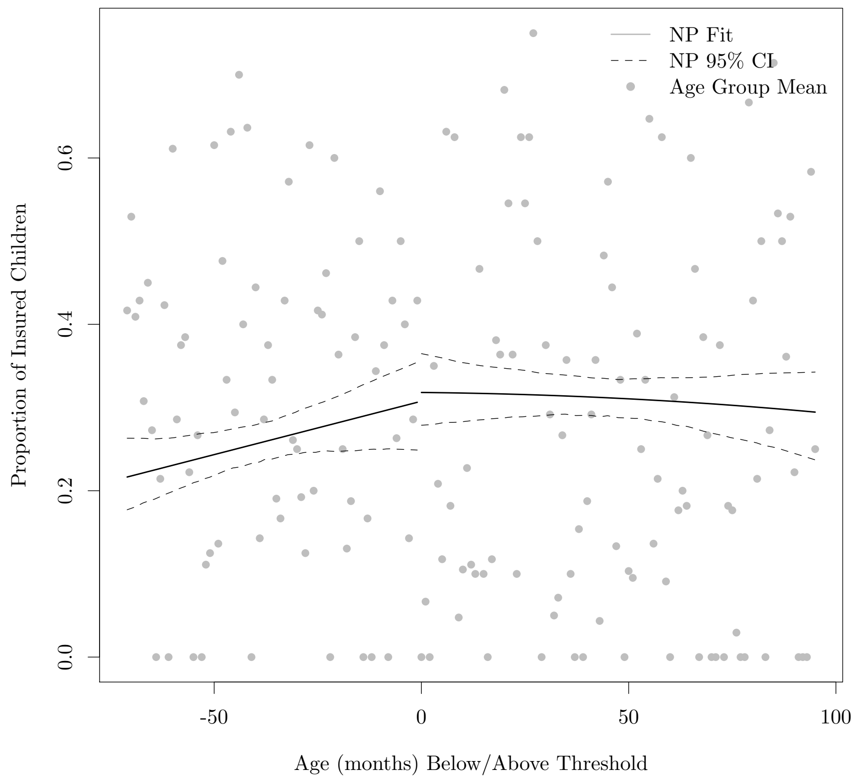

Figure A.1: Discontinuity in child's access to private insurance? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's access to private health insurance (0/1) against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 2.36 \times 10^7$) and above the threshold ($\hat{h}_a = 77.7$).

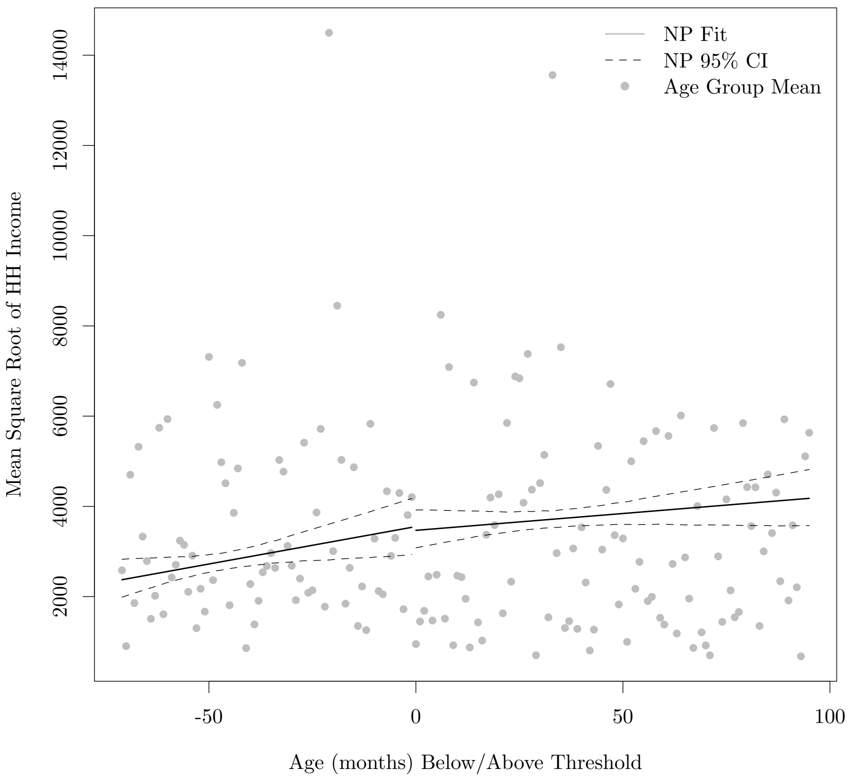

Figure A.2: Discontinuity in household expenditure? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's access to health insurance against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 1.12 \times 10^8$) and above the threshold ($\hat{h}_a = 2.8 \times 10^7$).

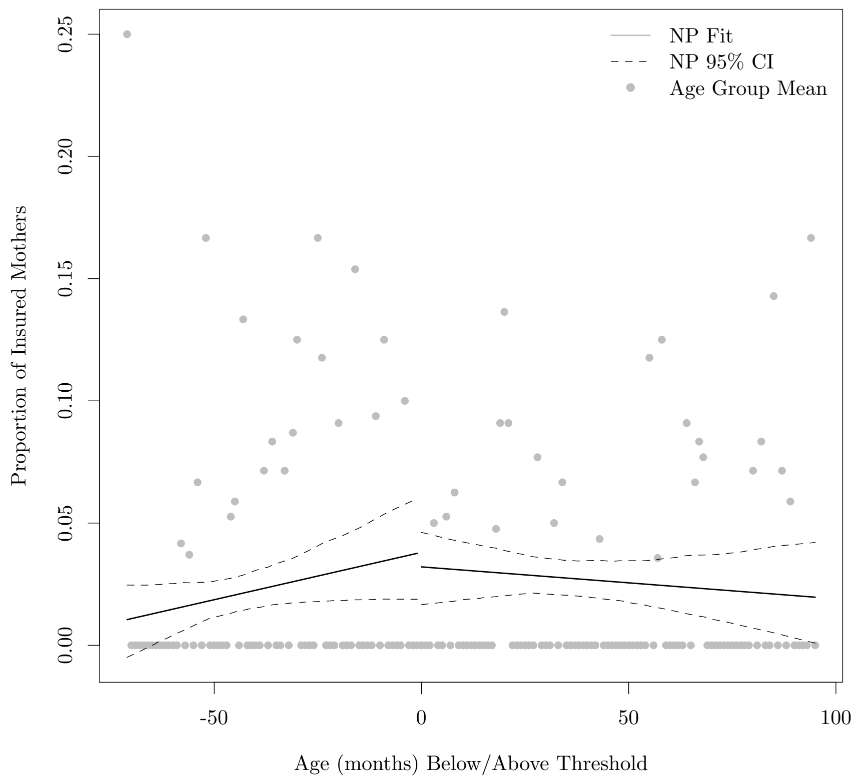

Figure A.3: Discontinuity in child's mother's access to private health insurance? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's mother's access to health insurance (0/1) against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 1.1 \times 10^8$) and above the threshold ($\hat{h}_a = 1.74 \times 10^7$).

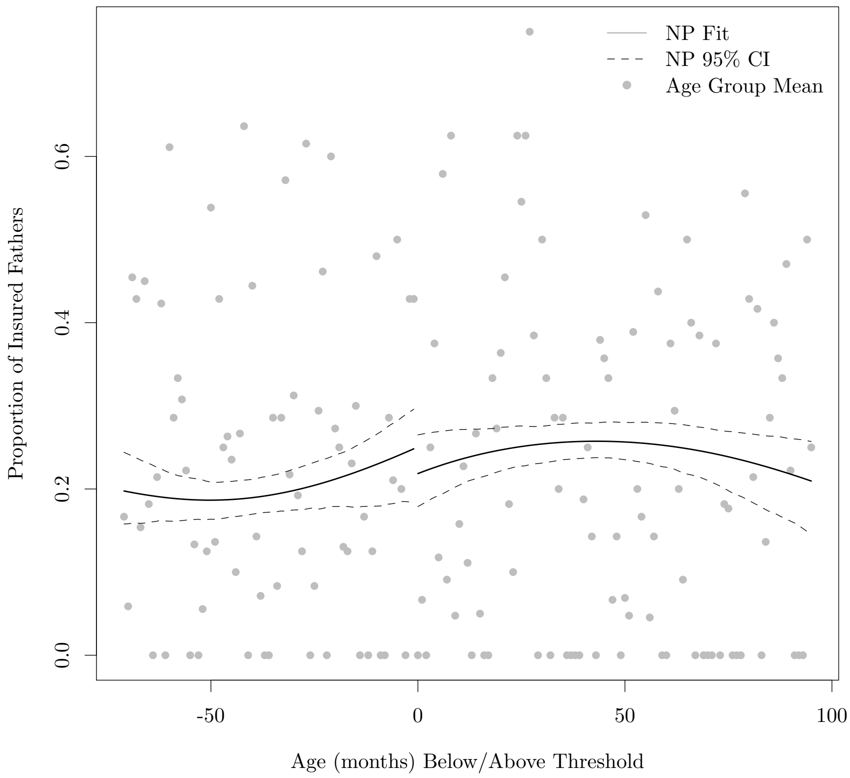

Figure A.4: Discontinuity in child's father's access to private health insurance? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's father's access to health insurance (0/1) against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 23.9$) and above the threshold ($\hat{h}_a = 28.4$).

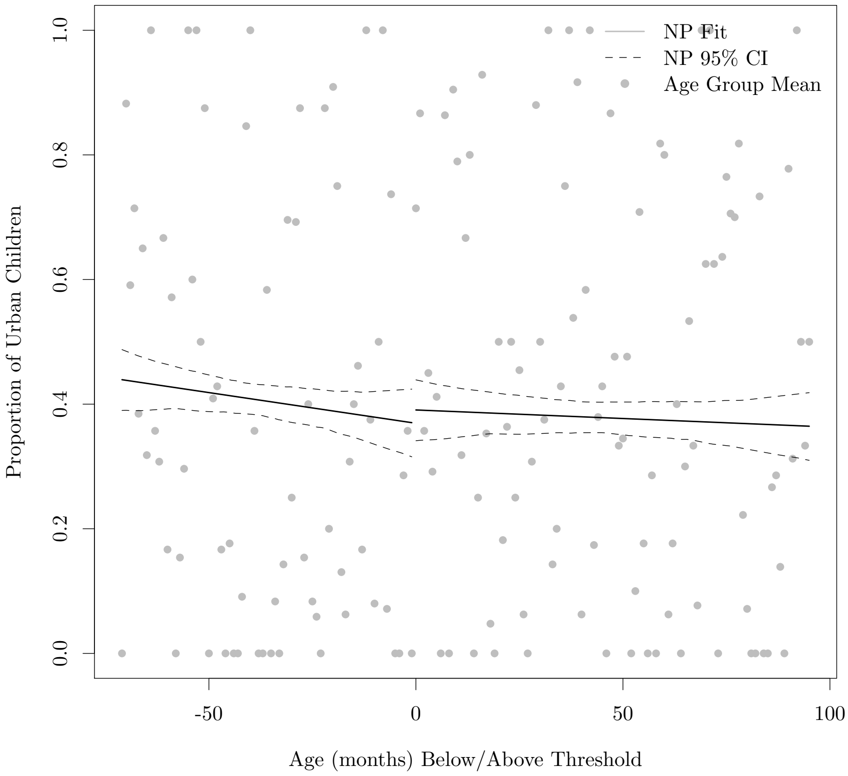

Figure A.5: Discontinuity in rural living? Illustration contains fits and 95% confidence intervals from nonparametric regressions of an urban child's access to health insurance (0/1) against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 5.68 \times 10^8$) and above the threshold ($\hat{h}_a = 1.35 \times 10^9$).

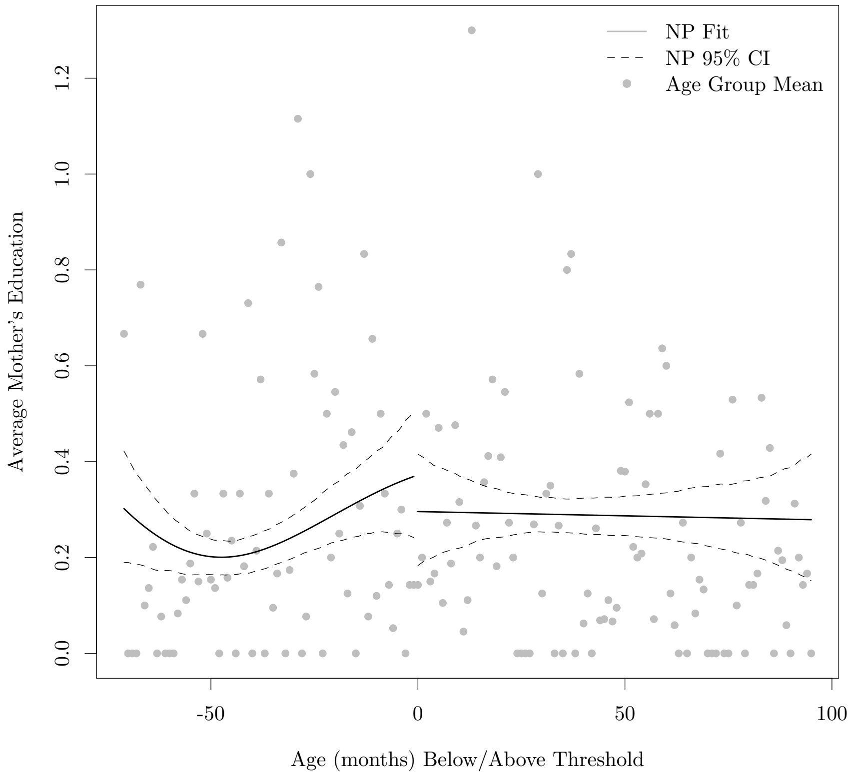

Figure A.6: Discontinuity in mother's education? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's access to health insurance against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 14.8$) and above the threshold ($\hat{h}_a = 4.3 \times 10^7$).

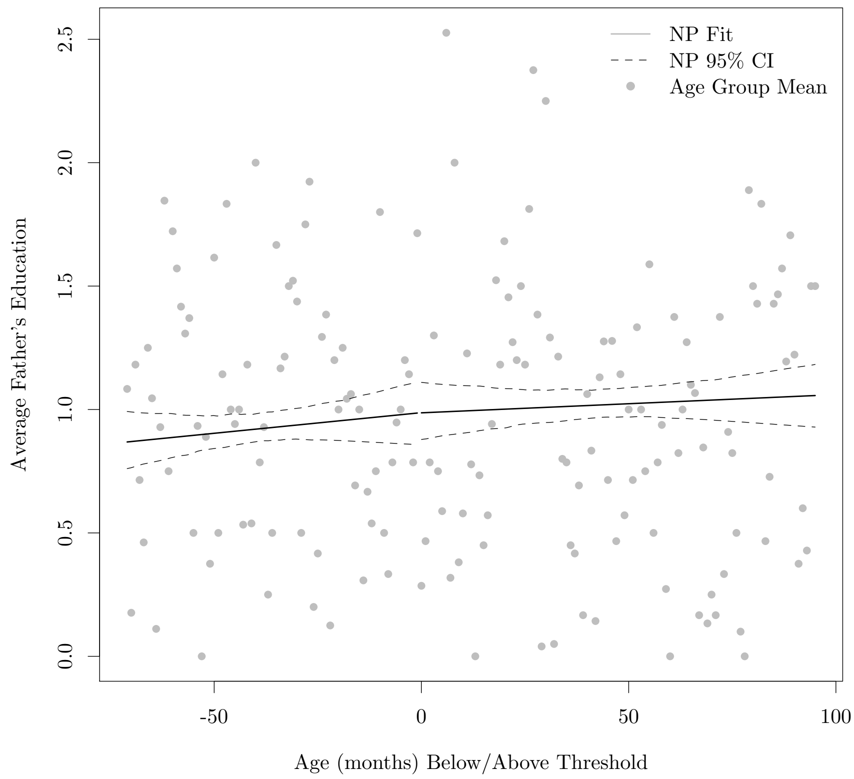

Figure A.7: Discontinuity in father's education? Illustration contains fits and 95% confidence intervals from nonparametric regressions of child's access to health insurance against age (threshold normalized to 0). Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 4.17 \times 10^8$) and above the threshold ($\hat{h}_a = 4.35 \times 10^7$).

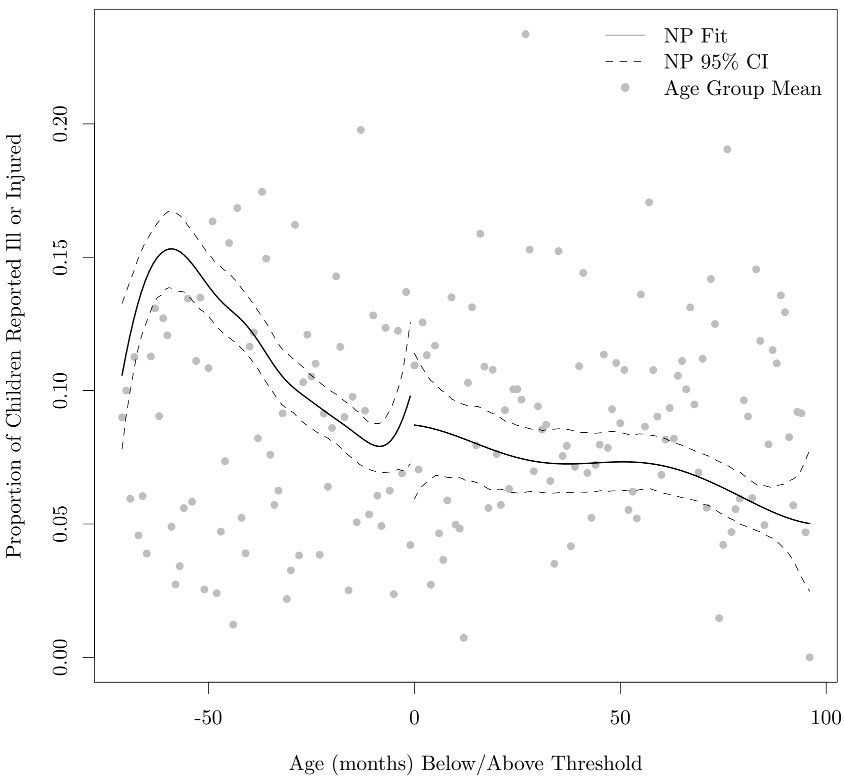

Figure A.8: Discontinuity in child illness reporting in non-filtered sample? Illustration contains fits and 95% confidence intervals from nonparametric regressions of children's reported illness or injury (0/1) against age (threshold normalized to 0) within the entire sample of children (aged 0-14) in the 1995 OHS. Nonparametric regressions estimated separately on either side of the threshold; least-squares cross-validated optimal bandwidths derived below the threshold ($\hat{h}_b = 5.35$) and above the threshold ($\hat{h}_a = 12.8$).

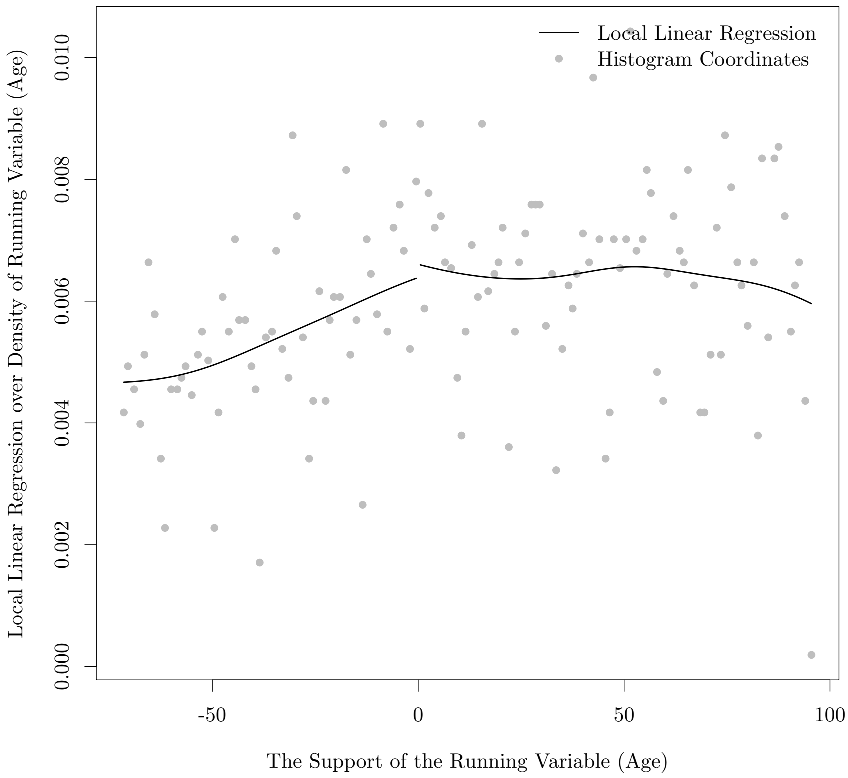

Figure A.9: Discontinuity in child health insurance coverage for the non-filtered Sample? Figure follows McCrary (2008), limited to all children (not just ill and injured children) covered by a medical aid scheme. Included is the fitted local linear regression and coordinate points from the histogram, where the $x-$coordinate is the histogram midpoint, and the $y-$coordinate is the histogram height. The estimated discontinuity is $\hat{\theta} = 0.345$, its standard error $\hat{\sigma} = 1.02$, the $p-$value= 0.74 from a $t-$test of discontinuity significance, histogram binsizes $\hat{b} = 1.29$ and local linear regression bandwidths $\hat{h} = 12.87$.

B Descriptive Statistics

Table B.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	Home Care
Insured Child	0.440	0.169	0.226
Black Child	0.545	0.759	0.728
Coloured Child	0.110	0.138	0.138
Asian Child	0.084	0.033	0.025
White Child	0.261	0.070	0.109
Northern Cape Province	0.118	0.088	0.088
Western Cape Province	0.148	0.205	0.167
Eastern Cape Province	0.026	0.058	0.033
Free State Province	0.107	0.061	0.142
KwaZulu-Natal Province	0.230	0.309	0.268
Northwest Province	0.072	0.094	0.046
Gauteng Province	0.171	0.096	0.134
Mpumalanga Province	0.090	0.061	0.092
Limpopo Province	0.038	0.029	0.029
Urban Locale	0.716	0.532	0.573
Med Center > 60min	0.074	0.158	0.138
30min < Med Center < 60min	0.087	0.201	0.180
15min < Med Center < 30min	0.345	0.296	0.322
Med Center < 15 min away from house	0.494	0.344	0.360
>10 in Household (HH)	0.061	0.128	0.117
8-9 in HH	0.097	0.146	0.117
7 in HH	0.069	0.125	0.134
6 in HH	0.123	0.132	0.142
5 in HH	0.210	0.172	0.201
4 in HH	0.256	0.197	0.172
<4 in HH	0.184	0.099	0.117
Mother: No Education	0.887	0.864	0.866
Mom: Some Education	0.049	0.047	0.046
Mom: Primary Education	0.023	0.061	0.063
Mom: Matric Completed	0.041	0.029	0.025
Mother: Insured	0.038	0.012	0.010
Mother and Father: Alive	0.954	0.897	0.904
Father: No Education	0.394	0.591	0.561
Dad: Some Education	0.171	0.178	0.176
Dad: Primary Education	0.143	0.138	0.146
Dad: Matric Completed	0.292	0.094	0.117
Father: Insured	0.350	0.113	0.197
(ln) HH Expenditure	7.9	7.2	7.3
Child Age	-41.3	-39.4	-34.7

Table B.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	Home Care
Insured Child	0.559	0.163	0.256
Black Child	0.415	0.724	0.646
Coloured Child	0.129	0.146	0.159
Asian Child	0.123	0.038	0.052
White Child	0.333	0.091	0.143
Northern Cape Province	0.171	0.105	0.140
Western Cape Province	0.108	0.181	0.195
Eastern Cape Province	0.045	0.057	0.042
Free State Province	0.066	0.046	0.088
KwaZulu-Natal Province	0.270	0.295	0.279
Northwest Province	0.115	0.091	0.062
Gauteng Province	0.144	0.106	0.088
Mpumalanga Province	0.052	0.089	0.084
Limpopo Province	0.029	0.030	0.023
Urban Locale	0.782	0.549	0.545
Med Center > 60min	0.060	0.143	0.153
30min < Med Center < 60min	0.063	0.165	0.175
15min < Med Center < 30min	0.322	0.363	0.305
Med Center < 15 min away from house	0.360	0.329	0.367
>10 in Household (HH)	0.068	0.095	0.140
8-9 in HH	0.024	0.135	0.172
7 in HH	0.079	0.118	0.114
6 in HH	0.121	0.183	0.182
5 in HH	0.310	0.190	0.188
4 in HH	0.289	0.198	0.146
<4 in HH	0.110	0.082	0.058
Mother: No Education	0.866	0.791	0.867
Mom: Some Education	0.034	0.101	0.084
Mom: Primary Education	0.042	0.084	0.039
Mom: Matric Completed	0.058	0.025	0.010
Mother: Insured	0.050	0.017	0.013
Mother and Father: Alive	0.916	0.861	0.890
Father: No Education	0.339	0.559	0.503
Dad: Some Education	0.178	0.175	0.188
Dad: Primary Education	0.226	0.169	0.201
Dad: Matric Completed	0.257	0.097	0.107
Father: Insured	0.433	0.112	0.253
(ln) HH Expenditure	8.2	7.3	7.4
Child Age	44.0	43.8	44.1

C Estimation Results

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

	Public Facility			Private Facility		
Variable	Coeff.	S.E.	Pr(> |t|)	Coeff.	S.E.	Pr(> |t|)
Intercept	-2.4662	0.012	0.00	-0.3208	0.012	0.00
Child Eligible (Age<6)	0.0417	0.002	0.00	-0.2922	0.005	0.00
Child Age	-0.0087	0.005	0.06	-0.0191	0.005	0.06
Child Age Squared	-0.0001	0.000	0.25	-0.0002	0.000	0.25
Child Insured	0.1583	0.047	0.00	-0.2938	0.047	0.00
(ln) HH Expenditure	0.8595	0.044	0.00	-0.4462	0.048	0.00
Squared Expenditure	-0.0624	0.005	0.00	0.0606	0.005	0.00
Child Age x Eligible	-0.0145	0.008	0.08	-0.0217	0.009	0.08
Squared Age x Eligible	-0.0001	0.000	0.50	-0.0001	0.000	0.50

Coefficient estimates (Coeff.), Standard Errors (S.E.) and significance probability (Pr(> |t|)) for multinomial logit estimates of health facility choice for children up to the age of 14 using 1995 South African Household Survey.

Table C.2: Nonparametric Bandwidth Summary for Model 1

Variable	Bandwidth	Scale Factor
Facility Choice	0.00	0.00
Child Eligible (Age<6)	0.58	2.79
Child Age	63.28	2.90
Child Insured	0.21	1.03
(ln) HH Expenditure	2.06	4.19
rd.age	0.00	0.00
age	0.58	2.79
insure	63.28	2.90
inc	0.21	1.03

Likelihood Cross-Validated bandwidths and resulting scale factors from kernel density estimation of nonparametric conditional mode model of health care facility choice for children up to the age of 14, using the 1995 South African October Household Survey.

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

	Public Facility			Private Facility		
Variable	Coeff.	S.E.	Pr(> |t|)	Coeff.	S.E.	Pr(> |t|)
Intercept	-2.1314	0.010	0.00	0.4065	0.011	0.00
Child Eligible (Age<6)	0.1380	0.004	0.00	-0.2252	0.004	0.00
Child Age (in months)	-0.0018	0.005	0.71	-0.0143	0.005	0.71
Child Age Squared	-0.0001	0.000	0.42	-0.0002	0.000	0.42
Child is Insured	-0.3173	0.038	0.00	-0.5771	0.038	0.00
(ln) HH Expenditure	0.7423	0.038	0.00	-0.6067	0.040	0.00
Squared (ln) Expenditure	-0.0527	0.005	0.00	0.0644	0.005	0.00
Black Child	0.1702	0.038	0.00	-0.1005	0.035	0.00
Coloured Child	0.0023	0.044	0.96	-0.3000	0.038	0.96
Asian Child	-0.1482	0.024	0.00	0.5079	0.026	0.00
Western Cape	-0.1550	0.022	0.00	-0.0118	0.019	0.00
Eastern Cape	-0.0741	0.030	0.01	-0.2181	0.023	0.01
Northern Cape	0.5132	0.016	0.00	-0.0844	0.013	0.00
Free State	-0.8868	0.017	0.00	-0.3277	0.014	0.00
KwaZulu-Natal	0.0916	0.049	0.06	-0.0569	0.040	0.06
Northwest Province	0.5277	0.019	0.00	0.6239	0.017	0.00
Gauteng Province	-0.0307	0.020	0.12	-0.0580	0.019	0.12
Mpumalanga Province	-0.2072	0.018	0.00	-0.1794	0.015	0.00
Urban Household	-0.0317	0.055	0.56	-0.0831	0.061	0.56
HH Size (8-9)	0.0272	0.007	0.00	-0.2960	0.004	0.00
HH Size (7)	0.1951	0.004	0.00	0.1760	0.003	0.00
HH Size (6)	0.2808	0.006	0.00	0.2943	0.004	0.00
HH Size (5)	0.3106	0.041	0.00	0.5853	0.038	0.00
HH Size (4)	0.6596	0.035	0.00	0.7682	0.034	0.00
HH Size (<4)	0.3421	0.008	0.00	0.8927	0.007	0.00
30 min < Distance < 60 min	0.0533	0.009	0.00	-0.3052	0.007	0.00
15 min < Distance < 30 min	0.0084	0.034	0.81	0.2931	0.030	0.81
Facility Distance < 15 min	0.0952	0.031	0.00	0.2778	0.028	0.00
Mother: Some Schooling	0.1468	0.004	0.00	0.0997	0.004	0.00
Mother: Primary School	0.4003	0.004	0.00	0.0356	0.004	0.00
Mother: Secondary School	0.6707	0.002	0.00	0.8556	0.002	0.00
Mother: Insured	-1.3373	0.003	0.00	-0.9597	0.003	0.00
Father: Some Schooling	0.2586	0.013	0.00	0.2393	0.013	0.00
Father: Primary School	0.1807	0.015	0.00	0.0828	0.018	0.00
Father: Secondary School	0.5004	0.011	0.00	0.5139	0.012	0.00
Father: Insured	-1.5410	0.015	0.00	-1.2363	0.015	0.00
Father Alive	-0.0731	0.025	0.00	0.0548	0.018	0.00
Child Age x Eligible	-0.0133	0.008	0.12	-0.0230	0.010	0.12
Age Squared x Elig	0.0000	0.000	0.63	-0.0001	0.000	0.63

Coefficient estimates (Coeff.), Standard Errors (S.E.) and significance probability (Pr(> |t|)) for multinomial logit estimates of health facility choice for children up to the age of 14 using 1995 South African Household Survey.

Table C.4: Nonparametric Bandwidth Summary for Model 2

Variable	Bandwidth	Scale Factor
Health Facility	0.04	0.17
Child Eligible (Age<6)	0.62	2.97
Child Age (in months)	24884061.63	1141481.62
Child is Insured	0.25	1.19
(ln) HH Expenditure	1.43	2.90
Population Group	0.07	0.36
Province of Residence	0.05	0.22
Urban Household	0.28	1.36
Household Size	0.21	1.00
Time to Facility	0.16	0.78
Mother's Education	0.40	1.92
Father Alive	0.17	0.82
Mother: Insured	1.00	4.81
Father's Education	0.28	1.36
Father: Insured	0.23	1.08

Likelihood Cross-Validated bandwidths and resulting scale factors from kernel density estimation of nonparametric conditional mode model of health care facility choice for children up to the age of 14, using the 1995 South African October Household Survey.