A Bayesian Nonparametric Causal Model for Regression Discontinuity Designs*

George Karabatsos† and Stephen G. Walker‡

September 26, 2013

Abstract

The regression discontinuity (RD) non-randomized design can identify and estimate causal effects for a "locally-randomized" subgroup of subjects, under relatively mild conditions. Though, the accurate estimation of causal effects still relies on the predictive accuracy of the statistical model. We propose a flexible Bayesian nonparametric regression model which can provide predictively-accurate estimates of causal effects, either in terms of the mean, variance, distribution function, quantile, probability density, or any other functional of the outcome variable. We illustrate the model through the analysis of two real educational sets.
Keywords: Bayesian Nonparametric Regression, Causal Inference, Sharp and Fuzzy Regression Discontinuity Designs.

1 Introduction

A basic objective in scientific research is to infer causal effects from empirical data. Randomized studies are the gold standard of causal inference (Rubin, 2008), because they ensure that any differences in treatment outcomes and non-treatment outcomes are only due to changes in the treatment variable. In a standard randomized study, the investigator randomly assigns each subject into one of the treatment conditions, usually with equal probability; each subject complies with her/his treatment assignment; and the Stable Unit Treatment Value Assumption (SUTVA) holds, such that any subject's outcome in response to treatment is independent of the treatments received by all the other subjects (e.g., Cox, 1958). Then unconfoundedness holds such that treatment outcomes and non-treatment outcomes are jointly independent of treatment assignments, conditionally on every possible value of all subject pre-treatment characteristics (covariates); and overlap holds such that for every value of the subject pre-treatment characteristics, there is a chance to receive either the treatment or the non-treatment. Then given unconfoundedness and overlap, the causal effect can be identified and estimated by a comparison of the mean of treatment outcomes, against the mean of non-treatment outcomes (e.g., Imbens, 2004).

*This research is supported by a National Science Foundation Grant SES-1156372.

†Corresponding author. Professor, University of Illinois-Chicago, U.S.A., Program in Measurement and Statistics. 1040 W. Harrison St. (MC 147), Chicago, IL 60607. E-mail: gkarabatsos1@gmail.com. Phone: 312-413-1816.

‡Professor, The University of Texas at Austin, Division of Statistics and Scientific Computation.

Often it is necessary to estimate causal effects from a non-randomized study, because a randomized study is often infeasible due to financial, ethical, or time constraints (Rubin, 2008). However, causal effect estimation is more challenging in a non-randomized study, because in such a setting, treated subjects almost surely differ in pre-treatment characteristics, compared to non-treated subjects. Popular causal models for non-randomized studies are often used to estimate causal effects on the basis of the SUTVA, unconfoundedness, and overlap assumptions (e.g., Imbens, 2004), while asserting no hidden bias such that none of the unobserved subject background characteristics serve as confounding variables (e.g., Rosenbaum 2002). Such models involve either a regression of outcomes, on the treatment receipt variable and on the observed pre-treatment characteristics, and/or involves matching/weighting subjects on the observed pre-treatment characteristics and/or on propensity scores (e.g., Imbens, 2004). The regression may also be on hypothesized unobserved pre-treatment covariates, in order to examine the sensitivity of the causal effect estimates over varying degrees of hidden bias (e.g., Rosenbaum & Rubin, 1983b). Though, arguably, for typical non-randomized studies, SUTVA, unconfoundedness, and overlap do not strictly hold for *every* possible value of observed (or even both observed and unobserved) subject pre-treatment characteristics (Imbens, 2004; Lee, 2008).

We propose a Bayesian nonparametric regression model (Karabatsos & Walker, 2012) for causal inference in non-randomized studies. It is an infinite-mixture model which allows for the entire probability density of the outcome variable to change flexibly as a function of the covariate(s), including the treatment assignment variable. For non-randomized studies, we propose the model for regression discontinuity (RD) designs (Thistlewaite & Campbell, 1960; Cook, 2008). This is because under a RD design, estimates of causal effects can be identified under assumptions that are weaker and thus more realistic than the joint assumptions of SUTVA, overlap, and unconfoundedness. The RD design employs a continuous-valued assignment variable (Lee & Lemieux, 2010). Each subject of the design is assigned to the treatment (non-treatment, respectively) when her/his observed value of the assignment variable equals or exceeds a cutoff value (is less than the cutoff, respectively)[1]. The RD design provides a "locally-randomized experiment", such that the causal effect of treatment versus non-treatment can be identified and estimated at the cutoff, under mild conditions, including the condition that every subject of the design has imprecise control over the assignment variable (Lee, 2008). As proven earlier (Goldberger, 2008/1972), the RD design can empirically produce causal effect estimates that are similar to those estimates of a randomized study (Aiken et al., 1998; Buddelmeyer & Skoufias, 2004; Black, et al. 2007; Berk et al. 2010; Shadish, et al., 2011).

Even though the RD design has existed for more than 50 years, it initially received limited interest from the fields of education, psychology, and statistics (Cook, 2008), and from the economics, political science, criminology, and health science fields. But between 1997 through 2013, a surge of at least 74 RD-based empirical studies emerged in these fields (Lee and Lemieux, 2010; Bloom, 2012; Wong et al. 2013; Li et al., 2013). There are at least three reasons why (van der Klaauw, 2008; Lee & Lemieux, 2010). First, many non-randomized studies employ treatment assignment rules that can be easily conceptualized as RD designs. Second, the

[1]The assignment variable and cutoff value may also be vector-valued (Imbens & Lemieux, 2008). The model proposed in this paper easily extends to vector-valued assignment variables and cutoffs.

empirical results of RD designs are intuitive and can be easily conveyed, say, via a plot of the outcomes against the assignment variable. Third, for non-randomized studies, causal effect identification and estimation in a RD design requires weaker and hence more credible assumptions, compared to the stronger assumptions (e.g., unconfoundedness, overlap) that are required by popular causal models. Therefore, the design gives the researcher the flexibility to consider from a range of different causal estimation methods.

The mainstream causal models that are used for non-randomized studies, including those involving RD designs, have primarily based causal inferences on comparisons of the mean of the treatment outcomes, and the mean of the non-treatment outcomes (Imbens, 2004; Lee & Lemieux, 2010). However in many causal inference settings, it may also be of interest to base causal inferences on comparisons of other features of the outcomes. Our Bayesian nonparametric regression model can provide inferences of causal effects in terms of how the treatment variable impacts the mean, variance, distribution function, a quantile, probability density, hazard function, and any other functional of the outcome variable. Finally, for either a randomized or a non-randomized study, the accurate estimation of causal effects relies on an appropriate model for the data. Karabatsos and Walker (2012) showed that their Bayesian nonparametric regression model tended to have better predictive performance than other parametric and flexible nonparametric regression models of common usage, over many real data sets. The other outperformed models include BART (Chipman et al., 2010) which has been proposed for causal inference (Hill, 2011). Moreover, we show how our Bayesian nonparametric model can be extended to handle causal inferences from a fuzzy RD design (Trochim, 1984), where there is imperfect treatment compliance among the subjects.

In Section 2, using the potential outcomes framework of causal inference (e.g., Neyman 1923), we review the assumptions that data from a RD design needs to meet in order to identify and estimate causal effects. In Section 3, we review the current models that are used to estimate causal effects from a RD designs. Then we describe our Bayesian nonparametric causal model for such designs. In Section 4, we illustrate our model through the analysis of two data sets, which were recently collected under a partnership between four Chicago university schools of education. The partnership collaborated to institute a new teacher education curriculum, that aims to graduate teachers who would improve Chicago public school education.

2 Identifying Causal Effects in a RD Design

The theory of potential outcomes (Neyman 1923; Rubin, 1974; Rubin, 1978; Angrist, et al., 1996) is arguably the most useful framework of causal inference, for either a randomized or a non-randomized study. The theory states that a study involves: a sample of *subjects*, indexed by $i = 1, \ldots, n$, and each described by *pre-treatment characteristics (covariates)* $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})^\intercal$ that the study recorded; a *treatment receipt variable* $T(A) \in \{t = 0, 1\}$, where for a given subject, $T_i(A) = 0$ indicates receipt of non-treatment (e.g., control), and $T_i(A) = 1$ indicates receipt of the treatment, when assigned treatment $A \in \{t = 0, 1\}$; *potential outcomes* $Y_i(\boldsymbol{A}_n, \boldsymbol{T}_n)$ defined at a common point in time, for all $(\boldsymbol{A}_n, \boldsymbol{T}_n) \in \{0, 1\}^n \times \{0, 1\}^n$ with $\boldsymbol{A}_n = (A_1, \ldots, A_n)^\top$ and $\boldsymbol{T}_n = (T_1, \ldots, T_n)^\top$. Under SUTVA, which implies no interference between all n subjects and no versions of treatments (e.g., Cox, 1958;

Rubin, 1990), and under perfect treatment compliance (i.e., $(T_i(1), T_i(0)) = (0, 1)$, $i = 1, \ldots, n$), the 2^{2n} potential outcomes $Y_i(\boldsymbol{A}_n, \boldsymbol{T}_n)$ can be reduced to the pair $Y_i(t), t = 0, 1$. Then the *causal effect* of T on Y is defined by a comparison of potential outcomes, such as $Y(1) - Y(0)$. The *fundamental problem of causal inference* is that at least one of the potential outcomes $(Y_i(1), Y_i(0))$ is missing, and thus the causal effect (e.g., $Y_i(1) - Y_i(0)$) is unobservable, because each subject can only be exposed to only one of the treatment conditions at a given time point (Holland, 1986).

In the non-randomized, *regression discontinuity* (RD) design (Thistlewaite & Campbell, 1960; Cook, 2008), causal effects can be identified under mild and arguably realistic assumptions. In a RD design, each subject is assigned to a treatment (non-treatment, respectively) if the subject's value of an *assignment variable*[2] (Berk & Rauma, 1983) equals to or exceeds a known and meaningful threshold (is less than the threshold, respectively). Specifically, if R_i is the treatment variable of a subject, and r_0 is the given threshold, then the treatment assignment mechanism is defined by $A_{r_0}^{(R_i)} = \mathbf{1}(R_i \geq r_0)$, corresponding to realization $a_{r_0}^{(r_i)} = \mathbf{1}(r_i \geq r_0)$ and r_i, where $\mathbf{1}(\cdot)$ denotes the indicator function. In the *sharp RD design* (Thistlewaite & Campbell, 1960), the treatment receipt probability is $\Pr(T = 1|R = r) = \mathbf{1}(r \geq r_0)$, and it thus jumps discontinuously from 0 to 1 at the cutoff r_0. In a *fuzzy RD design* (Trochim, 1984), the treatment receipt probability $\Pr(T = 1|R = r)$ has a discontinuous jump that is less than 1, at r_0, as a result of imperfect treatment compliance. Imperfect compliance can occur in settings where the assignment variable R measures the eligibility to receive a treatment, and some ineligible subjects (with $R_i < r_0$) decided to receive treatment (i.e., $T_i = 1$), and some eligible subjects (with $R_i \geq r_0$) decided receive the non-treatment (i.e., $T_i = 0$). Now, let $\mathcal{L}_{n_0}^{\epsilon}(r_0) = \{i : |r_i - r_0| < \epsilon\}$ be the subset of n_0 subjects whose assignment variables r_i are in a neighborhood of size $\epsilon > 0$ around r_0.

The identification of causal effects in a RD design, sharp or fuzzy, relies on data meeting the following five assumptions.

1. **Assumption (RD)** (Hahn, et al. 2001): $\lim_{r \uparrow r_0} \mathrm{E}(T|r) \neq \lim_{r \downarrow r_0} \mathrm{E}(T|r)$.

2. **Local SUTVA (LS)** (Cattaneo et al. 2013). There exists a $\epsilon > 0$ such that $T_i(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}_{n_0})}) = T_i(A_{r_0}^{(R_i)})$ and $Y_i(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}_{n_0})}, \boldsymbol{T}_{n_0}(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}_{n_0})})) = Y_i(A_{r_0}^{(R_i)}, T_i(A_{r_0}^{(R_i)}))$ for all $i \in \mathcal{L}_{n_0}^{\epsilon}(r_0)$ and all $\boldsymbol{R}_{n_0}$, i.e., the potential outcomes are unrelated to the treatment status of the other subjects within $\mathcal{L}_{n_0}^{\epsilon}(r_0)$. Then for all $i \in \mathcal{L}_{n_0}^{\epsilon}(r_0)$, $(T_i(0), T_i(1))$ are the potential treatment receipt outcomes; $T_i(1) - T_i(0)$ is the causal effect of A on T; a *complier* is a subject with $(T_i(1), T_i(0)) = (1, 0)$, for all $\epsilon > 0$ of $(T_i(A_{r_0}^{(r_0-\epsilon)}), T_i(A_{r_0}^{(r_0+\epsilon)}))$; and $ITT = Y_i(1, T_i(1)) - Y_i(0, T_i(0))$ is the Intention-to-Treat (ITT) causal effect of A on Y.

3. **Local Exclusion Restriction (LER)** (Hahn, et al. 2001): There exists a $\epsilon > 0$ such that for all subjects $i \in \mathcal{L}_{n_0}^{\epsilon}(r_0)$, any effect of A on Y must be via an effect of A on T, in the sense that $Y_i(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}_{n_0})}, \boldsymbol{T}_{n_0}) = Y_i(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}'_{n_0})}, \boldsymbol{T}_{n_0})$ for all $\boldsymbol{R}_{n_0}, \boldsymbol{R}'_{n_0}$ and all $\boldsymbol{T}_{n_0} \in \{0, 1\}^{n_0}$.

[2] Also called the *forcing variable* (Imbens & Lemieux, 2008), or the *running variable* (McCrary, 2008).

4. **Local Monotonicity (LM)** (Hahn, et al. 2001): $T_i(A_{r_0}^{(r_0+\epsilon)}) \geq T_i(A_{r_0}^{(r_0-\epsilon)})$ for some $\epsilon > 0$ and all $i \in \mathcal{L}_{n_0}^{\epsilon}(r_0)$.

5. **Local Randomization (LR)** (Lee, 2008): Every subject has "imprecise control" over the assignment variable R, in the sense that the c.d.f. $F_R(r|w) = \Pr(R \leq r|w)$ is continuous in r at r_0, and $0 < F_R(r_0|w) < 1$, for every subject that is uniquely identified by a value w of the latent variable W. Then, the treatment assignments $A_{r_0}^{(R_i)}$ are "as good as randomized" in a region around the cutoff (r_0), in the sense that $F_{\mathbf{x}}(\mathbf{x}|r)$ and $F_W(w|r)$ are each the same just below and just above r_0, and that conditional expectations $\mathrm{E}[h\{Y(0)\}|r]$ and $\mathrm{E}[h\{Y(1)\}|r]$ are continuous in r at r_0, for all $h\{\cdot\}$ (Imbens & Lemieux, 2008).

Since in a sharp RD design, $\mathrm{E}(T|R = r) = \Pr(T = 1|R = r) = \mathbf{1}(r \geq r_0) = A_{r_0}^{(r)} = T(A_{r_0}^{(r)})$, the design trivially satisfies RD, LER, and LM, with treatment or non-treatment receipt determined by a step function (implying assumption RD), and with full treatment compliance (implying LER and LM). With LS, a RD design only requires that SUTVA hold for only for the subset of subjects located around the cutoff, i.e., $\mathcal{L}_{n_0}^{\epsilon}(r_0)$ for some $\epsilon > 0$, as opposed to SUTVA holding for all subjects in a given study. LS and LER, together, imply that exists some $\epsilon > 0$, such that for the subset of subjects in $\mathcal{L}_{n_0}^{\epsilon}(r_0)$, the potential outcomes $Y_i(\boldsymbol{A}_{r_0}^{(\boldsymbol{R}_{n_0})}, \boldsymbol{T}_{n_0})$ can be re-written as $Y_i(t)$, $t = 0, 1$, so that $Y_i(1) - Y_i(0)$ gives the causal effect of T on Y; imply that $ITT_i = Y_i(1, T_i(1)) - Y_i(0, T_i(0)) = Y_i(T_i(1)) - Y_i(T_i(0)) = (Y_i(1) - Y_i(0))(T_i(1) - T_i(0))$, meaning that for a complier, $ITT_i = Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$ (Angrist, et al. 1996). Assumptions RD, LS and LM together imply that around the cutoff r_0, the treatment assignment A has a positive causal effect on treatment receipt T, with $\mathrm{E}[T_i(1) - T_i(0)|r_0] = \lim_{r \downarrow r_0} \mathrm{E}(T|r) - \lim_{r \uparrow r_0} \mathrm{E}(T|r) > 0$. The LR assumption also implies that the viability of a RD design can be evaluated through a test of the null hypothesis that the marginal density $f(r)$ of the running variable, over all subjects, is continuous at r_0, against the alternative hypothesis of discontinuity (McCrary, 2008). Though, for each data set we analyze for the present study, the assignment variable is not precisely manipulable by any one of the subjects. In the first data set, time is the assignment variable. In the second data set, the assignment variable is a test score. Arguably, each examinee does not know exactly how many correct item responses s/he obtained on the test, and how many are needed in order to attain the minimum (cutoff) passing score on the test.

All five assumptions, together, imply that, for the compliers around a cutoff (r_0), the causal effect (τ) of treatment ($T = 1$) versus non-treatment ($T = 0$) is:

$$\tau = \mathrm{E}[Y_i(1) - Y_i(0)|r_0 \text{ and } i \text{ is a complier}] = \frac{\mathrm{E}[Y_i(1, T_i(1)) - Y_i(0, T_i(0))|r_0]}{\Pr[T_i(1) - T_i(0)|R_i = r_0]} \quad (1a)$$

$$= \lim_{\epsilon \downarrow 0} \mathrm{E}[Y_i(1) - Y_i(0)|T_i(A_{r_0}^{(r_0+\epsilon)}) - T_i(A_{r_0}^{(r_0-\epsilon)}) = 1] \quad (1b)$$

$$= \frac{\lim_{\epsilon \downarrow 0} \mathrm{E}[Y_i(A_{r_0}^{(r_0+\epsilon)}, T_i(1)) - Y_i(A_{r_0}^{(r_0-\epsilon)}, T_i(0))]}{\lim_{\epsilon \downarrow 0} \Pr[T_i(A_{r_0}^{(r_0+\epsilon)}) - T_i(A_{r_0}^{(r_0-\epsilon)}) = 1]} \quad (1c)$$

$$= \frac{\lim_{r \downarrow r_0} \mathrm{E}[Y|r] - \lim_{r \uparrow r_0} \mathrm{E}[Y|r]}{\Pr[i \text{ is a complier } |r_0]} = \frac{\lim_{r \downarrow r_0} \mathrm{E}[Y|r] - \lim_{r \uparrow r_0} \mathrm{E}[Y|r]}{\lim_{r \downarrow r_0} \mathrm{E}[T|r] - \lim_{r \uparrow r_0} \mathrm{E}[T|r]}. \quad (1d)$$

For $\epsilon > 0$ sufficiently small, the conditioning event $T_i(A_{r_0}^{(r_0+\epsilon)}) - T_i(A_{r_0}^{(r_0-\epsilon)}) = 1$ in

(1) refers to the subgroup of subjects, the compliers, for whom treatment changes discontinuously at the cutoff r_0. The numerator of (1) is the average ITT causal effect of A on Y, for the subgroup of subjects located at r_0. RD and LM together ensure that the denominator of (1) is positive, and therefore is the probability that a given subject is a complier at r_0 (Hahn, et al. 2001). In a sharp RD design that involves perfect compliance, the denominator in (1) becomes 1. Then the causal effect of T on Y at r_0 coincides with the average ITT causal effect of A on Y at r_0, with $\tau = \lim_{r \downarrow r_0} \mathrm{E}[Y|R=r] - \lim_{r \uparrow r_0} \mathrm{E}[Y|R=r]$. The two real data sets of the present study involve sharp RD designs.

All five assumptions together, also imply that equation (1) provides an estimator of the causal effect of T on Y, for any chosen of functional $h\{\cdot\}$ of Y (Imbens & Lemieux, 2008). So causal effects not become only identifiable in terms of differences in mean outcomes as in (1) (i.e., when $h\{Y\} = Y$), but also in terms of the difference in the variances of the outcomes $\mathrm{Var}[Y_i(1)|r_0; i \text{ complier}] - \mathrm{Var}[Y_i(0)|r_0; i \text{ complier}]$; the difference in the c.d.f.s of the outcomes $F_{Y(1)}(y|r_0; i \text{ complier}) - F_{Y(0)}(y|r_0; i \text{ complier})$ at any chosen point y (corresponding to the choice $h\{Y\} = \mathbf{1}(Y < y)$); the difference in their quantiles $F^{-1}_{Y(1)}(u|r_0; i \text{ complier}) - F^{-1}_{Y(0)}(u|r_0; i \text{ complier})$ at any chosen point $u \in [0,1]$ (i.e., a c.d.f. inverse at a probability point); the difference in the probability densities $f_{Y(1)}(y|R=r) - f_{Y(0)}(y|R=r)$ at any chosen point y (i.e., a c.d.f. derivative at y); and so on for any other functional of Y.

3 Estimating Causal Effects In a RD Design

The early work on estimating causal effects from a sharp RD design relies on least-squares, linear polynomial modeling of observed outcomes y on observed values of the assignment variable r, which allows for a discontinuity at a cutoff r_0 (e.g., Cook, 2008). Such a linear model has the general form:

$$y_i = \beta_0 + \beta_1(r_i) + \tau a_{r_0}^{(r_i)} + \beta_2(r_i) a_{r_0}^{(r_i)} + \epsilon_i, \qquad \epsilon_i \sim \mathrm{n}(0, \sigma^2), \tag{2}$$

where the coefficient τ represents the average causal effect at r_0; while $\beta_k(r) = \sum_{l=1}^{q_k} \beta_{kl} r^l$ is the polynomial term of integer order $q \geq 1$; and $\mathrm{n}(\cdot|0, \sigma^2)$ is a density of the normal distribution. The (strictly) linear model corresponds to the assumption that $q_1 = 1$ and $\beta_2(r) = 0$; a more general polynomial model corresponds to a choice of integer $q_1 \geq 1$ and assumes $\beta_2(r) = 0$; whereas a model that allows for separate linear or polynomials for either side of the cutoff r_0 allows for non-zero $\beta_1(r)$ and $\beta_2(r)$ with general choices of positive integers q_1 and q_2, respectively (Bloom, 2012).

During the 1960s and 1970s it was recognized that the linear model (2) can produce biased estimates of the causal effect τ, even due to outliers of y_i that correspond to observations r_i that are located far away from a cutoff r_0 (Cook, 2008). Therefore, for estimating the causal effect τ from a sharp RD design, the local linear model (Fan & Gijbels, 1996) has been proposed, using a bandwidth parameter that assigns relatively higher weights to observations (y_i, r_i) corresponding to r_i values close to the cutoff r_0 (Hahn et al. 2002; Imbens & Lemieux, 2008). This model can be extended to outcome indicators $h\{Y\} = \mathbf{1}(Y < y)$, in order to provide a quantile regression and estimates of causal effects in terms of Y quantile comparisons (Frandsen et al., 2008). Also, these local linear models can be extended to estimate the causal effect τ in a fuzzy RD design. This estimation, equivalent to a

2-stage least-squares estimation of an instrumental variable model, involves a local linear regression for the outcome Y on R to estimate the limits in the numerator (1), and a separate local linear regression for the outcome T on R estimate the limits in the denominator of (1) (Imbens & Lemieux, 2008; Frandsen et al., 2008). Though, for either a sharp or a fuzzy RD design, while the accuracy of causal effect estimates from a local-linear model depends critically on the choice of bandwidth parameter, only large-sample justifications have been provided for such choices (Imbens & Kalyanaraman, 2012). Moreover, the quantile regression local linear model suffers from the "quantile crossing" problem (Bassett & Koenker, 1982), and relies on a large-sample ad-hoc resorting correction as a solution. Finally, methods have been proposed to estimate and perform randomization tests of causal effects, for the subset of locally-randomized subjects in $\mathcal{L}^{\epsilon}_{n_0}(r_0)$. The subset is based on the largest value of $\epsilon > 0$ that leads to a non-rejection of the null hypothesis of zero effect of T on X_k, for subject pre-treatment variables $k = 1, \ldots, p$ (Cattaneo, et al. 2013; Sales & Hansen, 2013; Li et al. 2013). However, the presence of hidden bias can affect the estimate of ϵ, and thus, affect the estimates and tests of the causal effect τ.

Our Bayesian nonparametric regression model (Karabatsos & Walker, 2012) can be extended to RD settings, in order to provide estimates of causal effects, in terms of how changes in the treatment variable impact the mean, variance, quantiles, c.d.f., p.d.f., hazard function, and other functionals of the potential outcome variables. For the sharp RD design, our Bayesian nonparametric model is given by:

$$
\begin{aligned}
f(y_i|r_i, a_{r_0}^{(r_i)}) &= \sum_{j=-\infty}^{\infty} \mathrm{n}(y|\mu_j, \sigma_j^2)\omega_j(\eta_\omega(r_i), \sigma_\omega(r_i)),\ i = 1, \ldots, n, && \text{(3a)}\\
\omega_j(\eta_\omega(r), \sigma_\omega(r)) &= \Phi(\{j - \eta_\omega(r)\}/\sigma_\omega(r)) - \Phi(\{j - 1 - \eta_\omega(r)\}/\sigma_\omega(r)) && \text{(3b)}\\
\eta_\omega(r) &= \beta_{0\omega} + \beta_{\omega 1} r + \beta_{\omega 2} a_{r_0}^{(r)} && \text{(3c)}\\
\sigma_\omega(r) &= \lambda_{0\omega} + \lambda_{\omega 1} r + \lambda_{\omega 2} a_{r_0}^{(r)} && \text{(3d)}\\
(\mu_j, \sigma_j^2) &\sim \mathrm{n}(\mu_j|\mu_\mu, \sigma_\mu^2)\mathrm{ig}(\sigma_j^2|1, b_\sigma) && \text{(3e)}\\
(\mu_\mu, \sigma_\mu^2) &\sim \mathrm{n}(\mu_\mu|\mu_0, \sigma_0^2)\mathrm{un}(\sigma_\mu|0, b_{\sigma\mu}) && \text{(3f)}\\
(b_\sigma, \boldsymbol{\beta}_\omega, \boldsymbol{\lambda}_\omega) &\sim \mathrm{ga}(b_\sigma|a_0, b_0)\pi(\boldsymbol{\beta}_\omega, \boldsymbol{\lambda}_\omega) && \text{(3g)}
\end{aligned}
$$

where the mixture weights $\omega_j(\eta_\omega(r), \sigma_\omega(r))$ sum to 1, at every value of r; $\mathrm{ga}(\cdot|a, b)$ is the density of the gamma distribution with shape and rate parameters (a, b); $\mathrm{ig}(\cdot|a, b)$ is the inverse gamma density; and $\mathrm{un}(\cdot|0, b)$ is the uniform distribution density. The model allows the entire probability density of the outcome variable to change flexibly as a function of covariates. Also, the model has a discontinuity at r_0 due to the presence of $a_{r_0}^{(r_i)}$ in both (3c) and (3d). The effect, controlled by the coefficients $(\boldsymbol{\beta}_\omega, \boldsymbol{\lambda}_\omega)$, is to reallocate the weights either side of r_0, resulting in different densities either side of this value. Obviously, there is a discontinuity if and only if each of the coefficients $(\lambda_{2\omega}, \beta_{2\omega})$ are non-zero. Moreover, when prior information is limited, we may specify vague prior hyper-parameters $\mu_0 = 0, \sigma_0^2 \to \infty$, $a_0 \to 0$, $b_0 \to 0$, with a large-variance multivariate normal prior density $\pi(\boldsymbol{\beta}_\omega, \boldsymbol{\lambda}_\omega) = \mathrm{n}(\boldsymbol{\beta}_\omega, \boldsymbol{\lambda}_\omega|\mathbf{0}, 10^5\mathbf{I}_{p+1})$, along with a choice of prior parameter $b_{\sigma\mu}$ corresponding to prior knowledge about the range of Y. Alternatively, given that in order to create a discontinuity we only need $\lambda_{\omega 2} = \beta_{\omega 2} = 1$, we may assign a $\beta \sim$ Bernoulli(1/2) prior to $\beta = \lambda_{\omega 2} = \beta_{\omega 2}$.

Our model (3) has infinite-dimensional parameter, $\Gamma = ((\mu, \sigma_j^2)_{j=-\infty}^{\infty},\ \mu_\mu,\ \sigma_\mu^2,$

b_σ, $\boldsymbol{\beta}_\omega$, $\boldsymbol{\lambda}_\omega$). A set of data $\mathcal{D}_n = \{y_i, r_i, a_{r_0}^{(r_i)}\}_{i=1}^n$ updates the joint prior density $\pi(\Gamma)$ to a posterior density, given by $\pi(\Gamma|\mathcal{D}_n) \propto \prod_{i=1}^n f(y_i|r_i, a_{r_0}^{(r_i)}; \Gamma)\pi(\Gamma)$ up to a proportionality constant. Then:

$$\mathrm{E}_n(y|r, a_{r_0}^{(r)}) = \int \left\{ \int y f(y|r, a_{r_0}^{(r)}; \Gamma) \mathrm{d}y \right\} \mathrm{d}\Pi(\Gamma|\mathcal{D}_n)$$

gives the posterior predictive expectation of the outcome y, conditionally on the assignment variable (r) and on the treatment assignment $(a_{r_0}^{(r)})$. If all five assumptions hold for the sharp RD design, then a posterior predictive estimate of the causal effect of T on Y is given by $\widehat{\tau}_h = \mathrm{E}_n(h\{y\}|r_0, 1) - \mathrm{E}_n(h\{y\}|r_0, 0)$, for any given choice of functional $h\{\cdot\}$. The posteriors $\pi(\Gamma|\mathcal{D}_n)$ and $\mathrm{E}_n(h\{y\}|r, a_{r_0}^{(r)})$ can be estimated by using the existing Markov Chain Monte Carlo (MCMC) Gibbs sampling methods (Karabatsos & Walker, 2012), along with a Metropolis sampling step for σ_μ^2.

Our model (3) can be easily extended to analyze data from a fuzzy RD design, where the key identifying LM and ER assumptions do not necessarily hold. This extension involves estimating bounds of the causal effects τ_h, with respect to a range of violations of these unverifiable (untestable) assumptions (Angrist et al., 1996). The Appendix provides more details and motivations.

4 Illustrative Applications

Here we report the analysis of two data sets, using our Bayesian model under the prior specification $b_{\sigma\mu} = 5$, and under the vague prior specifications for all other model parameters. For each data set, the reported posterior estimates are based on 40K MCMC samples, obtained by retaining every 5th sample of a 200K MCMC sample run after a 2K sample burn-in. As a result, univariate trace plots showed good mixing of all model parameters, and parameter estimates had 95% MCMC confidence intervals of size near .00.

For the first data set, the aim is to estimate the effect of the new teacher education curriculum on math teaching ability, among teacher education students attending one of the four Chicago universities. This data set involves a sharp RD design, specifically an interrupted time-series design (Cook & Campbell, 1979, Ch. 5), where time is the assignment variable, ranging from fall semester 2007 through spring semester 2013. Fall semester 2010 is the cutoff that represents the first time that the new curriculum (treatment) was instituted into the university. Thus, the old teacher curriculum (the non-treatment) was active before that time point. The outcome variable is the number-correct score on a 25-item, Learning Math for Teaching (LMT) test (University of Michigan). The LMT score was obtained from each undergraduate teacher education student, who had just completed a course in teaching algebra. A total of $n = 347$ undergraduate students completed the LMT test. Among all these students, the test has a Cronbach's alpha reliability estimate of .63; 135 students completed the test during the old curriculum, and 212 students completed the test during the new curriculum; most were female (89.9% female; 94.1% female pre-Fall 2010 intervention; 87.3% female post intervention), and the average LMT score was 12.9 (s.d.= 3.44; pre-Fall 2010 intervention: mean=13.53, s.d.=3.26; post-Fall 2010 intervention: mean=12.49, s.d.=3.49).

Using the Bayesian model, we analyze the data to estimate the effect of the new

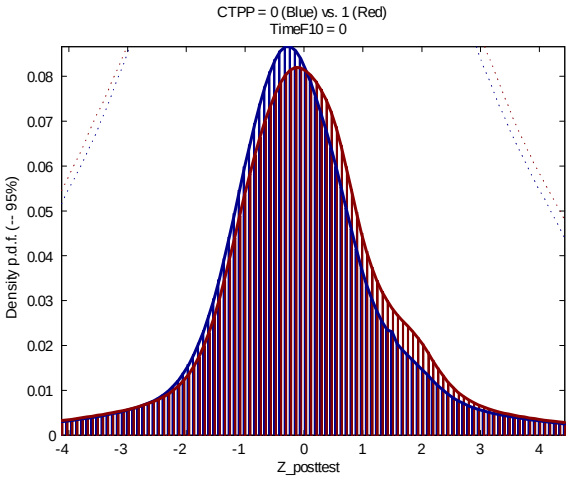

Figure 1: Posterior predictive density estimates of $Y(1)$ (red) and of $Y(0)$ (blue).

curriculum, on student ability to teach math (LMT assessment score), at the cutoff time of Fall 2010. For the model, we specified the z-transformed score on the LMT assessment, as the outcome (dependent) variable, and we specified covariates of the assignment variable TimeF10 = (Year − 2010.9)/10 and the treatment assignment variable CTPP = $\mathbf{1}$(Year $\geq$ 2010.9), with time point 2010.9 referring to Fall semester 2010. Our model displayed good fit to these data. Over all 347 observations, the standardized residuals ranged from −0.84 to 0.77, and the posterior mean estimate of R-squared was .92. From our model, Figure 1 presents the posterior predictive density estimate of the LMT outcome, for the new curriculum (treatment) and for the old curriculum (non-treatment), at Fall 2010. There, we see that the new curriculum, compared to the old curriculum, tended to increase the LMT scores, in terms of shifting the density of LMT scores to the right. This shift corresponds to an increase in the mean (from .17 to .20), the 10%ile (−1.43 to −1.35), the median (.07 to .15), and to a variance decrease (1.78 to 1.69).

The second data set, from another sharp RD design, involves $n = 205$ undergraduate teacher education students, each of whom entered into one of the four Chicago area schools of education during either the years of 2010, 2011, or 2012. It is of interest to investigate whether or not basic skills has an impact on teacher performance (e.g., Gitomer & Brown, 2011). This is because nearly all U.S. schools of education base their undergraduate admissions decisions on each applicant's ability to pass basic skills tests. Here, the assignment variable is defined by the score on an Illinois test of reading basic skills, with minimum cutoff passing score of 240. The outcome variable is the score on the 50-item Haberman (2008) Teacher Prescreener assessment of urban school teaching dispositions. A score in the 40-50 range indicates a very effective teacher. Haberman assessment scores have a test-retest reliability of .93, and have a 95% accuracy rate in predicting which teachers will stay and succeed in the teaching profession (Haberman, 2008). Many urban schools currently use the Haberman Pre-screener to score individuals who are applying for teaching positions. Among all the 205 students of the RD design, 90% were female; the average age is 22.5 (s.d.=5.35; min=19.1, max=56.6; $n = 203$); 47%, 21%, 10%, and 22% attended the four universities, respectively; 49%, 41% and 10% were re-

spectively in the 2010, 2011, or 2012 cohort; the average reading basic skills score is 204.69 (s.d.=33.7; min=137; max=293); and the average Haberman Pre-screener score is 29.82 (s.d.=4.32, min=14, max=42).

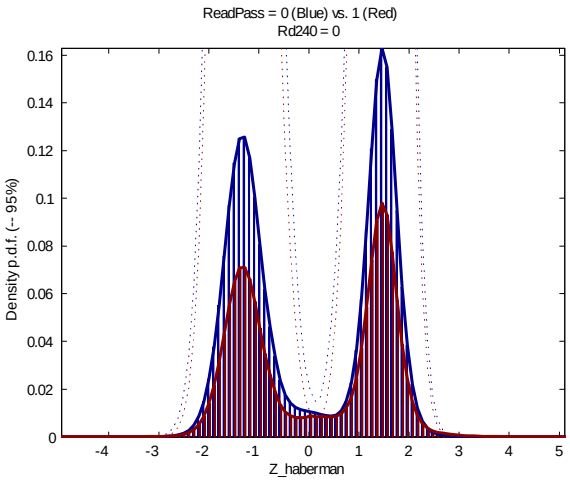

Figure 2: Posterior predictive density estimates of $Y(1)$ (red) and of $Y(0)$ (blue).

Using the Bayesian nonparametric regression model, we analyze the data to estimate the effect of passing the reading basic skills exam on students' ability to teach in urban schools. For the model, we specified the outcome (dependent) variable as the z-score-transformed Haberman score, and we specified covariates of the assignment variable Rd240 = (Read − 240)/10 and reading score passing (assignment) indicator ReadPass = $\mathbf{1}$(Read $\geq$ 240), where Read is the basic skills reading score with minimum passing cutoff score of 240. Our model displayed good fit to these data. Over all 205 observations, the standardized residuals ranged from −1.7 to 1.22, and the posterior mean estimate of R-squared was .98. From our model, Figure 2 presents the posterior predictive density estimates of the Haberman score, for passing the basic skills reading test (treatment) and for not passing the test (non-treatment), at basic skills score 240. There is a bimodal density of Haberman scores, for passing the test, and for not passing the test, with the two modes indicating groups of students who scored below and above average, respectively. For the Haberman score, the densities correspond to an increase in mean (from .13 to .26), median (.05 to .28), 75%ile (.97 to 1.43), 90%ile (1.66 to 2.21), 95%ile (2.13 to 2.82), and variance (1.60 to 3.36); and to a decrease in 5%ile (−1.74 to −2.38) and 10%ile (−1.30 to −1.70).

References

Aiken, L., West, S., Schwalm, D., Carroll, J., and Hsiung, S. (1998), "Comparison of a Randomized and Two Quasi-Experimental Designs in a Single Outcome Evaluation Efficacy of a University-Level Remedial Writing Program," *Evaluation Review*, 22, 207–244.

Angrist, J., Imbens, G., and Rubin, D. (1996), "Identification of Causal Effects Using

Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455.

Balke, A. and Pearl, J. (1997), "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176.

Bassett, G. and Koenker, R. (1982), "An empirical quantile function for linear models with iid errors," *Journal of the American Statistical Association*, 77, 407–415.

Berk, R., Barnes, G., Ahlman, L., and Kurtz, E. (2010), "When second best is good enough: A comparison between a true experiment and a regression discontinuity quasi-experiment," *Journal of Experimental Criminology*, 6, 191–208.

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

Black, D., Galdo, J., and Smith, J. (2005), "Evaluating the regression discontinuity design using experimental data," Unpublished manuscript.

Bloom, H. (2012), "Modern regression discontinuity analysis," *Journal of Research on Educational Effectiveness*, 5, 43–82.

Buddelmeyer, H. and Skoufias, E. (2004), *An evaluation of the performance of regression discontinuity design on PROGRESA*, no. 827, World Bank Publications.

Cattaneo, M., Frandsen, B., and Titiunik, R. (2013), "Randomization Inference in the Regression Discontinuity Design: An Application to the Study of Party Advantages in the U.S. Senate," *University of Michigan. February 19th. Unpublished manuscript.*

Chipman, H., George, E., and McCulloch, R. (2010), "BART: Bayesian Additive Regression Trees," *Annals of Applied Statistics*, 4, 266–298.

Cook, T. (2008), "Waiting for Life to Arrive: A history of the regression-discontinuity design in Psychology, Statistics and Economics," *Journal of Econometrics*, 142, 636–654.

Cook, T. and Campbell, D. (1979), *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand McNally.

Cox, D. (1958), *The Planning of Experiments*, New York: John Wiley.

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

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall/CRC.

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

Gitomer, D., Brown, T., and Bonett, J. (2011), "Useful signal or unnecessary obstacle? The role of basic skills tests in teacher preparation," *Journal of Teacher Education*, 62, 431–445.

Goldberger, A. (2008/1972), "Selection bias in evaluating treatment effects: Some formal illustrations," in *Modelling and evaluating treatment effects in economics*, eds. Millimet, D., Smith, J., and Vytlacil, E., Amsterdam: JAI Press, pp. 1–31.

Haberman, M. (2008), *The Haberman Star Teacher Pre-Screener*, Houston: The Haberman Educational Foundation.

Hahn, J., Todd, P., and der Klaauw, W. V. (2001), "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69, 201–209.

Hill, J. (2011), "Bayesian Nonparametric Modeling for Causal Inference," *Journal of Computational and Graphical Statistics*, 20, 217–240.

Holland, P. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960.

Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86, 4–29.

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

Imbens, G. W. and Lemieux, T. (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.

Karabatsos, G. and Walker, S. (2012), "Adaptive-modal Bayesian nonparametric regression," *Electronic Journal of Statistics*, 6, 2038–2068.

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

Lee, D. and Lemieux, T. (2010), "Regression Discontinuity Designs in Economics," *The Journal of Economic Literature*, 48, 281–355.

Li, F., Mattei, A., and Mealli, F. (2013), "Bayesian Inference for Regression Discontinuity Designs with Application to the Evaluation of Italian University Grants," Tech. rep., Department of Statistics, Duke University.

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

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9." *Annals of Agricultural Science; Translated in Statistical Science 1990*, 5, 465–472.

Robins, J. and Greenland, S. (1996), "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91, 456–458.

Rosenbaum, P. (2002), *Observational Studies (2nd Ed.)*, New York: Springer-Verlag.

Rosenbaum, P. and Rubin, D. (1983), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study With Binary Outcome," *Journal of the Royal Statistical Society, Series B*, 45, 212–218.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

— (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58.

— (1990), "Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science*, 5, 472–480.

— (2008), "For objective causal inference, design trumps analysis," *The Annals of Applied Statistics*, 808–840.

Sales, A. and Hansen, B. (2013), "Analyzing Regression-Discontinuity Designs as Randomized Experiments," Tech. rep., Department of Statistics, University of Michigan.

Shadish, W., Galindo, R., Wong, V., Steiner, P., and Cook, T. (2011), "A randomized experiment comparing random and cutoff-based assignment." *Psychological Methods*, 16, 179.

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

Trochim, W. (1984), *Research design for program evaluation: The regression-discontinuity approach*, Newbury Park, CA: Sage.

Wong, V., Steiner, P., and Cook, T. (2013), "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods," *Journal of Educational and Behavioral Statistics*, 38, 107–141.

Appendix: Bayesian Nonparametric Modeling of a Fuzzy RD Design

Our model (3) can be extended for the analysis of data arising from a fuzzy RD design, involving imperfect treatment compliance among the subjects. This extension involves estimating bounds of the causal effect τ, with respect to a range of violations of the LM and ER assumptions (Angrist et al., 1996). There are two compelling motivations to using this approach for a fuzzy design. First, the causal effect estimator (1) does not identify the subset of compliers around the cutoff r_0 (e.g., Angrist et al., 1996). Thus, it is not possible to empirically verify the accurate estimation of $\Pr[i \text{ is a complier}|r_0]$ in the denominator of the estimator (1). Second, the identifying local monotonicity (LM) and local exclusion restriction (LER) assumptions are questionable, and not even empirically verifiable in such a design. Therefore, it seems more useful to estimate upper and lower bounds for

the effect τ, instead of obtaining specific estimates of the causal effect τ under such explicit unverifiable assumptions (Robins & Greenland, 1996; Balke & Pearl, 1997). The methods for estimating such bounds are described as follows.

Recall that under LS, we may define for all $\epsilon > 0$ of the potential treatment receipt outcomes $(T_i(A_{r_0}^{(r_0-\epsilon)}), T_i(A_{r_0}^{(r_0+\epsilon)}))$, a *complier* as a subject with $(T_i(1), T_i(0)) = (1, 0)$; but also we may define a *defier* as a subject with $(T_i(1), T_i(0)) = (0, 1)$, a *never-taker* as a subject with $(T_i(1), T_i(0)) = (0, 0)$, and an *always-taker* as a subject with $(T_i(1), T_i(0)) = (1, 1)$ (Imbens & Lemieux, 2008). Also, recall that LS and LER, together, imply that there exists some $\epsilon > 0$ such that for all subjects $i \in \mathcal{L}_{n_0}^{\epsilon}(r_0)$, the following relation holds between ITT causal effect A on T, and the causal effect $Y_i(1) - Y_i(0)$ of T on Y: $ITT_i = Y_i(1, T_i(0)) - Y_i(0, T_i(0)) = Y_i(T_i(1)) - Y_i(T_i(0)) = (Y_i(1) - Y_i(0))(T_i(1) - T_i(0))$. Then by using the preceding formula, it is easy to show that for the same set of subjects, the ITT_i effect for a complier coincides with the causal effect $Y_i(1) - Y_i(0)$ of T on Y; the ITT_i effect for a defier is equal to the negative causal effect of T on Y, i.e., $-(Y_i(1) - Y_i(0))$; and for either a never-taker or an always-taker, the ITT_i effect and the causal effect of T on Y are equal zero. Also, the average ITT effect of A on T, and the average causal effect of T on Y, have the relationship:

$$\begin{aligned}
&\mathrm{E}[Y_i(1, T_i(1)) - Y_i(0, T_i(0))|r_0] = \mathrm{E}[(Y_i(1) - Y_i(0))(T_i(1) - T_i(0))|r_0] \\
&= \mathrm{E}[Y_i(1) - Y_i(0)|r_0,\ T_i(1) - T_i(0) = 1]\,\Pr[\ T_i(1) - T_i(0) = 1|r_0] \\
&-\mathrm{E}[Y_i(1) - Y_i(0)|r_0,\ T_i(1) - T_i(0) = -1]\,\Pr[T_i(1) - T_i(0) = -1|r_0] \\
&= \mathrm{E}[Y_i(1) - Y_i(0)|r_0 \text{ and } i \text{ is a complier}]\,\Pr[i \text{ is a complier}|r_0] \\
&-\mathrm{E}[Y_i(1) - Y_i(0)|r_0 \text{ and } i \text{ is a defier}]\,\Pr[i \text{ is a defier}|r_0],
\end{aligned}$$

with $\Pr[i \text{ is a complier}|r_0] + \Pr[i \text{ is a defier}|r_0] + \Pr[i \text{ is a never-taker}|r_0] + \Pr[i \text{ is an always-taker}|r_0] = 1$ (e.g., Angrist et al. 1996).

Given these facts, we may estimate the bounds of the causal effect of τ with respect to a plausible range of violations of the LER assumption. We have that under assumptions RD, LS and LM, $\mathrm{E}[T_i(1) - T_i(0)|r_0] > 0$, and also the causal effect of T on Y at r_0 becomes:

$$\begin{aligned}
\tau(c) \;=\; & \mathrm{E}[Y_i(1, T_i(1)) - Y_i(0, T_i(0))|r_0]/c \\
& + \frac{1-c}{c}\mathrm{E}[Y_i(1, T_i) - Y_i(0, T_i)|r_0 \text{ and } i \text{ is a non-complier}],
\end{aligned}$$

where $c = \Pr[i \text{ is a complier}|r_0]$, and $(1 - c) = \Pr[i \text{ is an always-taker}|r_0] + \Pr[i \text{ is a never-taker}|r_0] < 1/2$, since $\Pr[i \text{ is a defier}|r_0] = 0$ under LM. Then under the same assumptions, and for any functional $h\{\cdot\}$ of Y of interest, these bounds are estimated by evaluating:

$$\begin{aligned}
\widehat{\tau}_h(c) \;=\; & [\mathrm{E}_n(h\{y\}|r_0, 1) - \mathrm{E}_n(h\{y\}|r_0, 0)]/c \\
& + \frac{1-c}{c}\mathrm{E}[h\{Y_i(1, T_i)\} - h\{Y_i(0, T_i)\}|r_0 \text{ and } i \text{ is a non-complier}]
\end{aligned}$$

over a range of plausible hypothetical values for the probability $c = \Pr[i \text{ is a complier}|r_0] \geq 1/2$, given a hypothetical ITT effect $\mathrm{E}[h\{Y_i(1, T_i)\} - h\{Y_i(0, T_i)\}|r_0$ and i is a non-complier] for the always-takers and never-takers at r_0, and given

posterior predictive estimates $\mathrm{E}_n(h\{y\}|r_0, a)$ $(a = 0, 1)$ of the Bayesian model (3).

We may also estimate the bounds of the causal effect of τ with respect to a plausible range of violations of the LM assumption. Let

$$\lambda = \Pr[i \text{ is a defier}|r_0]/\{\Pr[i \text{ is a complier}|r_0] - \Pr[i \text{ is a defier}|r_0]\} = c/(c-d).$$

Then under RD, LS, and LER, and for any functional $h\{\cdot\}$ of Y of interest, these bounds are estimated by evaluating:

$$\begin{aligned}
\widehat{\tau}_h(c,d) &= (1+\lambda)\mathrm{E}[h\{Y_i(1)\} - h\{Y_i(0)\}|r_0 \text{ and } i \text{ is a complier}] \\
&\quad -\lambda\mathrm{E}[h\{Y_i(1)\} - h\{Y_i(0)\}|r_0 \text{ and } i \text{ is a defier}] \\
&= (1+\lambda)([\mathrm{E}_n(h\{y\}|r_0, 1) - \mathrm{E}_n(h\{y\}|r_0, 0)]/\Pr[i \text{ is a complier } |r_0]) \\
&\quad -\lambda([\mathrm{E}_n(h\{y\}|r_0, 1) - \mathrm{E}_n(h\{y\}|r_0, 0)]/\Pr[i \text{ is a defier } |r_0]),
\end{aligned}$$

over a range of plausible hypothetical values of the probabilities $0 < \Pr[i \text{ is a defier}|r_0] < \Pr[i \text{ is a complier}|r_0] \leq 1$ (Angrist et al., 1996), given the posterior predictive estimates $\mathrm{E}_n(h\{y\}|r_0, a)$ $(a = 0, 1)$ of our model (3).