JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **31**: 1026–1047 (2016)
Published online 19 August 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2481

BAYESIAN FUZZY REGRESSION DISCONTINUITY ANALYSIS AND RETURNS TO COMPULSORY SCHOOLING

SIDDHARTHA CHIB[a]* AND LIANA JACOBI[b]
[a] *Olin Business School, Washington University in St Louis, MO, USA*
[b] *Department of Economics, University of Melbourne, Victoria, Australia*

SUMMARY
This paper is concerned with the use of a Bayesian approach to fuzzy regression discontinuity (RD) designs for understanding the returns to education. The discussion is motivated by the change in government policy in the UK in April of 1947, when the minimum school leaving age was raised from 14 to 15—a change that had a discontinuous impact on the probability of leaving school at age 14 for cohorts who turned 14 around the time of the policy change. We develop a Bayesian fuzzy RD framework that allows us to take advantage of this discontinuity to calculate the effect of an additional year of education on subsequent log earnings for the (latent) class of subjects that complied with the policy change. We illustrate this approach with a new dataset composed from the UK General Household Surveys. Copyright © 2015 John Wiley & Sons, Ltd.

Received 13 June 2013; Revised 6 June 2015

1. INTRODUCTION

This paper is concerned with the use of a Bayesian approach to fuzzy regression discontinuity (RD) designs for understanding the returns to education. RD designs originate in Thistlethwaite and Campbell (1960) and Campbell (1969) and provide an environment for finding the causal effect of a binary treatment x on an outcome y. RD designs come in two flavors, called sharp and fuzzy RD designs. Illustrations and theory appear, for example, in Angrist and Lavy (1999), Hahn et al. (2001), Imbens and Lemieux (2008), Lee and Lemieux (2010), Frandsen et al. (2012), Calonico et al. (2014) and Cattaneo et al. (2015). Hahn et al. (2001) show that the fuzzy RD design is closely connected to an instrumental variable (IV) problem with a discrete instrument, and that the treatment effect of interest is a version of the local average treatment effect or complier average treatment effect (CATE).

In recent work, Chib and Greenberg (2014) draw on the latter interpretation and develop a fully Bayesian approach for fuzzy RD designs that is based on the principal stratification paradigm (Imbens and Rubin, 1997; Frangakis and Rubin, 2002; Chib and Jacobi, 2008). The essential idea is to model the imperfect compliance with the policy in place around the discontinuity point by a discrete confounder variable that represents subject types. Instead of two potential outcomes, as in the existing frequentist approach, the modeling involves four potential outcomes and leads to a mixture model for outcomes, averaged over subject types, and a Bayesian estimator of the RDD CATE under assumptions that are detailed below. Chib and Greenberg (2014) provide points of further contrast between this approach and the frequentist approach to the fuzzy RD design, and discuss the sampling performance of the Bayesian RDD CATE estimator in relation to the corresponding frequentist estimators.

In this paper we adapt this Bayesian approach to estimate the effects of changes in compulsory schooling laws on labor market outcomes. Specifically, we are interested in the effect of the change in government policy in the UK in April of 1947 when the minimum school leaving age was raised from 14 to 15. The question is what effect this had on the level of schooling and, subsequently, what effect the change in schooling had on labor market outcomes. In order to answer the second question

* Correspondence to: Siddhartha Chib, Olin Business School, Campus Box 1133, Washington University in St Louis, 1 Brookings Drive, St Louis, MO 63130, USA. E-mail: chib@wustl.edu

within the RD paradigm, it is first necessary to isolate what aspect of schooling was discontinuously impacted by the change in policy. The nature of the policy change suggests that the most likely impact would have been to lower the probability of leaving school at age 14 and to increase that of leaving at age 15, for cohorts who turned 14 around the time of the policy change, and that the likely impact on these probabilities would have dissipated in a few years once the new policy became the new norm. Note that it is less likely, though possible, that the policy affected the probabilities of leaving school at a later age, or even to suppose that it affected the entire schooling distribution. For example, the latter was assumed in Harmon and Walker (1995), Oreopoulos (2006, 2008) and Devereux and Hart (2010), and x was taken to be educational attainment, a continuous variable, whose effect on subsequent wages was calculated by two-stage least squares (2SLS) methods based on subjects both far and near (in years) away from the policy change. The assumption in these studies is that the instrument had an impact on schooling both across many years of schooling attainment and across time.

The approach we employ in this paper more closely exploits the nature and timing of the policy change, and the analysis is closely aligned to that of the fuzzy RD design. We begin by considering students who turned 14 around the time of the policy change, because these are the subjects for whom this policy change was directly relevant. For such a person, we model whether the person would drop out at 14 or the next year at 15. We are not interested in the people who were 14 around April of 1947 and dropped out in 1949 or later (at the age of 16 or higher) because they are a different subgroup of people (intuitively, for that group, the drop-out patterns cannot be as clearly connected to the change in policy). Thus, in our approach, we define the treatment x to be one if the subject, who turned 14 around April 1947, dropped out of school at age 15, and zero if the subject dropped out of school at age 14. We show below that the probability distribution of this x variable was discontinuously affected by the policy change near the time of the policy change. We then determine the effect of this x, which implies an additional year of compulsory schooling, on subsequent (log) earnings.

In our problem, compliance with the school-leaving age policy was less than perfect. Some students under the old policy regime left school at age 15 and some students under the new policy regime left school at age 14. We deal with this divergence between the new policy indicator variable and x by modeling x in terms of two covariates: the school-leaving policy in effect (an observed quantity) and an individual-specific variable that represents subject type (an unobserved quantity). In effect, we suppose that the person who was 14 in 1947, and drops out at 15, is doing so because either he is complying with the new policy or because he is always going to drop out at 15. Likewise, the person who was 14 in 1947, and dropped out at 14, is doing so because he is either complying with the existing policy or because he is always going to drop out at 14.

There are thus three types of subjects—*compliers* (C), *always drop out 14* (AD14), and *always drop out 15* (AD15)—and four potential outcomes—$(y_{0,\mathrm{C}}, y_{1,\mathrm{C}})$ for the compliers, and $(y_{0,\mathrm{AD14}}, y_{1,\mathrm{AD15}})$ for the remaining types. In our modeling we specify the distributions of these four potential outcomes under the assumptions that the distribution of the complier outcomes depends on both the forcing variable z and observed confounders $\boldsymbol{w}$, those of $(y_{0,\mathrm{AD14}}, y_{1,\mathrm{AD15}})$ depend only on $\boldsymbol{w}$, and that the marginal distributions of z, $\boldsymbol{w}$ and types are smooth around the cut-off. With a properly reasoned prior distribution on the parameters we are then able to find the effect on later earnings for subjects that are comparable, namely the compliers with the policy in place, at the time of the policy change. In our analysis, we carefully construct the form of the prior distribution and variables that go into $\boldsymbol{w}$. Specifically, in formulating our prior, we adopt a novel approach that combines information from auxiliary samples and simulations of the prior and outcome models and, in specifying the potential outcome models, we consider several alternative formulations and report results from the best-supported formulation in terms of the marginal likelihood/Bayes factor comparisons.

The rest of the paper is organized as follows. In Section 2 we first provide additional details of our empirical setting and then describe the fuzzy RD approach with compliance types that is relevant for our setting. In Section 3 we provide the inferential approach and a discussion of the marginal likelihood computation by the method of Chib (1995). Section 4 provides evidence on the performance of our approach in simulated data experiments and Section 5 deals with the analysis of our real data. The paper concludes in Section 6.

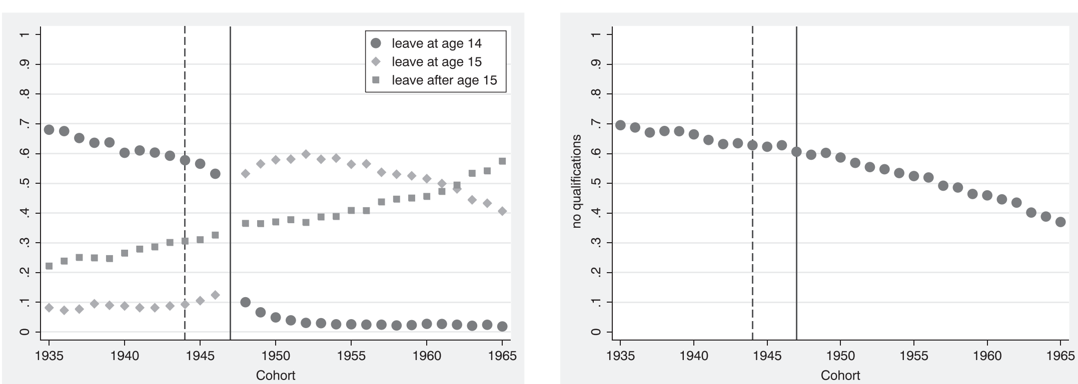
2. BAYESIAN FUZZY RD APPROACH WITH TYPES

2.1. The Setting

The 1947 increase in the school-leaving minimum age was implemented as part of the 1944 Education Act. This enactment introduced a system of free secondary education for all students within a tripartite system that consisted of grammar schools, secondary modern schools and technical schools. Admission to the more academically oriented grammar schools was based on an ability test administered at age 11. Most of those students who failed to secure admission to a grammar school attended a secondary modern school, while a small proportion of those transferred to a technical school at age 12 or 13 (Halsey *et al.*, 1980). The technical schools mostly provided lower=level academic education. Some of the students in these schools were those that had left school at the minimum school leaving age. Thus the 1947 policy change primarily affected the level of schooling of students at the lower end of the schooling distribution. We can see this reflected in the new data that we have hand composed from various iterations of the UK General Household surveys. Figure 1(a) shows the school-leaving behavior within cohorts of students who turned 14 in the years between 1935 and 1965. The three graphs present the proportion of students within each cohort that left school at age 14, 15 or beyond. The solid vertical line in 1947 refers to the increase in the school-leaving age and the dashed line to the year in which the tripartite schooling system and access to free secondary education was introduced.

It is clear from the figure that not all students complied with the new school-leaving minimum age. It is also clear that within the cohort of students who had turned 14 between 1946 and 1948 there was a large jump in the proportion of students who left school at age 15. Conversely, one sees that within this period there was a sharp decline in the proportion of students who left at age 14. For example, in the cohort of students who turned 14 in 1946, over 50% dropped out of school at age 14, and only around 15% at age 15. In comparison, over 50% of the students that were 14 in 1948 dropped out at age 15, while roughly 10% dropped out at age 14. In contrast, the graph indicates a smooth trend for the proportion of subjects leaving school after the age of 15.

Another aspect of the policy change was that it had little immediate effect on student qualifications. Figure 1(b) shows a smooth downward trend in the proportion of students with no qualifications



(a) (b)

Figure 1. Schooling and qualifications of students leaving school at age 14, age 15 and beyond age 15 in the 1935–1965 cohorts: (a) proportion leaving school age 14, 15 and beyond; (b) Proportion with no qualifications

within each cohort. At the time of the policy change, students could take two external examinations (the school certificate and the higher school certificate), which were usually taken by grammar schools students at age 16 and age 18 and replaced by the general certificate for education (O-level, A-level) in 1951. While students completing the full course of secondary schooling (4 years before the policy change and 5 years afterwards) could obtain a school-leaving certificate, secondary modern students often left school at the earliest possible moment without taking a school-leaving certificate (Halsey *et al.*, 1980).

Given these facts, it appears that the change in the minimum school-leaving age had a discontinuous impact on the probability of leaving school at age 14 for cohorts who turned 14 around the time of the policy change. This leads us to develop a framework for calculating the average causal effect of an additional year of compulsory schooling, for those students who left school at age 15 rather than at age 14, on later earnings, for subjects that were compliant with the policy change. In order to ensure that students in the analysis faced a comparable secondary schooling system and similar labor market conditions at entry, we focus our main analysis on students who turned 14 within a 3-year window around the policy change (turning 14 between October 1945 and September 1948). We also provide a secondary analysis for a wider 4-year window.

2.2. Assumptions

Let T_0 denote April 1933, the time point exactly 14 years prior to the policy change. Let z denote the distance in months of the birth month and year relative to T_0. For example, a person born in April 1933 has a z value of 0; someone born in January of 1934 has a z value of 9. Let τ denote the value 0. Now, under these definitions, a person with a z value less than 0 would have turned 14 before the policy change and, therefore, would have been exposed to the old regime, whereas a person with a z value equal to or above 0 would have turned 14 under the new policy. Thus the policy that a subject faces is given by $I(z \geq \tau)$, which is 1 for the new policy and 0 for the old policy. In the language of the RD designs, z is the forcing variable and τ is the threshold.

The treatment variable x in our problem is binary. In particular, for subjects who turned 14 between October 1945 and September 1948 (i.e. for subjects with a z value ranging from -18 to 17), we let

$$x = \begin{cases} 1 & \text{if subject dropped out of school at age 15} \\ 0 & \text{if subject dropped out of school at age 14} \end{cases} \tag{1}$$

Finally, we let the outcome y denote subsequent log earnings. The goal of the analysis to calculate the effect of this x (i.e. an extra year of education) on subsequent earnings.

There are two central complications in answering this questions. The first is that x is not a deterministic function of the policy in place, $I(z \geq \tau)$, as would occur in a sharp RD design. Some students under the old policy regime left school at age 15 and some students under the new policy regime left school at age 14. Even if one assumes that the policy change was random, the divergence between the policy $I(z \geq \tau)$ and x implies that x cannot be viewed as randomly assigned. The second complication is that there are observed confounders, which we label as $\boldsymbol{w} : k_w \times 1$, which likely affect both the outcome and the treatment x.

In dealing with these complications, we make the following key assumptions that are standard in the RD framework.

Assumption 1. The distribution of z (the forcing variable) is smooth on either side of τ.

Assumption 2. The distribution of the observed confounders $\boldsymbol{w}$ is smooth on either side of τ and its marginal impact on the potential outcomes is the same in each policy regime.

Under these assumptions, any observed discontinuity in the relation between y and z, after controlling for the observed confounders $\boldsymbol{w}$ and unobserved confounders, can be attributed to the effect of x on the outcomes.

As in Chib and Greenberg (2014), we deal with unobserved confounders by taking recourse to the principal stratification (Frangakis and Rubin, 2002) framework. In particular, let $x_l = j$ denote the potential treatment intake under assignment $I(z \geq \tau) = l$, where $j = 1$ refers to leaving school at age 15. Then our third assumption is about the nature of the unobserved confounder.

Assumption 3. The unobserved confounder is a discrete random variable that represents subject type. A subject can be of three types—a complier, always drop out 14 or always drop out 15—defined as follows:

$$\begin{array}{ll} \text{complier (C)} & \text{if } x_l = l \\ \text{always drop out 14 (AD14)} & \text{if } x_0 = x_1 = 0 \\ \text{always drop out 15 (AD15)} & \text{if } x_0 = x_1 = 1 \end{array}$$

In other words, a *complier* is a person who complies with the policy in place, an *always drop out 14* is a person who always leaves school at age 14 regardless of the policy in place, and an *always drop out 15* is a person who always leaves at age 15 regardless of the policy in place. We employ the standard monotonicity assumption and rule out defiers.

In the sequel we denote subject type by the discrete random variable s, which takes the values $k \in \{\text{C}, \text{AD14}, \text{AD15}\}$. Our assumption about the a priori distribution of types is summarized in the next assumption.

Assumption 4. Subject types are distributed smoothly around τ with unknown distribution $\Pr(s = k) = q_k$, where $q_{\text{C}} + q_{\text{AD14}} + q_{\text{AD15}} = 1$.

The model for the type probability in Assumption 4 encapsulates the assumption that the distribution of type around τ is independent of z. In the frequentist RD literature, types are not modeled explicitly, but the implied distribution of type can be derived from the distribution of the treatment assumed in that literature. The assumption that the distribution of x is free of z around τ (see, for example, Assumption 3 in Hahn et al. (2001) implies that the type probabilities are free of z, consistent with our latter assumption. Note that we could let our distribution of type depend on $\boldsymbol{w}$ but we do not model this dependence because of to data limitations.

2.3. Outcome Modeling

It follows from the definition of subject type and treatment state that four potential outcomes are possible. We indicate these potential outcomes by

$$\{y_{0,\text{C}},\ y_{1,\text{C}},\ y_{0,\text{AD14}},\ y_{1,\text{AD15}}\}$$

where $y_{j,\text{C}}$ is the outcome for compliers when $x = j$, and the other two outcomes are those for subject types AD14 and AD15 in the states $x = 0$ and $x = 1$, respectively. As in Chib and Greenberg (2014), we model the distributions of these outcomes by noting that for compliers

$$\Pr(x = 0 | z < \tau, s = \text{C}) = 1$$

and

$$\Pr(x = 1 | z \geq \tau, s = \text{C}) = 1$$

or, in other words, that for compliers the policy regime and the value of x agree, as in the sharp RD model, and one can define the potential outcomes in terms of the forcing variable z and the observed confounders $\boldsymbol{w}$. For the other two types of subjects, namely types AD14 and AD15, we assume that the outcome models satisfy an exclusion restriction with respect to z. We summarize these assumptions as follows.

Assumption 5. The marginal distributions of $(y_{0,\mathrm{C}},\ y_{1,\mathrm{C}})$ depend on both z and $\boldsymbol{w}$, whereas those of $(y_{0,\mathrm{AD14}},\ y_{1,\mathrm{AD15}})$ depend only on $\boldsymbol{w}$.

The latter assumption can be weakened because, as shown in Chib and Greenberg (2014), the mixture likelihood with z present in the marginal distributions of $(y_{0,\mathrm{AD14}},\ y_{1,\mathrm{AD15}})$ does not suffer from label switching, and the complier average treatment effect at the discontinuity is estimable. We enforce the exclusion restriction here, however, because the paucity of unique z values in our application makes it difficult to estimate the general case.

Another important point to note is that because the confounder in our case is discrete, the modeling entails four potential outcomes and a model of types, with no explicit modeling of the treatment. Instead, in Hahn et al. (2001), and the rest of the frequentist literature on fuzzy RD designs, the confounder is continuous and therefore the modeling is in terms of two potential outcomes, one for each level of the binary treatment, and a model of the binary treatment, with no explicit modeling of types. Furthermore, in the frequentist context, for identification purposes, the *difference* of these two potential outcomes (the subject level treatment effect) must be free of z around τ, for example, Assumption 3 of Hahn et al. (2001), which means that the two potential outcomes must have the *same* dependence on z. A reader of this paper points out that the latter assumption can be relaxed by letting this difference be continuous in z but, to the best of our knowledge, the consequences of this for current windowed IV estimators in finite-sample settings (with limited data around the cut-off and no possibility of getting any data on one of the potential outcomes at the cut-off) have not been examined in any simulation study. In our case, because the assumption about the confounder is different, z can play a different role in each of the four potential outcome models. Further discussion of this and related points of contrast between our modeling based on a discrete confounder, and the frequentist modeling based on a continuous confounder, is contained in Chib and Greenberg (2014).

We now specialize our assumptions by supposing that the potential outcomes for compliant subjects are generated parametrically as

$$\begin{aligned} y_{0,\mathrm{C}} &= g_{0,\mathrm{C}}(z) + \boldsymbol{w}'\boldsymbol{\beta}_{\mathrm{C}} + \varepsilon_{0,\mathrm{C}} \\ y_{1,\mathrm{C}} &= g_{1,\mathrm{C}}(z) + \boldsymbol{w}'\boldsymbol{\beta}_{\mathrm{C}} + \varepsilon_{1,\mathrm{C}} \end{aligned} \tag{2}$$

where $g_{j,\mathrm{C}}(z)$ are smooth functions of the forcing variable, and $\boldsymbol{\beta}_{\mathrm{C}} : k_w \times 1$ is the effect of $\boldsymbol{w}$ on the potential outcomes which is the same in the two treatment states following Assumption 2. For the idiosyncratic shocks $\varepsilon_{j,\mathrm{C}}$, we suppose that

$$\varepsilon_{j,\mathrm{C}} \sim t_\nu \left(0, \sigma^2_{j\mathrm{C}}\right)$$

Student-t random variables with $\nu > 2$ degrees of freedom, mean zero and variance $\nu\sigma^2_{j\mathrm{C}}/(\nu-2)$. The Student-$t$ assumption is a simple generalization of the Gaussian and one that we have found is better supported in our application. Furthermore, our benchmark assumption is that the functions $g_{j,\mathrm{C}}$ are linear in z:

$$\begin{aligned} g_{j,\mathrm{C}}(z) &= \beta_{j\mathrm{C}0} + \beta_{j\mathrm{C}1} z \\ &\stackrel{\text{def}}{=} (1, z)\boldsymbol{\beta}_{j,\mathrm{C}} \end{aligned} \tag{3}$$

and that these functions include an intercept term (therefore, no intercept is included in $\boldsymbol{w}$). More general models of $g_{j,\mathrm{C}}(z)$ can be specified as discussed in Chib and Greenberg (2014) but are difficult to estimate here given the limited support of the distribution of z in our empirical example. Therefore, each complier model is defined by the $(3 + k_w)$ parameters $\left(\boldsymbol{\beta}_{j\mathrm{C}}, \boldsymbol{\beta}_{\mathrm{C}}, \sigma^2_{j\mathrm{C}}\right)$.

Next, the outcome model of an AD14 subject, who is only observed in the $x = 0$ state, is defined similarly in terms of a Student-t distribution as

$$y_{0,\mathrm{AD14}} = (1, \boldsymbol{w}')\boldsymbol{\beta}_{0,\mathrm{AD14}} + \varepsilon_{0,\mathrm{AD14}} \tag{4}$$

where the 1 represents the intercept:

$$\varepsilon_{0,\mathrm{AD14}} \sim t_\nu\left(0, \sigma^2_{0,\mathrm{AD14}}\right)$$

depending on the $(k_w + 2)$ parameters $\left(\boldsymbol{\beta}_{0,\mathrm{AD14}}, \sigma^2_{0,\mathrm{AD14}}\right)$. Finally, the outcome model of an AD15 subject, who is only observed in the $x = 1$ state, is specified as

$$y_{1,\mathrm{AD15}} = (1, \boldsymbol{w}')\boldsymbol{\beta}_{1,\mathrm{AD15}} + \varepsilon_{1,\mathrm{AD15}} \tag{5}$$

where

$$\varepsilon_{1,\mathrm{AD15}} \sim t_\nu\left(0, \sigma^2_{1,\mathrm{AD15}}\right)$$

defined by the $(k_w + 2)$ parameters $\left(\boldsymbol{\beta}_{1,\mathrm{AD15}}, \sigma^2_{1,\mathrm{AD15}}\right)$.

It should be appreciated that these parametric assumptions are not necessary in our treatment of this problem but we make these assumptions to learn more from the data, though at the risk of possible misspecification. We use marginal likelihoods below to compare alternative models in an attempt to minimize some of this risk, at least within the parametric class.

2.4. RDD Complier Average Treatment Effect

Under our assumptions, the relevant treatment effect is the average treatment effect for the compliant subjects at the point of discontinuity τ. This effect, which we refer to as the RDD CATE (short for regression discontinuity design complier average treatment effect), is given by

$$\begin{aligned}\text{RDD CATE} &= \mathbb{E}[y_1 | z = \tau, s = \mathrm{C}, \boldsymbol{w}] - \mathbb{E}[y_0 | z = \tau, s = \mathrm{C}, \boldsymbol{w}] \\ &= g_{1,\mathrm{C}}(\tau) - g_{0,\mathrm{C}}(\tau)\end{aligned}$$

and is identified under Assumptions 1–5. This is simply because the model we have specified is a mixture model that is not subject to label switching (since the component distributions are different and arise in different subsamples of the data), a point that becomes more clear in the following section. It should be emphasized that the RDD CATE that we estimate is at the discontinuity because this is the only point at which the compliers in the two treatment states are comparable. A reader questioned whether (under our parametric assumption) it was possible to find the treatment effect at other values of z (say for $z < \tau$), but such a calculation would require the (unwarranted) assumption that the function $g_{1,\mathrm{C}}$ holds on the support on which it was not estimated.

An important point to note is that, because $\tau = 0$ in the way we have defined the forcing variable z, the RDD CATE simplifies to

$$\text{RDD CATE} = \beta_{1\mathrm{C}0} - \beta_{0\mathrm{C}0} \tag{6}$$

We take advantage of this fact in formulating a prior distribution that is 'neutral' about the a priori expected size of the RDD CATE, as discussed below.

3. INFERENCE

The model we have described is a mixture model once averaged over the unknown subject types. To see this, it is useful to think of a (2×2) table in which the rows indicate the policy epochs $I(z < \tau)$ and $I(z \geq \tau)$ and the columns indicate the treatment levels $x = 0$ and $x = 1$. In each cell, which we henceforth indicate by (lj), one can now display the possible subject types, as shown in Table I. Specifically, an individual in cell (00), in the old policy regime who leaves at age 14, can be either a complier or AD14; a person in cell (10), the new policy regime who leaves school at age 14, is of type AD14; a subject in cell (01), the old policy regime who leaves at age 15, is of type AD15; while a person in cell (11), the new regime who leaves at age 15, can be either a complier or AD15. Succinctly, these possible subject types by cell can be delineated by the sets

$$K_{00} = \{\mathrm{C}, \mathrm{AD14}\}, \; K_{10} = \{\mathrm{AD14}\}, \; K_{01} = \{\mathrm{AD15}\}, \; K_{11} = \{\mathrm{C}, \mathrm{AD15}\}$$

and the possible types in each treatment state $x = j$ by

$$K_0 = \{\mathrm{C}, \mathrm{AD14}\}, \; K_1 = \{\mathrm{C}, \mathrm{AD15}\}$$

Thus the distribution of the outcomes in each cell is a mixture of distributions (though the mixture in the (1,0) and (0,1) cells is a one-component mixture).

Table I also shows that this model allows inference about types given the data, without the typical label-switching problem of mixture models, because of the assumption that z does not affect the distribution of the subjects in the (10) and (01) cells. Thus, given the data, the component distributions in the (00) cells cannot be permuted without changing the probability distribution of the data in that cell. Similarly, the component distributions in the (11) cell cannot be permuted. Finally, the complier models appear in different subsamples of the data, namely the (00) and (11) cells, and can be inferred from the data. Thus it is possible to revise prior beliefs about the distribution of types in these four cells and to estimate the component models and, therefore, to learn about the RDD CATE.

Another facet of this model, again transparent from Table I, is that, given the policy state $I(z \geq \tau)$ and the subject type, the treatment x is deterministic. Define the covariate vector by

$$\boldsymbol{v}_k' = \begin{cases} (1, z, \boldsymbol{w}') & \text{if} \quad k = \mathrm{C} \\ (1, \boldsymbol{w}') & \text{if} \quad k = \mathrm{AD14}, \mathrm{AD15} \end{cases} \tag{7}$$

Also define the regression parameters by

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_{0\mathrm{C}}, \boldsymbol{\beta}_{1\mathrm{C}}, \boldsymbol{\beta}_{0,\mathrm{AD14}}, \boldsymbol{\beta}_{0,\mathrm{AD15}}\right)$$

the regression scale parameters by

$$\boldsymbol{\sigma}^2 = \left(\sigma^2_{0\mathrm{C}}, \sigma^2_{1\mathrm{C}}, \sigma^2_{0,\mathrm{AD14}}, \sigma^2_{0,\mathrm{AD15}}\right)$$

and the type probabilities by

Table I. Distribution of subject type by observed policy regime and schooling intake

	Schooling intake	
Policy indicator	$x = 0$ (leave age 14)	$x = 1$ (leave age 15)
$I(z < 0)$ (old policy)	C, AD14	AD15
$I(z \geq 0)$ (new policy)	AD14	C, AD15

DOI: 10.1002/jae

$$\boldsymbol{q} = (q_{\mathrm{C}}, q_{\mathrm{AD14}}, q_{\mathrm{AD15}})$$

and let

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{q}) \tag{8}$$

denote the full set of unknown parameters. Then the distribution of a particular observation $(y, x = j)$, given the policy regime l, and subject type $s = k$ relevant for that regime and treatment x, is

$$p(y, x = j|\boldsymbol{\theta}, s = k, I[z \geq \tau] = l) = t_\nu\left(y|\boldsymbol{v}_k'\boldsymbol{\beta}_{jk}, \sigma_{jk}^2\right), \; j = 0,1;\; k \in K_{lj};\; l = 0,1 \tag{9}$$

where, here and henceforth, the dependence on the covariates $(z, \boldsymbol{w})$ is not emphasized in the notation. This essentially shows that the cell, along with the subject type, determines which of the four outcome distributions is relevant. In practice, therefore, the first step in the inferential analysis is to sort the observations $(y_i, z_i, \boldsymbol{w}_i), i \leq n$, into groups that conform to this cell structure. This simplifies both the conceptual formulation of the inferential procedure and the coding of the Markov chain Monte Carlo (MCMC) algorithm. One can also let n_{lj} denote the number of observations in cell (lj) and let I_{lj} denote the indices of the observations in these cells.

3.1. Posterior Simulation

The posterior distribution of the outcomes and the latent subject types is proportional to

$$\begin{aligned}
&\pi(\boldsymbol{\theta})\prod_{i=1}^{n}\text{Gamma}\left(\lambda_i|\frac{\nu}{2},\frac{\nu}{2}\right)\\
&\prod_{i\in S_{00}}\left\{I[s_i=\mathrm{C}]\,q_{\mathrm{C}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{0\mathrm{C}},\lambda_i^{-1}\sigma_{0\mathrm{C}}^2\right)+I[s_i=\mathrm{AD14}]\,q_{\mathrm{AD14}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{AD14}}'\boldsymbol{\beta}_{0\mathrm{AD14}},\lambda_i^{-1}\sigma_{0\mathrm{AD14}}^2\right)\right\}\\
&\times\prod_{i\in S_{10}}I[s_i=\mathrm{AD14}]\,q_{\mathrm{AD14}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{AD14}}'\boldsymbol{\beta}_{0\mathrm{AD14}},\lambda_i^{-1}\sigma_{0\mathrm{AD14}}^2\right)\\
&\times\prod_{i\in S_{01}}I[s_i=\mathrm{AD15}]\,q_{\mathrm{AD15}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{AD15}}'\boldsymbol{\beta}_{1\mathrm{AD15}},\lambda_i^{-1}\sigma_{1\mathrm{AD15}}^2\right)\\
&\times\prod_{i\in S_{11}}\left\{I[s_i=\mathrm{C}]\,q_{\mathrm{C}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{1\mathrm{C}},\lambda_i^{-1}\sigma_{1\mathrm{C}}^2\right)+I[s_i=\mathrm{AD15}]\,q_{\mathrm{AD15}}\,\mathrm{N}\left(y_i|\boldsymbol{v}_{i\mathrm{AD15}}'\boldsymbol{\beta}_{1\mathrm{AD15}},\lambda_i^{-1}\sigma_{1\mathrm{AD15}}^2\right)\right\}
\end{aligned}$$

where $\pi(\boldsymbol{\theta})$ is the prior density and $\boldsymbol{\lambda} = \{\lambda_i\}$ are Gamma-distributed positive random variables that are introduced into the scale mixture of normals representation of the Student-t distribution. This is sampled easily by MCMC methods. The key steps are given next. At the outset, in each round of the MCMC algorithm one samples the subject types in each of the (00) and (11) cells. This sampling is from the two-point discrete mass distributions:

$$\Pr(s_i = \mathrm{C}|y_i, x_i = 0, \boldsymbol{\theta}) = \frac{q_{\mathrm{C}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{0\mathrm{C}},\sigma_{0\mathrm{C}}^2\right)}{q_{\mathrm{C}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{0\mathrm{C}},\sigma_{0\mathrm{C}}^2\right)+q_{\mathrm{AD14}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{AD14}}'\boldsymbol{\beta}_{0\mathrm{AD14}},\sigma_{0\mathrm{AD14}}^2\right)}, \; i \in I_{00} \tag{10}$$

and

$$\Pr(s_i = \mathrm{C}|y_i, x_i = 1, \boldsymbol{\theta}) = \frac{q_{\mathrm{C}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{1\mathrm{C}},\sigma_{1\mathrm{C}}^2\right)}{q_{\mathrm{C}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{C}}'\boldsymbol{\beta}_{1\mathrm{C}},\sigma_{1\mathrm{C}}^2\right)+q_{\mathrm{AD15}}t_\nu\left(y_i|\boldsymbol{v}_{i\mathrm{AD15}}'\boldsymbol{\beta}_{1\mathrm{AD15}},\sigma_{1\mathrm{AD15}}^2\right)}, \; i \in I_{11} \tag{11}$$

which are then assembled in an n-vector s whose components in one MCMC iteration might look like

$$\boldsymbol{s} = \left(\underbrace{c, c, n, \ldots, c}_{n_{00}\times 1}, \underbrace{n, n, n, \ldots, n}_{n_{10}\times 1}, \underbrace{a, a, a, \ldots, a}_{n_{01}\times 1}, \underbrace{c, a, a, \ldots, c}_{n_{11}\times 1} \right)$$

Based on this $\boldsymbol{s}$, let $I_s = \left(I_{s,0\mathrm{C}},\ I_{s,1\mathrm{C}},\ I_{s,0\mathrm{AD14}},\ I_{s,1\mathrm{AD15}}\right)$, where

$$I_{s,jk} = \{i : x_i = j \text{ and } s_i = k\},\ j = 0, 1;\ k \in K_j$$

are the indices of observations in each treatment state that are currently assigned to each possible type. If the prior of the type probabilities $\boldsymbol{q}$ is Dirichlet

$$\pi(\boldsymbol{q}) = \mathrm{Dir}(\boldsymbol{q}|a_{\mathrm{C},0}, a_{\mathrm{AD},0}, a_{\mathrm{AD15},0}), \tag{12}$$

with known hyperparameters $\left(a_{\mathrm{C},0},\ a_{\mathrm{AD},0},\ a_{\mathrm{AD15},0}\right)$ as formalized in the examples, then one samples the type probabilities from an updated Dirichlet distribution

$$\mathrm{Dir}(\alpha_{\mathrm{C}}, \alpha_{\mathrm{AD14}}, \alpha_{\mathrm{AD15}}) \tag{13}$$

where $\alpha_k = a_{k,0} + \sum_{i=1}^{n} I[s_i = k], k \in \{\mathrm{C}, \mathrm{AD14}, \mathrm{AD15}\}$, and then the $\boldsymbol{\lambda} = \{\lambda_i\}$ from the n posterior distributions

$$\lambda_i|y_i, x_i, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, s_i \sim \mathrm{Gamma}\left(\frac{\nu+1}{2}, \frac{\left(y_i - \boldsymbol{v}'_{ik}\boldsymbol{\beta}_{jk}\right)^2}{\sigma_k^2}\right),\ i \in \cup_{j=0,1} \cup_{k\in K_j} I_{s,jk} \tag{14}$$

If the prior of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}) = \prod_{j=0}^{1} \prod_{k\in K_j} \mathrm{N}(\boldsymbol{\beta}_{jk}|\boldsymbol{\beta}_{jk,0}, \boldsymbol{A}_{jk,0}), \tag{15}$$

for known hyperparameters $\left\{\boldsymbol{\beta}_{jk,0}, \boldsymbol{A}_{jk,0}\right\}$ again formalized in the examples, then the posterior sampling of $\boldsymbol{\beta}_{jk}$ is from

$$\mathrm{N}(\hat{\boldsymbol{\beta}}_{jk}, \boldsymbol{A}_{jk}),\ j = 0, 1;\ k \in K_j \tag{16}$$

where

$$\hat{\boldsymbol{\beta}}_{jk} = \boldsymbol{A}_{jk}\left(\boldsymbol{A}_{jk,0}^{-1}\boldsymbol{\beta}_{jk,0} + \sigma_{jk}^{-2} \sum_{i\in I_{s,jk}} \lambda_i \boldsymbol{v}_{ik} y_i\right)$$

and

$$\boldsymbol{A}_{jk} = \left(\boldsymbol{A}_{jk,0}^{-1} + \sigma_{jk}^{-2} \sum_{i\in I_{s,jk}} \lambda_i \boldsymbol{v}_{ik}\boldsymbol{v}'_{ik}\right)^{-1}$$

In the last step, if the prior of $\boldsymbol{\sigma}^2$ is the inverse-gamma density:

$$\pi(\boldsymbol{\sigma}^2) = \prod_{j=0}^{1} \prod_{k \in K_j} \mathrm{IG}\left(\sigma_{jk}^2 | \frac{\nu_{jk,0}}{2}, \frac{\delta_{jk,0}}{2}\right) \tag{17}$$

then the posterior sampling of σ_{jk}^2 is from

$$\mathrm{IG}\left(\sigma_{jk}^2 | \frac{\nu_{kj,0} + n_{jk}}{2}, \frac{\delta_{kj,0} + \sum_{i \in I_{s,jk}} \lambda_i \left(y_i - \boldsymbol{v}_{ik}' \boldsymbol{\beta}_{jk}\right)^2}{2}\right), \ j = 0, 1; \ k \in K_j \tag{18}$$

where n_{jk} is the cardinality of $I_{s,jk}$.

The above steps are repeated a large number of times. After the first 1000 burn-in draws, the subsequent 10,000 draws are used as drawings from the posterior distributions of the parameters.

We use the MCMC draws of the parameters to compute the difference $(\beta_{1\mathrm{C}0} - \beta_{0\mathrm{C}0})$ in equation (6) at each iteration of the MCMC algorithm. These differences are drawings from the posterior distribution of the RDD CATE:

$$\pi(\text{RDD CATE}|\boldsymbol{y}) = \int \pi(\text{RDD CATE}|\boldsymbol{y}, \beta_{1\mathrm{C}0}, \beta_{0\mathrm{C}0}) \pi(\beta_{1\mathrm{C}0}, \beta_{0\mathrm{C}0}|\boldsymbol{y}) \, d\beta_{1\mathrm{C}0} \, d\beta_{0\mathrm{C}0}$$

and are a surrogate for the posterior distribution of the RD treatment effect. These draws can be summarized in terms of the sample mean, sample standard deviation and sample quantiles (with the latter providing the posterior quantiles of the treatment effect).

3.2. Marginal Likelihood Computation

We also use the MCMC draws to calculate the model marginal likelihood by the method of Chib (1995). The marginal likelihood is used to choose the model with the best supported degrees of freedom (and set of covariates). The starting point of the calculation is the expression of the log-marginal likelihood in terms of the likelihood ordinate, prior ordinate and the posterior ordinate as

$$\ln m(\boldsymbol{y}, \boldsymbol{x}) = \ln f(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}|\boldsymbol{\theta}^*) + \ln \pi(\boldsymbol{\theta}^*) - \ln \pi(\boldsymbol{\theta}^*|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z})$$

where $\boldsymbol{\theta}^*$ refers to the posterior expectation of the parameters. In this expression, the likelihood ordinate can be evaluated directly as

$$\begin{aligned}
&\prod_{i \in S_{00}} \left\{q_{\mathrm{C}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{C}}' \boldsymbol{\beta}_{0\mathrm{C}}, \sigma_{0\mathrm{C}}^2\right) + q_{\mathrm{AD14}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{AD14}}' \boldsymbol{\beta}_{0\mathrm{AD14}}, \sigma_{0\mathrm{AD14}}^2\right)\right\} \\
&\times \prod_{i \in S_{10}} q_{\mathrm{AD14}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{AD14}}' \boldsymbol{\beta}_{0\mathrm{AD14}}, \sigma_{0\mathrm{AD14}}^2\right) \\
&\times \prod_{i \in S_{10}} q_{\mathrm{AD15}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{AD15}}' \boldsymbol{\beta}_{1\mathrm{AD15}}, \sigma_{1\mathrm{AD15}}^2\right) \\
&\times \prod_{i \in S_{11}} \left\{q_{\mathrm{C}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{C}}' \boldsymbol{\beta}_{1\mathrm{C}}, \sigma_{1\mathrm{C}}^2\right) + q_{\mathrm{AD15}} \, t_\nu\left(y_i | \boldsymbol{v}_{i\mathrm{AD15}}' \boldsymbol{\beta}_{1\mathrm{AD15}}, \sigma_{1\mathrm{AD15}}^2\right)\right\}
\end{aligned}$$

and the prior ordinate can also be computed directly as

$$\pi(\boldsymbol{\theta}^*) = \mathrm{Dir}(\boldsymbol{q}^*|a_{\mathrm{C},0}, a_{\mathrm{AD14},0}, a_{\mathrm{AD15},0}) \prod_{j=0}^{1} \prod_{k \in K_j} \mathrm{N}_p\left(\boldsymbol{\beta}^*_{jk}|\boldsymbol{\beta}_{jk,0}, \boldsymbol{A}_{jk,0}\right) \mathrm{IG}\left(\sigma^{2*}_{jk}|\frac{\nu_{jk,0}}{2}, \frac{\delta_{jk,0}}{2}\right)$$

Finally, the posterior ordinate can be evaluated using the following decomposition:

$$\pi\left(\boldsymbol{\theta}^*|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}\right) = \pi\left(\boldsymbol{\sigma}^{2^*}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}\right) \pi\left(\boldsymbol{\beta}^*|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}, \boldsymbol{\sigma}^{2^*}\right) \pi\left(\boldsymbol{q}^*|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}, \boldsymbol{\beta}^*, \boldsymbol{\sigma}^{2^*}\right)$$

where an estimate of the first ordinate is obtained as

$$\hat{\pi}\left(\boldsymbol{\sigma}^{2^*}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{W}, \mathbf{z}\right) = \prod_{j=0}^{1} \prod_{k \in K_j} \frac{1}{M} \sum_{g=1}^{M} \mathrm{IG}\left(\sigma^{2*}_{jk}|\frac{\nu_{kj,0} + n^{(g)}_{jk}}{2}, \frac{\delta_{kj,0} + \sum_{i \in I^{(g)}_{s,jk}} \lambda^{(g)}_i \left(y_i - \boldsymbol{v}'_{ik}\boldsymbol{\beta}^{(g)}_{jk}\right)^2}{2}\right)$$

using the draws on λ_i's and $\boldsymbol{\beta}_{jk}$ from the main run of the MCMC algorithm, an estimate of the second ordinate is obtained as

$$\prod_{j=0}^{1} \prod_{k \in K_j} \frac{1}{M} \sum_{g=1}^{M} \mathrm{N}\left(\boldsymbol{\beta}^*_{jk}|\hat{\boldsymbol{\beta}}^{(g)}_{jk}, \boldsymbol{A}^{(g)}_{jk}\right)$$

where

$$\hat{\boldsymbol{\beta}}^{(g)}_{jk} = \boldsymbol{A}^{(g)}_{jk}\left(\boldsymbol{A}^{-1}_{jk,0}\boldsymbol{\beta}_{jk,0} + \sigma^{-2*}_{jk} \sum_{i \in I^{(g)}_{s,jk}} \lambda^{(g)}_i \boldsymbol{v}_{ik} y_i\right)$$

and

$$\boldsymbol{A}^{(g)}_{jk} = \left(\boldsymbol{A}^{-1}_{jk,0} + \sigma^{-2*}_{jk} \sum_{i \in I^{(g)}_{s,jk}} \lambda^{(g)}_i \boldsymbol{v}_{ik}\boldsymbol{v}'_{ik}\right)^{-1}$$

where the draws are from a reduced run of the MCMC algorithm with σ^2_{jk} fixed at $\sigma^{2^*}_{jk}$ and, finally, the last ordinate is obtained as

$$\frac{1}{M} \sum_{g=1}^{M} \mathrm{Dir}\left(\boldsymbol{q}^*|a_{\mathrm{C},0} + n^{(g)}_{0\mathrm{C}} + n^{(g)}_{1\mathrm{C}}, a_{\mathrm{AD14},0} + n^{(g)}_{0,\mathrm{AD14}}, a_{\mathrm{AD15},0} + n^{(g)}_{1,\mathrm{AD15}}\right)$$

using a second reduced run of the MCMC algorithm with σ^2_{jk} fixed at $\sigma^{2^*}_{jk}$ and $\boldsymbol{\beta}_{jk}$ fixed at $\boldsymbol{\beta}^*_{jk}$.

4. PERFORMANCE IN SIMULATED DATA

In this section we present results from a small repeated-sampling simulation study to illustrate the sampling performance of the approach introduced in the previous sections. Much further study comparing the performance Bayesian approach and the frequentist estimators common in the literature is provided by Chib and Greenberg (2014). Here we employ a simple data-generating process (DGP) based on the framework from the previous section. The true values of the parameters are loosely based on the results from the analysis of the real data. The DGP has considerable heterogeneity in the regression coefficients across types and schooling intake, and in the variances across types. In addition, the design leads to outliers.

In simulating the data, for each individual we first generate the forcing variable z from a discrete uniform distribution on the integers from -24 to 24 (assuming a 4-year window), assigning the policy indicator a value of 1 if z in non-negative and zero otherwise. Next, we generate the w confounder for each subject by randomly drawing an integer between 85 and 95 to mimic the survey year variable in the data. In generating the outcomes, we assume that $\boldsymbol{\beta}_{0\mathrm{C}} = (4.5, -0.20, 0.03)$ and $\boldsymbol{\beta}_{1\mathrm{C}} = (4.55, 0.4, 0.03)$, which implies that the true treatment effect as defined in equation (6) is 0.05. For the remaining two types we assume that $\boldsymbol{\beta}_{\mathrm{AD14}} = (6.8, -0.02)$ and $\boldsymbol{\beta}_{\mathrm{AD15}} = (5.5, -0.04)$. The variances are set at $\sigma^2_{\mathrm{C}} = 0.1$, $\sigma^2_{\mathrm{AD14}} = 0.15$ and $\sigma^2_{\mathrm{AD15}} = 0.2$. We assume here and in our subsequent fitting that the variances $\sigma^2_{0\mathrm{C}}$ and $\sigma^2_{1\mathrm{C}}$ are equal. Finally, the type probabilities are set at $\Pr(s = \mathrm{C}) = 0.70, \Pr(s = \mathrm{AD14}) = 0.15 = \Pr(s = \mathrm{AD15}) = 0.15$.

We now generate 100 datasets for each combination of $n = 1500, 3000$ and $\nu = 5, 10, \infty$, where the last value reflects the normal model. For each dataset we find the Bayes posterior estimate of the model parameters, RDD CATE and the log marginal likelihood fitting our model under each of the three values of degrees of freedom.

Our results are summarized in Table II. The first block of results shows the value of the RDD CATE estimate averaged over the repeated samples as well as the value of the posterior standard deviations averaged over those samples. The second block of results shows the average absolute deviations of the posterior RDD CATE estimate from the true value of 0.05. As one can see, the true RDD CATE is well recovered under each model specification for each sample size. The true specification fit with the correct degrees of freedoms is always closest to the true value, although we observe only very small differences in the average treatment effects when the model with the incorrect ν is estimated. The true specification always displays the lowest average absolute deviation, although the results for the specifications with $\nu = 5$ and $\nu = 10$ are in many cases identical. While the treatment effect is recovered well under both sample sizes, we do observe a decrease in the precision of the RDD CATE estimate as the sample size is halved to from 3000 to 1500, with the latter reflecting a size closer to the samples used in our empirical analysis.

In our empirical example we observe larger differences between the treatment effects across the different model specifications. In this case model comparison based on the marginal likelihood criteria plays an important role in choosing the correct model and treatment effect estimate. In Table III we provide some summary results on the performance of the marginal likelihood criteria. We see that the true model is picked in all replications under the larger sample size, and in almost all cases (96% and above) under the smaller sample size.

Table II. RDD CATE estimates (posterior means and standard deviations) and absolute Deviations of RDD CATE estimates from true effect for compliers averaged over 100 replications

Data		Average RDD CATE (STD)			Average absolute deviations		
n	ν	$\nu = 5$	$\nu = 10$	$\nu = \infty$	$\nu = 5$	$\nu = 10$	$\nu = \infty$
3000	$\nu = 5$	0.047 (0.032)	0.047 (0.033)	0.045 (0.036)	0.026	0.027	0.031
	$\nu = 10$	0.048 (0.030)	0.047 (0.030)	0.046 (0.031)	0.025	0.025	0.026
	$\nu = \infty$	0.049 (0.028)	0.049 (0.028)	0.050 (0.028)	0.022	0.021	0.020
1500	$\nu = 5$	0.051 (0.046)	0.050 (0.047)	0.044 (0.050)	0.035	0.035	0.038
	$\nu = 10$	0.056 (0.043)	0.055 (0.043)	0.055 (0.044)	0.036	0.035	0.036
	$\nu = \infty$	0.054 (0.040)	0.052 (0.040)	0.053 (0.040)	0.037	0.035	0.035

DOI: 10.1002/jae

Table III. Results on model comparison: number out of 100 replications that model is preferred based on the marginal likelihood criterion when the data are generated under the degrees of freedom as specified in the first column

Data	Number of replications model is preferred (out of 100)					
	$n = 3000$			$n = 1500$		
	$\nu = 5$	$\nu = 10$	$\nu = \infty$	$\nu = 5$	$\nu = 10$	$\nu = \infty$
$\nu = 5$	100	0	0	96	4	0
$\nu = 10$	0	100	0	2	98	0
$\nu = \infty$	0	0	100	0	2	98

5. ANALYSIS OF RETURNS TO COMPULSORY SCHOOLING

5.1. Data

The data for our analysis are new and are constructed by us from information contained in the UK General Household Surveys (UKGHS), a panel of cross-section surveys carried out by the Social Survey Division of the Office for National Statistics (ONS) since 1971. Following Devereux and Hart (2010) we exclude the pre-1979 surveys to avoid problems due to a different reporting scheme of earnings. For the analysis under the Bayesian type-based RDD analysis of returns to schooling we focus on subjects who turned 14 within a narrow (symmetric) window around the policy change. Our main sample is based on males who turned 14 within a 3-year window around April 1947, e.g. born between October 1931 and September 1934. In addition, we restrict our analysis to subjects from the 1986 to 1995 surveys and the 1998 survey since only these survey years contain information on both birth year and birth month of subjects needed for the construction of the forcing variable z. The remaining survey years only report age at the time of the survey. Based on the reported birth year and month we compute an individual's forcing variable z as the distance of the birth date in months to April 1933. As discussed in the model assumptions, this implies that an individual born in April 1933 has $z = 0$ and turns 14 in April 1947 when the new policy was introduced. In our 3-year window sample z thus ranges from -18 for an individual born in October 1931 to 17 for an individual born in September 1934. We also consider a sample based on a slightly wider 4-year window around the policy change that includes males born between April 1931 and March 1935, with z ranging from -24 to 23.

The binary schooling intake variable x_i for each subject is defined based on reported age when a subject left school. We assign a value of zero to the schooling intake if a subject reported leaving school at age 14, and a value of one if the subject reported leaving school at age 15. As discussed earlier, we exclude subjects from the sample who reported a school-leaving age above 15 as they were not affected by the policy change. For the definition of the earnings variable y_i we follow Devereux and Hart (2010) and construct a variable for real log weekly earnings based on the reported gross weekly earnings (including earnings from self-employment), deflated by the UK retail price index with base year 1998. We also construct an hourly wage variable based on the weekly earnings and the reported work hours. Further, to avoid retirement-related issues, we omit subjects aged 60 and above. As shown in Banks and Blundell (2005), many low-skilled men quit working before the age of 65 in this time period, leading to an employment rate of low-skilled men age 60–64 of about 40%. Following Devereux and Hart (2010) we also drop subjects whose weekly working hours is missing or above 84 hours.

It should be noted that our sample for the type-based RD analysis is different from previous work based on the UKGHS data, as: (i) subjects who turned 14 outside a narrow window around the policy change are excluded; (ii) we only include those for whom the policy regime can be determined;

DOI: 10.1002/jae

Table IV. Summary statistics for male sample

Variable					
Description	[short]	Mean	SD	Min.	Max.
Leaving school at age 15	[x]	0.53		0	1
Under new Policy	[$z \geq \mathbf{0}$]	0.52		0	1
Age	[age]	55.74	2.06	52	59
Married	[married]	0.85		0	1
Survey year[a]	[year]	88.87	2.11	86	94
Log weekly earnings[b]	[learn]	5.61	0.49	2.55	7.27
Log hourly wage[b]	[lwage]	1.92	0.47	0.03	3.56
Work hours		40.807	8.36	6	80

[a]Calender years are reported as (year − 1900).
[b]Real earnings in 1998 pounds.

and (iii) subjects leaving school beyond the age of 15 are excluded. Restricting our attention to males we obtain a sample of 1015 subjects aged between 52 and 59. Of these, 484 turn 14 under the old policy regime with 380 subjects leaving school at age 14 and 104 at age 15. From the 531 subjects who turned 14 under the new school-leaving minimum age of 15, 101 subjects dropped out at age 14 and the remaining 430 at age 15. Table IV provides some descriptive statistics for the sample.

The alternative sample based on the wider 4-year window with 1327 males exhibits almost identical summary statistics, with a slightly higher proportion of subjects under the new policy (0.55) and leaving school at 15 (0.55).

Given our focus on subjects within a narrow window around the policy change, a respondent's age and the survey year when earnings are reported are highly correlated. By construction, we only observe subjects from four different cohorts in each survey year, with subjects under the new policy being slightly younger. Hence we can only include controls for either age or survey year in our earnings models, and not controls for both as is commonly done in the empirical literature with large cohort studies. Since the subjects in our final samples are mature employees above the age of 50 and holding mainly blue-collar jobs, we expect a flat earnings profile and therefore include controls for survey year.

5.2. Models and Prior Specification

For our formal analysis, we consider two different specifications of the covariate vector, one with a control for survey year (specification 1: S1) and one with an additional control for marital status (specification 2: S2). In particular:

$$\boldsymbol{w}_i = \begin{pmatrix} 1 & \text{year} \end{pmatrix} \text{ [Specification 1] }, \quad \boldsymbol{w}_i = \begin{pmatrix} 1 & \text{year} & \text{married} \end{pmatrix} \text{ [Specification 2]}$$

We exclude age controls for reasons mentioned in Section 4. In our work, the degrees of freedom in the Student-t model is not fixed at the outset and models with different values of the degrees of freedom parameter are estimated. These models are then compared, through (log) marginal likelihoods, to see which assumption about the distribution tails provides the best fit for the data at hand.

Our analysis is conducted with the prior distribution

$$\pi(\boldsymbol{\theta}) = \text{Dir}(\boldsymbol{q}|a_{\text{C},0}, a_{\text{AD14},0}, a_{\text{AD15},0}) \prod_{j=0}^{1} \prod_{k \in K_j} \text{N}_p(\boldsymbol{\beta}_{jk}|\boldsymbol{\beta}_{jk,0}, \boldsymbol{A}_{jk,0}) \mathcal{IG}\left(\boldsymbol{\sigma}_{jk}^2|\nu_{0.jk}, \delta_{0,jk}\right)$$

where the hyperparameters in each model specification are carefully formulated on the basis of an auxiliary sample and the sampling of the prior approach, along the lines of Chib and Ergashev (2009). In

DOI: 10.1002/jae

the latter approach, parameters are simulated from the prior and the outcomes are simulated given the parameters. This process is repeated many times and the resulting prior distributions of the simulated outcomes, the treatment and the treatment effect are inspected. If these distributions seem reasonable, the prior is accepted; otherwise it is revised and the process repeated until a reasonable prior on this basis is found. The approach of sampling the prior shows that uninformed priors do not lead to reasonable distributions of the quantities of interest.

In more detail, we formalize our prior as follows. In the case of our 3-year window, we construct an auxiliary sample of 998 males who turned 14 either between April 1944 and September 1945 or between October 1948 and March 1950 (which is a 1.5-year band just outside the main sample window). For the extended 4-year sample, the auxiliary sample consists of 996 males who turned 14 either between October 1943 and March 1945, or between April 1949 and September 1950. We then estimate our models on these auxiliary data with largely uninformed priors, setting the prior means of the normal prior of the regression coefficients at zero and the standard deviations at 5; the prior means in the inverse gamma distributions for the variance parameters at 0.5 and the standard deviations at 1; the parameters of the Dirichlet prior for the type probabilities at 50, 30 and 20.

Next, we simulate the posterior distribution of the model parameters using the algorithm described in the previous section. The results from this simulation are used to construct the prior for the analysis of the main sample. Specifically, for the prior of the regression coefficients, we set the prior means at the posterior means of the auxiliary sample analysis, and the prior standard deviations at $k = 1$ or $k = 2$, times the posterior standard deviations. Further, we adjust the prior for the intercepts β_{0C0} and β_{1C0} in $g_{jC}(z)$ to imply a neutral prior. Finally, we also recognize that the posterior distribution of compliance types found from the auxiliary sample analysis is potentially skewed given that the auxiliary samples involve individuals further away from the policy change. Therefore, we set the prior parameters in the Dirichlet distribution to imply a more balanced distribution of types by letting $(\alpha_{\mathrm{C},0} = 50, \alpha_{\mathrm{AD14},0} = 25, \alpha_{\mathrm{AD15},0} = 25)$.

The resulting prior parameters are given in Table V for the model of log weekly earnings when $\nu = 5$ is assumed. That this prior is reasonable can be seen from the simulated outcome distributions that are summarized in Table VI. These distributions are based on 10,000 draws of the parameters from the prior distributions in Table V. The results confirm that the prior settings are reasonable. Our prior for the other specifications and the other values of ν are determined in the same way. In each case we find that our prior settings based on the auxiliary samples suggest reasonable distributions of the implied outcomes, treatment intake and treatment effect, and are suitable for our empirical investigation.

5.3. Results

As mentioned above, we consider two different samples: one based on a 3-year window and another based on a wider 4-year sample around the introduction of the policy change. Estimated log marginal likelihoods for our various models and sample datasets are given in Table VII. It can be seen that the log marginal likelihoods provide strong support for the Student-t model with 5 degrees across samples and specifications. Further, in all cases, including a control for marital status increases the marginal likelihood and is the preferred specification.

In Table VIII we summarize the posterior parameter estimates in the preferred earnings model for the 3-year sample (columns 1–4) and the 4-year sample (columns 5–8). The first block of results refers to the regression coefficients in the intake and type-specific earnings models. Overall, the results indicate a considerable variation in the parameter estimates across the types with similar patterns across the two samples. In both samples we observe slightly higher intercept estimates for compliers who leave school at age 15, which are above the intercepts for AD14 but below those for AD15 subjects. The second coefficient measures the effect of the forcing variable in the two complier models,

DOI: 10.1002/jae

Table V. Prior distribution used in the estimation of the models with 5 degrees of freedom

Prior distribution parameter values (estimation 4-year window)								
	Specification 1				Specification 2			
	C0	C1	AD14	AD15	C0	C1	AD14	AD15
3-year window: log weekly earnings								
$\boldsymbol{\beta}_{kj,0}/\boldsymbol{B}_{kj,0}^{0.5}$	5.815/1.387	6.119/1.368	1.462/2.190	7.948/2.661	5.644/1.368	5.950/1.361	1.495/2.205	7.968/2.725
	−0.005/0.011	−0.002 0.007			−0.006/0.011	−0.002/0.007		
	−0.004/0.015		0.046/0.025	−0.027/0.030	−0.004/0.015		0.041/0.025	−0.029/0.031
					0.143/0.095		00.478/0.193	0.263/0.188
$\nu_{kj,0}/\delta_{kj,0}$	49.979/5.123		49.979/12.133	49.979/11.753	48.905/4.878		48.905/11.501	48.905/11.551
4-year window: log weekly earnings								
$\boldsymbol{\beta}_{kj,0}/\boldsymbol{B}_{kj,0}^{0.5}$	5.845/1.555	6.058/1.573	2.715/2.202 0	7.995/2.902	5.790/1.519	5.932/1.538	3.271/2.197	7.958/2.860
	−0.006/0.011	0.002/0.008			−0.005/0.011	0.003/0.008		
	−0.005/0.017		0.032/0.025	−0.026/0.033	−0.005/0.017		0.023/0.025	−0.028/0.032
					0.115/0.094		0.319/0.176	0.297/0.176
$\nu_{kj,0}/\delta_{kj,0}$	49.979/5.123		49.979/12.133	49.979/11.753	48.994/4.987		48.994/10.319	48.994/12.199

DOI: 10.1002/jae

Table VI. Mean, standard deviation and quantiles of simulated outcome y as well as mean of the implied treatment intake and treatment effect

Results from prior simulation									
			Quantiles						
		Mean (SD)	5%	25%	50%	75%	95%	$\Pr(x=1)$	RDD CATE
3-year	Specification 1	5.59 (2.27)	2.35	3.84	5.60	7.32	8.83	0.51	−0.002
	Specification 2	5.57 (2.28)	2.29	3.82	5.57	7.32	8.83	0.52	0.051
4-year	Specification 1	5.58 (2.40)	2.19	3.71	5.57	7.42	9.01	0.53	0.000
	Specification 2	5.56 (2.36)	2.21	3.74	5.55	7.37	8.96	0.53	0.009

Table VII. Estimated log marginal likelihoods for various degrees of freedom and the two covariate specifications

		Model comparison					
		Specification 1			Specification 2		
	Sample	$\nu = 5$	$\nu = 10$	$\nu = \infty$	$\nu = 5$	$\nu = 10$	$\nu = \infty$
Earnings	3-year	−1151.42	−1161.99	−1173.31	**−1139.56**	−1150.16	−1161.96
	4−year	−1468.62	−1476.38	−1486.10	**−1452.78**	−1460.49	−1470.09
Wage	3-year	−1159.76	−1162.87	−1171.31	**−1151.10**	−1153.84	−1162.43
	4−year	−1472.31	−1475.10	−1483.28	**−1459.79**	−1462.76	−1471.32

Note: bold numbers indicate model preferred based on marginal likelihood criterion). Results are based on the draws from the MCMC sampler with 10,000 iterations and 1000 burn-in iterations.

Table VIII. Posterior means and standard deviations for model parameters for the preferred earnings model

Model parameter estimates							
3-year window				4-year window			
$C(x=0)$	$C(x=1)$	AD14	AD15	$C(x=0)$	$C(x=1)$	AD14	AD15
$\boldsymbol{\beta}_{jk}$							
5.688(0.56)	5.734(0.56)	4.542(1.04)	8.969(1.39)	5.826(0.59)	5.891(0.59)	4.703(1.09)	8.491(1.39)
−0.001(0.003)	0.004(0.003)			−0.004(0.005)	0.001(0.004)		
−0.003(0.006)		0.011(0.01)	−0.039(0.02)	−0.004(0.007)		0.009(0.01)	−0.033(0.02)
0.158 (0.043)		0.158(0.09)	0.179(0.09)	0.168(0.049)		0.132(0.09)	0.174(0.10)
σ^2_{jk}							
0.109(0.01)		0.161(0.03)	0.203(0.03)	0.111(0.01)		0.148(0.03)	0.188(0.03)
p_k							
0.611(0.02)		0.181(0.01)	0.208(0.02)	0.571(0.02)		0.203(0.016)	0.226(0.017)

which is positive for compliers leaving at age 15 but smaller and negative for compliers leaving at age 14. The remaining coefficients measure the effect of the survey year and marital status. While we observe a marriage premium for all types, the effect of the survey year varies, being negative for AD15 subjects and for compliers, but positive for AD14 subjects. We also observe some differences in the variance estimates across types in both samples. Finally, in the 3-year sample we have a higher proportion of compliers (61%) compared with the 4-year sample (57%), and lower proportions of AD14 and AD15 subjects.

DOI: 10.1002/jae

Next, we consider the potential earnings distributions by type implied by the model. By definition, the two predictive distributions for compliers are given by

$$p_j(y_{n+1}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w}_{n+1}, s_{n+1} = \mathrm{C}) \\ = \int t_\nu \left(\boldsymbol{w}'_{n+1}\boldsymbol{\beta}_{j,\mathrm{C}}, \sigma^2_{\mathrm{C}}\right) \pi\left(\boldsymbol{\beta}_{j,\mathrm{C}}, \sigma^2_{\mathrm{C}}|\boldsymbol{y}, \boldsymbol{x}\right) p(\boldsymbol{w}_{n+1}|\boldsymbol{y}, \boldsymbol{x})\, d\boldsymbol{\beta}_{j,\mathrm{C}}\, d\sigma^2_{\mathrm{C}}\, d\boldsymbol{w}_{n+1}$$

where the first term in the integrand is the sampling density, the second is the posterior distribution of the model parameters and the third (following Chib and Jacobi (2008)) is the empirical distribution of the sample covariates. The predictive distributions for compliers can be obtained in a straightforward manner by the method of composition. Following the same approach we obtain the predictive distributions for AD14 and AD15 subjects. In Figure 2 we provide graphs of the four kernel-smoothed densities of these marginal predictive distributions by treatment intake and type for both samples. The upper panel shows the graphs of the potential earnings densities for compliers, in each treatment state, as well as the densities for the AD14 and AD15 subjects.

The graphs reveal that in both samples the earning densities of compliers in the $x = 1$ state are shifted to the right of the density for compliers in the $x = 0$ state, which points to a positive earnings effect for compliers pointing towards higher earnings of compliers under $x = 1$. We can further see that the corresponding densities for the AD14 and AD15 subjects display larger variances and thicker tails, with the density of AD15 subjects somewhat skewed to the right. The difference is more pronounced in the 3-year sample. While the lower panel of Figure 2 with the predictive densities for wages also exhibits densities for the AD15 subjects that are skewed to the right, these graphs show

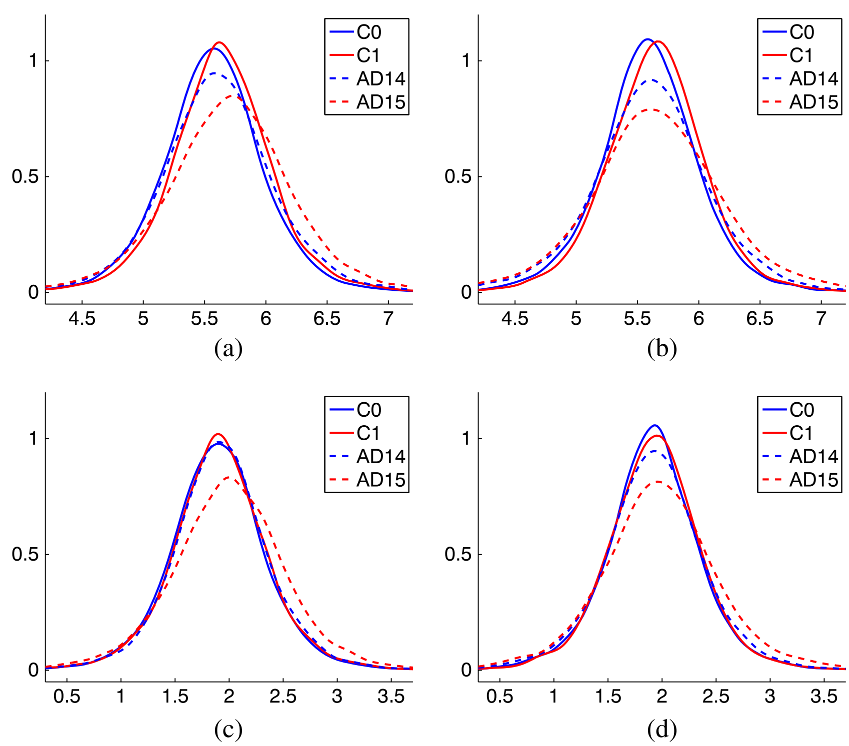


Figure 2. Kernel-smoothed predictive densities of potential earnings and wage by type and treatment for the preferred model: (a) earnings (3-year window); (b) earnings (4-year window); (c) wages (3-year window); (d) wages (4-year window)

almost overlapping densities for the compliers, suggesting almost no wage effects for compliers for staying until age 15. Also, the densities for AD14 is essentially overlapping with those of compliers. As in the case for earnings, we again observe a predictive distribution for AD15 subjects that is skewed to the right. This may suggest that individuals who left school at age 15 irrespective of the policy in place may have left school only after obtaining a school-leaving certificate. Thus these students would be different in terms of their skill sets compared to those students who stayed on only to fulfill the minimum schooling requirement. Consequently, such students would perform better in the labor market. Recent work by Grenet (2009) provides empirical evidence based on schooling reforms in France and England that a positive earnings effect of compulsory schooling emerges when qualifications of students increase.

To evaluate the causal effect of an extra year of compulsory schooling at the age of 14, we first present the estimated average earnings and wage effects for compliers. Table IX presents the posterior estimates of the complier average earnings and wage effects for all model specifications in both samples summarized in terms of the mean of the posterior distribution (standard deviations in parentheses). The results of the preferred model are marked with an asterisk. The first block of results refers to the weekly earnings variable. In our main 3-year sample the estimated RDD CATE for earnings ranges from 4.3% to 6.6%, with 6.4% in the preferred model. While the RDD CATE from the preferred model in the 4-year sample is only slightly lower, at 4.5%, overall the estimates vary between 1% and 5.8% across the two model specifications and different degrees of freedom. For the larger sample size, the distribution of RDD CATE is less dispersed. As in previous work, our estimated effects are lower for hourly wages, ranging between 1.6% and 3.7% in the 3-year sample, and between 0.7% and 3.4% in the 4-year sample. The sizes of these effects from the preferred models are 1.6% and 0.07% in the two samples, respectively. As in the case of the log earnings outcome, the estimates from the 4-year sample are generally lower and more varied across specifications.

It should be noted that because of the difference in approach and the difference in the sample data, our estimates of the effect of schooling are not directly comparable to results from previous work that exploited the same policy change. However, our results are qualitatively similar to those from two large cohort studies (Oreopoulos, 2008; Devereux and Hart, 2010), where the average effect of an additional year of schooling on earnings was small but positive, with a lower effect on wages.

We conclude the analysis by exploiting a further feature of the Bayesian approach and examine the complete posterior distribution of the complier average earnings and wage effects. Figure 3 shows the kernel-smoothed graphs of the posterior densities of the complier average returns to schooling

Table IX. Posterior means and standard deviations of the RDD CATE based on the fittings of the various model specifications

Estimates of average returns to schooling for compliers					
	Sample	Model	$\nu = 5$	$\nu = 10$	$\nu = \infty$
Earnings	3-year	S1	0.066 (0.074)	0.048 (0.068)	0.045 (0.063)
		S2	0.064 (0.074)*	0.049 (0.068)	0.043 (0.062)
	4-year	S1	0.058 (0.065)	0.036 (0.060)	0.021 (0.054)
		S2	0.045 (0.063)*	0.027 (0.059)	0.008 (0.055)
Wage	3-year	S1	0.018 (0.080)	0.037 (0.078)	0.030 (0.071)
		S2	0.016 (0.079)*	0.035 (0.078)	0.026 (0.072)
	4-year	S1	0.021 (0.070)	0.034 (0.068)	0.017 (0.060)
		S2	0.007 (0.069)*	0.022 (0.067)	0.005 (0.061)

* Indicates estimates from preferred model.

DOI: 10.1002/jae

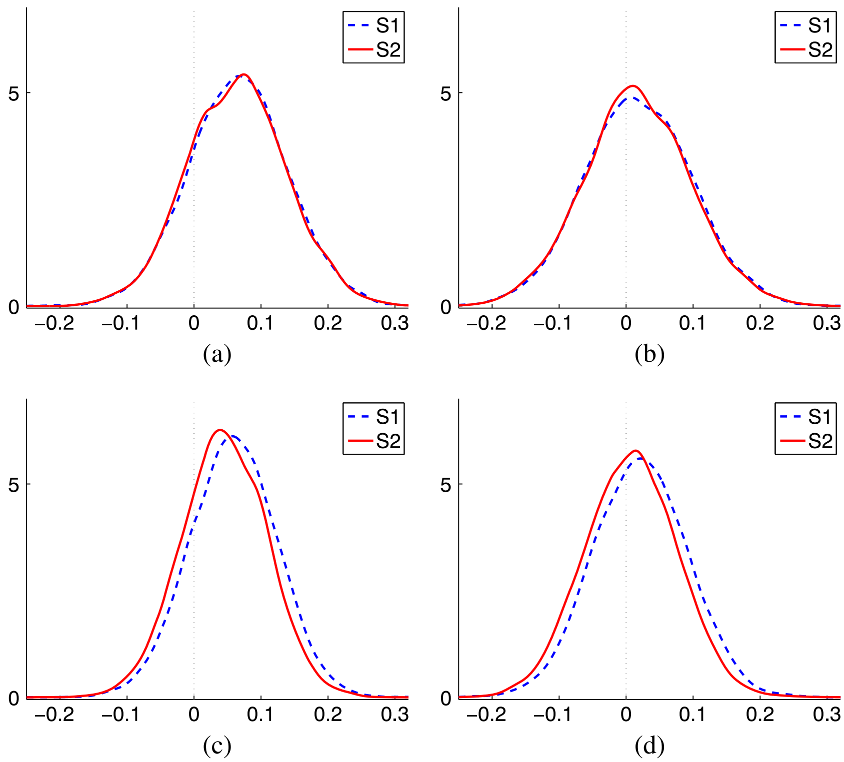


Figure 3. Kernel-smoothed densities of the posterior density of RDD CATE for the preferred specifications 1 and 2 (under 5 degrees of freedom) for earnings and wages: (a) earnings (3-year window); (b) wages (3-year window); (c) earnings (4-year window); (d) wages (4-year window)

for the two-covariate specification (specifications 1 and 2) and 5 degrees of freedom (the model with the highest marginal likelihood). The upper panel contains the graphs for the 3-year window and the lower panel the 4-year window. The RDD CATE densities for earnings are centered to the right of zero, with more mass over positive values. The densities for earnings are located further to the left. The RDD CATE posterior densities are overall symmetric, though that under the 3-year sample is centered slightly further to the right and displays a slightly larger spread. While the densities under both specifications are almost identical under the 3-year sample, those of the preferred specification 2 are to the left of specification 1 under the 4-year sample.

6. CONCLUDING REMARKS

In this paper we evaluate the returns to education from the increase in the compulsory schooling age from 14 to 15 in the UK in 1947. Motivated by the specific features of the enactment, and the discontinuous impact the new policy had on the probability of leaving school at age 14 for cohorts who turned 14 around the time of the change, we develop a Bayesian fuzzy RD framework based on compliance types to estimate the complier average earnings effect from an additional year of schooling at age 14. Interestingly, our findings on the treatment effect parallel those in a number of recent studies (Lindeboom *et al.*, 2009; Clark and Royer, 2007; Galindo-Rueda and Vignoles, 2007) where the impact of this policy change on socio-economic outcomes that are considered to be correlated with earnings, such as children's health outcomes and mortality, was found to be small.

DOI: 10.1002/jae

ACKNOWLEDGEMENTS

The authors are grateful to the editor and two referees for their diligence and constructive comments.

REFERENCES

Angrist JD, Lavy V. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2): 533–575.

Banks J, Blundell R. 2005. Private pension arrangements and retirement in Britain. *Fiscal Studies* **26**(1): 35–53.

Calonico S, Cattaneo MD, Titiunik R. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82**(6): 2295–2326.

Campbell DT. 1969. Reforms as experiments. *American Psychologist* **24**(4): 409–429.

Cattaneo MD, Frandsen B, Titiunik R. 2015. Randomization inference in the regression discontinuity design: an application to party advantages in the U.S. senate, *Journal of Causal Inference*. (in press).

Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313–1321.

Chib S, Ergashev B. 2009. Analysis of multi-factor affine yield curve models. *Journal of the American Statistical Association* **104**: 1324–1337.

Chib S, Greenberg E. 2014. Nonparametric Bayes analysis of sharp and fuzzy regression discontinuity designs. Technical report, Washington University in St Louis.

Chib S, Jacobi L. 2008. Analysis of treatment response data from eligibility designs. *Journal of Econometrics* **144**: 465–478.

Clark D, Royer H. 2007. The effect of education on adult mortality and health: evidence from the United Kingdom. Working paper, Centre for Economic Performance.

Devereux PJ, Hart RA. 2010. Forced to be rich? Returns to compulsory schooling in Britain. *Economic Journal* **120**: 1345–1364.

Frandsen BR, Froelich M, Melly B. 2012. Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics* **168**: 382–395.

Frangakis CE, Rubin DB. 2002. Principal stratification in causal inference. *Biometrics* **58**(1): 21–29.

Galindo-Rueda F, Vignoles A. 2007. The heterogeneous impact of selection in secondary schools. In *Schools and the Equal Opportunity Problem*, Woessman L, Peterson P (eds). MIT Press: Cambridge, MA.

Grenet J. 2009. Is it enough to increase compulsory education to raise earnings? Evidence from French and British compulsory schooling laws. CEP working paper.

Hahn J, Todd P, der Klaauw WV. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* **69**(1): 201–209.

Halsey AH, Heath AF, Ridge JM. 1980. *Origins and Destinations: Family, Class and Destinations in Modern Britain*. Oxford University Press: New York.

Harmon C, Walker I. 1995. Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* 1278–1286.

Imbens GW, Lemieux T. 2008. Regression discontinuity designs: a guide to practice. *Journal of Econometrics* **142**(2): 615–635.

Imbens GW, Rubin DB. 1997. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**: 305–327.

Lee DS, Lemieux T. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* **48**: 281–355.

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

Oreopoulos P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* **96**(1): 152–175.

Oreopoulos P. 2008. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *Corrigendum* **96**(152): 139.

Thistlethwaite DL, Campbell DT. 1960. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. *Journal of Educational Psychology* **51**(6): 309–317.