The Annals of Applied Statistics
2015, Vol. 9, No. 4, 1906–1931
DOI: 10.1214/15-AOAS881
© Institute of Mathematical Statistics, 2015

EVALUATING THE CAUSAL EFFECT OF UNIVERSITY GRANTS ON STUDENT DROPOUT: EVIDENCE FROM A REGRESSION DISCONTINUITY DESIGN USING PRINCIPAL STRATIFICATION

BY FAN LI[*,1], ALESSANDRA MATTEI[†,2] AND FABRIZIA MEALLI[†]

Duke University[*] and University of Florence[†]

Regression discontinuity (RD) designs are often interpreted as locally randomized experiments for units with a realized value of a pretreatment variable falling around a threshold. Motivated by the evaluation of Italian university grants, we consider a fuzzy RD design where the treatment status is based on both eligibility criteria and a voluntary application status. Resting on the fact that grant application and grant receipt statuses are post-assignment (post-eligibility) intermediate variables, we use the principal stratification framework to define causal estimands within the Rubin Causal Model. We propose a probabilistic formulation of the assignment mechanism underlying RD designs, by reformulating the Stable Unit Treatment Value Assumption (SUTVA) and making an explicit local overlap assumption for a subpopulation around the threshold. We invoke a local randomization assumption instead of the more standard continuity assumptions. We also develop a Bayesian approach to select the target subpopulation(s) with adjustment for multiple comparisons, and to draw inference for the target causal estimands within this framework. Applying the method to the data from two Italian universities, we find evidence that university grants are effective in preventing students from low-income families from dropping out of higher education.

1. Introduction. Amid the recent economic crisis in Europe, there has been a heated debate on how to arrange college students financial support, especially in terms of the instruments used, for example, loans, grants, tuition waiver. Accurate evaluation of the effectiveness of the existing financial aid systems is crucial for providing information to policy makers to choose between different instruments. In Italy state universities offer financial aid every year to a limited number of eligible freshmen. This intervention aims to give equal opportunity to achieve higher education to motivated students irrespective of their economic background. Dropout from university is a relevant phenomenon in Italy: indeed, the low rate of university graduates among Italian youths is mainly due to the high dropout rate (about 30%) rather than to a low enrollment rate. In this paper, we will investigate

Received December 2014; revised September 2015.

[1]Supported by NSF Grants SES-1155697 and SES-1424688.

[2]Part of this paper was written when Alessandra Mattei was a senior fellow of the Uncertainty Quantification program of the U.S. Statistical and Applied Mathematical Sciences Institute (SAMSI).

Key words and phrases. Bayesian, causal effects, intermediate variables, principal stratification, randomization, regression discontinuity, university grants.

the causal effects of Italian university grants on student dropout rate, using data on first-year enrollees from two universities.

In the Italian university system, only students who both meet a prefixed eligibility criteria and apply for a grant can receive the grant, consisting of tuition waiver, free meals and accommodation, and a limited stipend of around 3000 Euros. Eligibility depends on an economic measurement of the student's family income and assets falling below or above a predetermined threshold. This allocation rule motivates us to adopt the regression discontinuity (RD) design framework for evaluation. RD design—a quasi-experimental design for causal inference—was first introduced in psychology by Thistlethwaite and Campbell (1960) and has became increasingly popular since the late 1990s in economics and other fields. Recent surveys can be found in Cook (2008), Imbens and Lemieux (2008), Lee and Lemieux (2010), van der Klaauw (2008). There are two general setups in RD designs, the sharp and the fuzzy RD designs. In the sharp design, the original form of the design, the treatment status is assumed to be a deterministic step function of a so-called *forcing variable* or *running variable*. All units with a realized value of the forcing variable on one side of a prefixed threshold are assigned to one regime and all units on the other side are assigned to the other regime. The basic idea underlying a RD analysis is that one can compare units with very similar values for the forcing variable, but different levels of treatment, to draw causal inference of the treatment at the threshold. Examples of sharp RD designs can be found, among others, in Berk and de Leuuw (1999), Lee (2008), Mealli and Rampichini (2012). In the fuzzy design, the realized value of the forcing variable does not alone determine the receipt of the treatment, although a value of the forcing variable falling above or below the threshold acts as an encouragement or incentive to participate in the treatment. In those cases, the receipt of the treatment depends also on individual choices, which may confound treatment receipt. Hahn, Todd and Van der Klaauw (2001) establish a connection between fuzzy RD designs and instrumental variables (IV), and show that in a fuzzy RD setting one can identify the local average treatment effect [Imbens and Angrist (1994)] for a subpopulation of compliers at the threshold. Examples of fuzzy RD designs can be found, among others, in Battistin and Rettore (2008), Garibaldi et al. (2012), van der Klaauw (2002).

The Italian university grant allocation rule defines a fuzzy RD design because not all eligible students get a grant; they must apply first, and application is voluntary. Also, ineligible students may apply, even if they will not receive any grant. Comparing to standard fuzzy RD designs where only assignment (eligibility) and receipt of the treatment (grant) are available, the additional data on the application status in this study may provide valuable information for policy makers. In this article we develop a framework for RD analysis that is embedded in the Rubin Causal Model [RCM, Rubin (1974, 1978)]. Resting on the fact that grant application and grant receipt statuses are post-assignment (post-eligibility) variables, we adopt the principal stratification framework [Frangakis and Rubin (2002)]—a generalization of the IV approach to noncompliance [Angrist, Imbens and Rubin

(1996), Imbens and Rubin (1997)]—to define causal estimands and lay the basis for inference.

Causal inference in RD designs is usually based on comparisons of units with close but distinct values of the forcing variable and relies on smoothness assumptions about the relationship between outcomes and the forcing variable around the threshold, which imply randomization at the single threshold value. For example, in fuzzy RDs, estimands are usually specified as ratio of differences of regression functions at the threshold, and inference generally relies on asymptotic approximations [e.g., Imbens and Lemieux (2008)]. In real applications, large-sample approximations might be unreliable due to the small sample size, and exact inference would be preferable. RD designs have been often described as leading to locally randomized experiments around the threshold [Dinardo and Lee (2011), Lee (2008), Lee and Lemieux (2010)]. Expanding on this interpretation, a recent strand of the literature [e.g., Cattaneo, Frandsen and Titiunik (2015), Sales and Hansen (2014)] is moving toward a formal and well-structured definition of the conditions under which RD designs are equivalent to local randomized experiments.

We further develop the idea of local randomization; we aim to provide a formal definition of the hypothetical experiment underlying RD designs, formalizing the assignment mechanism, that is, the process that describes which units got assigned to which treatment. The core of our framework is to assume there exists at least one subpopulation around the threshold where a *local overlap* assumption holds. For this subpopulation we explicitly introduce a *local randomization* assumption. Though our framework is not tied to any mode of inference, we choose the Bayesian approach for reasons explained later. We also develop a Bayesian hierarchical modeling approach to adjust for multiple comparisons in selecting the target subpopulation(s). Our work adds to the limited literature on Bayesian analysis of RD [Chib and Greenberg (2014), Chib and Jacobi (2015)].

In Section 2 we introduce the basic setup and the causal estimands. In Section 3 we propose a probabilistic formulation of the assignment mechanism for general RD designs, explicitly formulating the key assumptions, and elaborate it for the particular RD design used in the Italian university grants. Selection of the subpopulations where these assumptions hold is also discussed. A Bayesian approach for inferring causal effects in RD designs is developed in Section 4. We then apply the proposed approach to evaluate causal effects of Italian university grants on student dropout in Section 5. Section 6 concludes.

2. Causal estimands.

2.1. *Basic setup*. We introduce the notation in the context of Italian university grants. Let Z be the eligibility status, which is the initial assignment and plays the role of an "instrument" or an "encouragement," as in randomized experiments with noncompliance. Consider a sample or population of N units; each can be either eligible to receive a treatment, $z = 1$, or ineligible, $z = 0$. In the Italian grants

system, eligibility depends on the value of a combined measurement of one's assets including income and properties, adjusted for family size, denoted by S. If a student, satisfying preliminary grade criteria, has a value of S falling below a predetermined threshold, for example, $s_0 = 15{,}000$ euros, he/she is eligible, and he/she is not eligible otherwise. That is, the eligibility status Z_i for student i is a deterministic function of S: $Z_i = \mathbf{1}(S_i \leq s_0)$, where $\mathbf{1}(\cdot)$ is the indicator function. Using the terminology in RD designs, S is the forcing variable.

Three variables were measured after each student i is assigned eligibility Z_i: the application status, the receipt of the grant and the dropout status; and, in principle, eligibility may affect them. Thus, we can define potential outcomes for these variables: for each student i ($i = 1, \ldots, N$), given eligibility status z ($z = 0, 1$), let $A_i(z)$ be an indicator for the potential grant application status (equal to 1 if student i applies for a grant and 0 otherwise), $W_i(z)$ be an indicator for the potential treatment received (equal to 1 if student i receives a grant and 0 otherwise), and $Y_i(z)$ be the potential indicator for dropout (1 if student i drops out of university, 0 otherwise). This notation, with only two potential outcomes for each post-treatment variable for each unit, reflects the acceptance of the Stable Unit Treatment Value Assumption [SUTVA, Rubin (1980)], which implies that there is no interference between units and that there are no levels of the eligibility status other than zero and one. A more explicit formulation of SUTVA will be introduced in Section 3.1.

For each student i, given the observed eligibility status Z_i, the following variables are observed: $A_i^{\text{obs}} = A_i(Z_i)$, the observed application status; $W_i^{\text{obs}} = W_i(Z_i)$, the observed treatment received; and $Y_i^{\text{obs}} = Y_i(Z_i)$, the observed dropout status. The remaining potential outcomes are missing: $A_i^{\text{mis}} = A_i(1 - Z_i)$, $W_i^{\text{mis}} = W_i(1 - Z_i)$, and $Y_i^{\text{mis}} = Y_i(1 - Z_i)$. A vector of p pretreatment variables, $\mathbf{X}_i$, is also observed for each unit. We use boldface upper-case letters to denote the vector of values of a variable for all units from hereon. For example, $\mathbf{Z} = (Z_1, \ldots, Z_N)'$, $\mathbf{A}^{\text{obs}} = (A_1^{\text{obs}}, \ldots, A_N^{\text{obs}})'$.

2.2. *The role of principal stratification for causal inference in fuzzy RD designs.* In the RCM, a causal effect is defined as a comparison of the potential outcomes $Y_i(1)$ and $Y_i(0)$ for *a common set of units*. Obviously, in our study, such comparisons between potential dropout statuses only measure the effect of the eligibility status. To draw inference about the causal effect of *receiving a grant*, additional structure and assumptions are required. Since both the application status and receipt of the grant are post-assignment variables, we adopt the Principal Stratification framework [Frangakis and Rubin (2002)].

For each post-assignment variable, principal stratification defines a cross-classification of subjects into groups, namely, principal strata, defined by the joint potential values of that post-assignment variable under each of the assignments being compared. In our study, based on the application status A, subjects

are classified into four principal strata, $G_i \equiv (A_i(0), A_i(1))$: compliant-applicants $G_i = (0,1) = \text{CA}$, students who would not apply if ineligible, but would apply if eligible; always-applicants $G_i = (1,1) = \text{AA}$, students who would apply irrespective of their eligibility status; never-applicants $G_i = (0,0) = \text{NA}$, students who would not apply irrespective of their eligibility status; and defiant-applicants $G_i = (1,0) = \text{DA}$, students who would not apply if eligible, but would apply if ineligible. Because principal strata are not affected by assignment, we can define *population-average* causal effects conditional on the principal strata, known as principal causal effects:

$$\tau_g \equiv \mathbb{E}\big[Y_i(1) - Y_i(0) | G_i = g\big], \tag{1}$$

for $g = \text{AA}, \text{CA}, \text{NA}, \text{DA}$. Then the average causal effect of eligibility on dropout is a weighted average of these principal causal effects:

$$\mathbb{E}\big[Y_i(1) - Y_i(0)\big] = \sum_{g=\text{AA},\text{CA},\text{NA},\text{DA}} \pi_g \tau_g,$$

where π_g is the proportion of units in principal stratum g.

Never-applicants and defiant-applicants never receive a grant, so for them we always observe the outcome in the absence of the grant. By contrast, for always-applicants and compliant-applicants we can observe $Y_i(1)$ for some eligible students who receive a grant and $Y_i(0)$ for some other ineligible students who do not receive a grant. Therefore, always-applicants and compliant-applicants are the only groups where we can learn information about the effect of receiving a grant in this study, and thus the corresponding principal causal effects, τ_{AA} and τ_{CA}, are the causal estimands of primary interest.

In the standard IV approach to noncompliance [Angrist, Imbens and Rubin (1996), Imbens and Rubin (1997)] as well as the standard setting of fuzzy RD designs [e.g., Imbens and Lemieux (2008)], data on application status are not utilized, either because they are not available or because they are ignored. Instead, the analysis is based on the principal strata formed by the post-assignment variable of grant receipt status. Specifically, there are four principal strata based on the joint potential grant receipt statuses, $R_i = (W_i(0), W_i(1))$: compliers $R_i = (0,1)$, students who would receive the grant if eligible and would not receive the grant if ineligible; always-takers $R_i = (1,1)$, students who would receive the grant regardless of eligibility; never-takers $R_i = (0,0)$, student would not receive the grant regardless of eligibility; and defiers $R_i = (1,0)$, students who would not receive the grant if eligible and would receive the grant if ineligible. The focus is generally on the causal effect for compliers:

$$\tau \equiv \mathbb{E}\big[Y_i(1) - Y_i(0) | R_i = (0,1)\big].$$

We now establish the connection between these two sets of principal strata. The Italian grant assignment rule implies that $W_i(0) = 0$ for all i, as ineligible

units have no access to a grant, and $W_i(1) = 0$ if $A_i(1) = 0$, as eligible units need to apply for a grant to receive a grant. Therefore, by design, there are no always-takers or defiers, and the remaining principal strata R's can be expressed as unions of principal strata G's: never-takers comprise never-applicants and defiant-applicants, and compliers comprise always-applicants and compliant-applicants. As such, τ can be rewritten as the weighted average of the causal effects for always-applicants and compliant-applicants:

$$\tau = \mathbb{E}\big[Y_i(1) - Y_i(0) | G_i \in \{\mathrm{AA}, \mathrm{CA}\}\big] = \frac{\pi_{\mathrm{AA}}\tau_{\mathrm{AA}} + \pi_{\mathrm{CA}}\tau_{\mathrm{CA}}}{\pi_{\mathrm{AA}} + \pi_{\mathrm{CA}}}. \tag{2}$$

This illustrates that principal strata defined by the application status leads to a finer partition of the units than principal strata defined by the grant-receipt status. Indeed, the standard IV causal estimand—the causal effect for compliers τ—provides information on a "marginal" (with respect to application behavior) causal effect. If causal effects are homogeneous, marginalizing over application behavior does not critically affect the evaluation analysis. Conversely, if causal effects are heterogeneous, ignoring application behavior may represent a loss of potentially useful information. For example, if the grants are found to have a higher positive effect on always-applicants than on compliant-applicants, then it would be useful and cost-effective to study the characteristics of ineligible applicants and include those into the eligibility rule to allocate additional resources.

The estimands τ_{AA}, τ_{CA} and τ represent effects of eligibility, rather than effects of the receipt of a grant. However, "the receipt of a grant" is completely confounded with "the eligibility status": $W(z) = z \times A(z) = z$ for always-applicants and compliant-applicants. To *attribute* these effects to "the receipt of a grant," below we can make an exclusion restriction assumption:

ASSUMPTION 1 (Exclusion restriction for compliant-applicants and always-applicants). For all units with $G_i \in \{\mathrm{AA}, \mathrm{CA}\}$ or, equivalently, $R_i = (0, 1)$, the effect of eligibility is only through the receipt of the grant.[3]

[3]This assumption could be formalized by introducing potential outcomes of the form $Y_i(z, a, w)$, that is, potential outcomes for Y if eligibility status Z was set to z, application status A was set to a, and grant status W was set to w; $z, a, w \in \{0, 1\}$. Specifically, in our setting, the (stochastic) exclusion restriction assumption for compliant-applicants and always-applicants would require that for each $a', a'' \in \{0, 1\}$, $\Pr(Y_i(0, a', w) | G_i \in \{\mathrm{AA}, \mathrm{CA}\}) = \Pr(Y_i(1, a'', w) | G_i \in \{\mathrm{AA}, \mathrm{CA}\})$, $w = 0, 1$. This exclusion restriction assumes that for compliant-applicants and always-applicants, that is, compliers, the potential outcome that would realize if they were eligible and received a grant (did not receive) is equal to the potential outcome that would realize if they were ineligible and received (did not receive) a grant, irrespective of their application status. The potential outcomes $Y_i(z, a, w)$ are a priori counterfactuals for units who exhibit a value of the application status, A_i^{obs}, and a value of the grant status, W_i^{obs}, under treatment z not equal to a and w, respectively, because in one specific experiment, they can never be observed for such types of units. Here we prefer to avoid potential outcomes of the form $Y_i(z, a, w)$ by focusing on observable potential outcomes $Y_i(z)$.

Assumption 1 attributes the intention-to-treat (ITT) effect for compliers to the receipt of grant rather than eligibility. A more formal version of this assumption, which requires double-indexed notation, is given in Footnote 3 and also Imbens and Rubin (2015) (Chapter 23, Assumption 23.4). This type of exclusion restriction is routinely made, often implicitly, in randomized experiments with full compliance [Mealli and Pacini (2013), Mealli and Rubin (2002)].

In real studies, the *sample-average* counterpart of the *population-average* estimands may also be of interest:

$$\tau_g^S \equiv \frac{1}{N_g} \sum_{i:G_i=g} \big[Y_i(1) - Y_i(0)\big], \tag{3}$$

where $g = \text{AA}, \text{CA}, \{\text{AA}, \text{CA}\}$ and N_g is the number of units in stratum g. Usually the *sample-average* effects can be estimated more precisely than their *population-average* counterparts. The subtle difference between them in Bayesian inference is explained in Section 4. More details can be found, for example, in Imbens and Rubin (1997), Rubin (1978) and Imbens (2004). For simplicity, we do not distinguish between population-average and sample-average estimands in the methodological discussion, but will present both estimates in the application.

3. The basis for inference.

3.1. *Probabilistic treatment assignment mechanism in RD designs.* The complex selection process in the Italian university grants system implies that the mechanism governing the receipt of the grant, which depends on both institutional and individual choices, is not ignorable. Below we introduce a probabilistic assignment mechanism underlying the RD design considered here, which is also applicable to general RD settings with minor modifications.

We first define the assignment mechanism, which is a row-exchangeable function that assigns probabilities to all 2^N possible N-dimensional vectors of assignments $\mathbf{Z}$, as a row-exchangeable function that assigns probabilities to all possible N-dimensional vectors of realizations of the forcing variable, $\mathbf{S}$, above or below the threshold value, s_0. Formally,

$$\begin{aligned} &\Pr\big(\mathbf{Z} = \mathbf{z} | \mathbf{A}(0), \mathbf{A}(1), \mathbf{W}(0), \mathbf{W}(1), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\big) \\ &\quad = \Pr\big(\mathbf{S} \in \mathbf{\Lambda} | \mathbf{A}(0), \mathbf{A}(1), \mathbf{W}(0), \mathbf{W}(1), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\big), \end{aligned} \tag{4}$$

where $\mathbf{z} \in \{0,1\}^N$ and $\mathbf{\Lambda} \in \{(-\infty, s_0]^N, (-\infty, s_0]^{N-1} \times (s_0, \infty), (s_0, \infty) \times (-\infty, s_0]^{N-1}, \ldots, (-\infty, s_0] \times (s_0, \infty)^{N-1}, (-\infty, s_0]^{N-1} \times (s_0, \infty), (s_0, \infty)^N\}$. Since Z is a deterministic function of S, the assignment mechanism can be formulated with respect to either Z or S. Here we prefer S because it is the underlying random variable that describes the reasons for the missing and observed values of potential outcomes: a value of S is assigned, which in turn determines a value for Z.

Statistical inference for causal effects requires assumptions on the assignment mechanism. We introduce assumptions that allow us to describe RD settings as classical randomized experiments around the threshold. The assignment mechanism in equation (4) is a classical randomized experiment if (i) it is individualistic,

$$\Pr\big(\mathbf{S} \in \mathbf{\Lambda} | \mathbf{A}(0), \mathbf{A}(1), \mathbf{W}(0), \mathbf{W}(1), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{X}\big) = \prod_{i=1}^{n} \Pr\big(S_i \leq s_0 | A_i(0), A_i(1), W_i(0), W_i(1), Y_i(0), Y_i(1), \mathbf{X}_i\big);$$

(ii) it is probabilistic, which implies that for each unit, i, both events $S_i \leq s_0$ and $S_i > s_0$ have a priori a nonzero probability of occurring; (iii) it is unconfounded, that is, free of dependence of any potential outcomes; and (iv) it is a known function of its arguments.

The particular assignment rules underlying RD designs suggest that these assumptions are more reasonable for subpopulations of units who have a relatively large probability that the realized values of the forcing variable fall in a neighborhood around the threshold, s_0. For these subpopulations, we can reasonably assume that the distribution of the forcing variable is unrelated to observed and unobserved characteristics of students. On the other hand, students with a very small (close to zero) or a very large (close to one) probability that $S_i \leq s_0$ are likely systematically different from other students. For example, potential outcomes observed for very rich students, who do not receive any grant, are plausibly different from potential outcomes for poor students with a realized value of S around the threshold, who do not receive a grant, and vice versa. Therefore, we focus on subpopulations of students who have a probability that $S_i \leq s_0$ strictly between zero and one, and sufficiently far away from zero and one. The following assumption guarantees that at least one such subpopulation of units exists.

ASSUMPTION 2 (Local overlap). Let $\mathcal{U}$ be the random sample (or population) of units in the study. There exists a subset of units, $\mathcal{U}_{s_0}$, such that for each $i \in \mathcal{U}_{s_0}$, $\Pr(S_i \leq s_0) > \varepsilon$ and $\Pr(S_i > s_0) > \varepsilon$ for some sufficiently large $\varepsilon > 0$.

Assumption 2 assumes that there exists a subpopulation of units, each of whom has a nonzero probability of being assigned to either treatment level. This represents a main distinction between our framework and the existing RD literature that often describes RD designs as settings where the overlap assumption is violated. Now within the subpopulation $\mathcal{U}_{s_0}$ we can formally introduce a modified SUTVA specific to the RD settings:

ASSUMPTION 3 (Local RD-SUTVA). For each $i \in \mathcal{U}_{s_0}$, consider two eligibility statuses $Z_i' = \mathbf{1}(S_i' \leq s_0)$ and $Z_i'' = \mathbf{1}(S_i'' \leq s_0)$, with possibly $S_i' \neq S_i''$. If $Z_i' = Z_i''$, that is, if either $S_i' \leq s_0$ and $S_i'' \leq s_0$, or $S_i' > s_0$ and $S_i'' > s_0$, then $A_i(\mathbf{Z}') = A_i(\mathbf{Z}'')$, $W_i(\mathbf{Z}') = W_i(\mathbf{Z}'')$, and $Y_i(\mathbf{Z}') = Y_i(\mathbf{Z}'')$.

Local RD-SUTVA rules out interference between units, implying that potential outcomes for a student cannot be affected by the eligibility status of other students. Local RD-SUTVA also assumes that there are no levels of the eligibility status other than zero and one. This component of RD-SUTVA implies that values of the forcing variable leading to the same eligibility status cannot alter potential outcomes for any unit, and thus allows us to avoid defining potential outcomes as functions of the forcing variable. Under the local RD-SUTVA for each unit within $\mathcal{U}_{s_0}$ there exist only two potential outcomes for each post-assignment variable, corresponding to the realized value of the forcing variable falling *below* and *above* the threshold, respectively.

Finally, we formalize the concept of RD design as a local randomized experiment:

ASSUMPTION 4. (Local randomization). For each $i \in \mathcal{U}_{s_0}$,

$$\Pr\big(S_i | A_i(0), A_i(1), W_i(0), W_i(1), Y_i(0), Y_i(1), \mathbf{X}_i\big) = \Pr(S_i).$$

Assumption 4 states that within the subpopulation $\mathcal{U}_{s_0}$ a Bernoulli trial has been conducted, with individual assignment probabilities depending only on the distribution of the forcing variable, $\Pr(Z_i = 1) = \Pr(S_i \leq s_0)$, but not on either the potential outcomes or pre-treatment variables. This assumption is crucial in justifying the key idea underlying any RD design. It implies that the eligibility statuses are randomly assigned in some small neighborhood, $\mathcal{U}_{s_0}$, around s_0.

Assumption 4 may not always be plausible. For instance, when the forcing variable is a deterministic variable, which conceptually cannot be interpreted as a random variable with a nondegenerate probability distribution (such as time), the underlying design cannot, in general, be interpreted as a locally randomized experiment [see Section 6.3 in Lee and Lemieux (2010), page 347].

There are subtle but substantive differences between local RD-SUTVA and local randomization. Local RD-SUTVA is an exclusion restriction assumption and it is required to make the representation of potential outcomes as functions of the eligibility status adequate. Local randomization is an independence assumption that is crucial for inference. RD-SUTVA is different from independence assumptions: it does not imply that the probability that we observe a value of the forcing variable above or below the threshold does not depend on potential outcomes. RD-SUTVA simply implies that the exposure to assignment level z specifies a well-defined potential outcome, for all unit i and assignment levels z. In other words, considering potential outcomes as random variables, RD-SUTVA does not imply that potential outcomes have the same distribution for each value of the forcing variable. In order to make the forcing variable independent of potential outcomes, we need an additional assumption, such as Assumption 4.

Following Assumption 2, we can define a local version of the target estimands,

$$\tau_{g,s_0} \equiv \mathbb{E}\big[Y_i(1) - Y_i(0) | G_i = g, i \in \mathcal{U}_{s_0}\big], \tag{5}$$

for $g = \text{AA}, \text{CA}, \{\text{AA}, \text{CA}\}$ and their *finite-sample* counterparts, and we have

$$\tau_{\{\text{AA,CA}\},s_0} \equiv \tau_{s_0} = \frac{\tau_{\text{AA},s_0}\pi_{\text{AA},s0} + \tau_{\text{CA},s_0}\pi_{\text{CA},s0}}{\pi_{\text{AA},s_0} + \pi_{\text{CA},s_0}},$$

where $\pi_{g,s_0} = \Pr(G_i = g | i \in \mathcal{U}_{s_0})$ for $g = \text{AA}, \text{CA}, \text{NA}, \text{DA}$, denote the proportion of principal strata in $\mathcal{U}_{s_0}$. A special case of $\mathcal{U}_{s_0}$ contains the subpopulation of units with a realized value of the forcing variable *exactly* equal to s_0.

Also, Assumption 4 implies that

$$\mathbb{E}\big[Y_i(1) - Y_i(0) | G_i = g, i \in \mathcal{U}_{s_0}\big] = \mathbb{E}\big[Y_i(1) - Y_i(0) | Z_i = 1, G_i = g, i \in \mathcal{U}_{s_0}\big].$$

Under the allocation rule of the Italian university grants, $Z_i = W_i^{\text{obs}}$ for always-applicants and compliant-applicants. Therefore, the local randomization assumption allows the estimands τ_{AA,s_0}, τ_{CA,s_0} and τ_{s_0} to be interpreted as causal effects of receiving a grant for subpopulations of students who actually receive a grant, analogous to the notion of average treatment effect for the treated (ATT).

3.2. *Two additional assumptions*. The following two assumptions—though not necessary for Bayesian inference—are plausible in our study and can sharpen the inference.

ASSUMPTION 5. Monotonicity of application status:

$$A_i(1) \geq A_i(0), \qquad \text{for all } i \in \mathcal{U}_{s_0}.$$

ASSUMPTION 6. Stochastic exclusion restriction for never-applicants:

$$\Pr\big(Y_i(1) | G_i = \text{NA}, i \in \mathcal{U}_{s_0}\big) = \Pr\big(Y_i(0) | G_i = \text{NA}, i \in \mathcal{U}_{s_0}\big).$$

Monotonicity rules out the existence of defiant-applicants. The exclusion restriction rules out direct effects of eligibility on dropout for never-applicants. Never-applicants are students who would never apply for a grant irrespective of their eligibility status. These students would not receive the grant in any case. The exclusion restriction for never-applicants (Assumption 6) is of a very different nature from the exclusion restriction for compliant-applicants and always-applicants (Assumption 1): Assumption 6 has implications for inference but not for interpretation, whereas Assumption 1 is made solely for interpreting causal effects of assignment on the outcome as causal effects of treatment on the outcome. More discussions on the difference can be found in Imbens and Rubin (2015), Chapter 23 and Mealli and Pacini (2013).

3.3. *Selection of the subpopulations*. An important issue in practice is how to select the subpopulation $\mathcal{U}_{s_0}$. There can be a diverse choice of the shape of the subpopulation. In this paper, we limit our choice to symmetric intervals about s_0 for convenience. Specifically, we assume the following:

ASSUMPTION 7. There exists $h > 0$ such that for each $\varepsilon > 0$, $\Pr(s_0 - h \leq S_i \leq s_0 + h) > 1 - \varepsilon$, for each $i \in \mathcal{U}_{s_0}$.

Assumption 7 allows us to focus on the specific subsets of symmetric intervals among all neighborhoods of different shape around the threshold, s_0. Note that Assumptions 2 and 7 do not imply that $\mathcal{U}_{s_0}$ is unique. They only require that there exists at least one subpopulation, $\mathcal{U}_{s_0}$. Consequently, we are not interested in finding the largest h, but we only aim at determining plausible values for h.

Our approach for selecting bandwidth h exploits the "local" nature of Assumption 4, in the sense that it holds for a subset of units, but may not hold in general for other units. Therefore, under Assumption 4, covariates should be well balanced in the two subsamples defined by assignment in $\mathcal{U}_{s_0}$, and thus any test of the null hypothesis of no effect of assignment on covariates should fail to reject the null.

Assessing balance in the observed covariates raises problems of multiple comparisons, which may lead to a much higher than planned type I error if they are ignored [e.g., Benjamini and Hochberg (1995)]. We account for multiplicities using a Bayesian hierarchical mixed model, which provides an explicit method for borrowing information across covariates [e.g., Berry and Berry (2004), Scott and Berger (2006)]. Following Berry and Berry (2004), we use a mixture for the prior distribution of the eligibility parameters by assigning a point mass on equality of the means of the covariates between eligible and ineligible units. This Bayesian procedure provides a measure of the risk (posterior probability) that a chosen interval around s_0 defines a subpopulation of units that does not exactly match any true $\mathcal{U}_{s_0}$, including units for which our RD assumptions do not hold. More details are given in Section 5. Cattaneo, Frandsen and Titiunik (2015) also exploit balance tests of covariates to select $\mathcal{U}_{s_0}$, but their approach aims at selecting the *largest* subpopulation and does not account for multiple comparisons.

Conventional RD approaches based on local polynomial regression also involve bandwidth selection, but for a very different objective, namely, finding an optimal balance between precision and bias at the threshold for local polynomials [e.g., Imbens and Kalyanaraman (2012), Lee and Lemieux (2010), Ludwig and Miller (2007)], whereas our objective is to find a subpopulation where the RD assumptions are plausible and the selected subpopulation defines the target population.

3.4. *Mode of inference*. Once the subpopulation $\mathcal{U}_{s_0}$ is chosen, and under the RD Assumptions 2–4, one can choose different modes of inference. For example, under the additional Assumptions 5 and 6, the average causal effect for compliers, τ_{s_0}, is nonparametrically identified and can be estimated using standard

moment-based methods. But the average causal effects for always-applicants and compliant-applicants, τ_{AA,s_0} and τ_{CA,s_0}, are only nonparametrically partially identified [Mealli and Pacini (2013)]. One can also use likelihood approaches to parametrically estimate causal effects [e.g., Frumento et al. (2012), Mercatanti (2013)] or use randomization-based inference [Cattaneo, Frandsen and Titiunik (2015)].

We choose Bayesian inference for the following reasons. First, causal inference in RD designs usually involves complex observational data, with multiple sources of uncertainties, including the missing potential outcomes; the Bayesian approach is particularly useful for accounting for uncertainties and for pooling information from the data in such complex settings. Second, RD analysis usually relies on a sample of units with values of the forcing variable close to a single point, the size of which may be small; Bayesian methods, not relying on asymptotic approximations, are attractive in dealing with small samples. Third, in the Bayesian paradigm, the missing potential outcomes are treated as random variables, and all inferences are based on the posterior distributions of causal estimands, which are functions of potential outcomes. Thus, inference about sample-average and population-average estimands can be drawn using the same inferential procedures. Finally, covariates can be easily incorporated in the Bayesian approach, which may reduce posterior variability of the estimates.

4. Bayesian inference. Our Bayesian analysis follows the seminal works of Rubin (1978) and Imbens and Rubin (1997). Below we give a brief outline for conducting Bayesian principal stratification analysis; the readers may refer to the existing literature for more details [e.g., Elliott, Raghunathan and Li (2010), Mattei, Li and Mealli (2013), Schwartz, Li and Mealli (2011)]. Throughout the discussion, we use $p(\cdot|\cdot)$ and $\boldsymbol{\theta}_{\cdot|\cdot}$ to denote generic conditional distributions and the corresponding parameters, respectively.

Nine quantities are associated with each unit: $Y_i(0)$, $Y_i(1)$, $W_i(0)$, $W_i(1)$, $A_i(0)$, $A_i(1)$, $\mathbf{X}_i$, Z_i, S_i. Among these, S_i completely determines Z_i; the principal stratum $G_i = (A_i(0), A_i(1))$ and S_i completely determine $(W_i(0), W_i(1))$. Therefore, inference for causal effects involves only $Y_i(0)$, $Y_i(1)$, $A_i(0)$, $A_i(1)$, $\mathbf{X}_i$, S_i, of which four are observed: S_i, $\mathbf{X}_i$, $A_i^{\mathrm{obs}} = A_i(Z_i)$, $Y_i^{\mathrm{obs}} = Y_i(Z_i)$, and two are unobserved: $A_i^{\mathrm{mis}} = A_i(1 - Z_i)$, $Y_i^{\mathrm{mis}} = Y_i(1 - Z_i)$.

Bayesian inference considers the observed values to be realizations of random variables and the unobserved values to be unobserved random variables. Let $p(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{A}(0), \mathbf{A}(1), \mathbf{X}, \mathbf{S}; \mathcal{U}_{s_0})$ denote the joint probability density function of these random variables for all units in $\mathcal{U}_{s_0}$. We assume this distribution is unit-exchangeable, that is, it is invariant under a permutation of the unit indices. Then, with essentially no loss of generality, by appealing to de Finetti's theorem [de Finetti (1937)], we can assume that there exists an unknown parameter vector $\boldsymbol{\theta}$, which is itself a random variable having a known prior distribution $p(\boldsymbol{\theta})$

such that

$$p\big(\mathbf{Y}(0),\mathbf{Y}(1),\mathbf{A}(0),\mathbf{A}(1),\mathbf{X},\mathbf{S};\mathcal{U}_{s_0}\big) = \int \prod_{i\in\mathcal{U}_{s_0}} p\big(Y_i(0),Y_i(1),A_i(0),A_i(1),\mathbf{X}_i,S_i|\boldsymbol{\theta}\big)p(\boldsymbol{\theta})\,d\boldsymbol{\theta}.$$

Bayesian inference of the causal estimands, which are functions of $Y_i(z)$'s and $A_i(z)$'s, centers around deriving the posterior distribution for the parameter vector of their distribution, denoted by $\boldsymbol{\theta}_{Y,G}$. Under Assumption 4, and assuming the parameters governing the distributions of the covariates, the forcing variable and the potential outcomes are a priori distinct and independent, the posterior distribution of $\boldsymbol{\theta}_{Y,G}$ can be written as follows:

$$\begin{aligned} &p\big(\boldsymbol{\theta}_{Y,G}|\mathbf{Y}^{\text{obs}},\mathbf{A}^{\text{obs}},\mathbf{X},\mathbf{S};\mathcal{U}_{s_0}\big) \\ &\propto p(\boldsymbol{\theta}_{Y|G})\times p(\boldsymbol{\theta}_G) \\ &\quad\times \prod_{i\in\mathcal{U}_{s_0}}\left[\int\!\!\int p\big(Y_i(0),Y_i(1)|G_i,\mathbf{X}_i;\boldsymbol{\theta}_{Y|G}\big)p(G_i|\mathbf{X}_i;\boldsymbol{\theta}_G)\,dY_i^{\text{mis}}\,dA_i^{\text{mis}}\right]. \end{aligned} \tag{6}$$

The above decomposition suggests that two models need to be specified for model-based inference: (i) the model for potential outcomes conditional on principal strata and covariates, and (ii) the model for principal strata conditional on covariates, as well as the prior distribution for the parameters, $p(\boldsymbol{\theta}_{Y,G})$, with $\boldsymbol{\theta}_{Y,G}=(\boldsymbol{\theta}_G,\boldsymbol{\theta}_{Y|G})$.

Let $\pi_{i,g}=\Pr(G_i=g|\mathbf{X}_i;\boldsymbol{\theta}_G)$ and $f_{i,gz}=p(Y_i(z)|G_i=g,\mathbf{X}_i;\boldsymbol{\theta}_{Y|G})$. Then the posterior distribution of $\boldsymbol{\theta}_{Y,G}$ given the observed data can be written as follows:

$$\begin{aligned} &p\big(\boldsymbol{\theta}_{Y,G}|\mathbf{Y}^{\text{obs}},\mathbf{A}^{\text{obs}},\mathbf{X},\mathbf{S};\mathcal{U}_{s_0}\big) \\ &\propto p(\boldsymbol{\theta}_{Y,G})\times \prod_{i\in\mathcal{U}_{s_0}:S_i>s_0,A_i^{\text{obs}}=0}(\pi_{i,\text{CA}}f_{i,\text{CA},0}+\pi_{i,\text{NA}}f_{i,\text{NA}}) \\ &\quad\times \prod_{i\in\mathcal{U}_{s_0}:S_i>s_0,A_i^{\text{obs}}=1}\pi_{i,\text{AA}}f_{i,\text{AA},0}\times \prod_{i\in\mathcal{U}_{s_0}:S_i\le s_0,A_i^{\text{obs}}=0}\pi_{i,\text{NA}}f_{i,\text{NA}} \\ &\quad\times \prod_{i\in\mathcal{U}_{s_0}:S_i\le s_0,A_i^{\text{obs}}=1}(\pi_{i,\text{AA}}f_{i,\text{AA},1}+\pi_{i,\text{CA}}f_{i,\text{CA},1}), \end{aligned} \tag{7}$$

where $f_{i,\text{NA}}=f_{i,\text{NA},0}=f_{i,\text{NA},1}$ by the exclusion restriction (Assumption 6). The likelihood function, specified by the four products, does not depend on the association between the potential outcomes $Y_i(0)$ and $Y_i(1)$. Therefore, the posterior distribution of the association parameters equal their prior distribution as long as the association parameters are a priori independent of the other parameters, as we assume henceforth. The population-average causal estimands τ_{AA,s_0}, τ_{CA,s_0} and

τ_{s_0} are functions of the parameter vector $\boldsymbol{\theta}_{Y,G}$, which is free of the association parameters, therefore, inference for them does not involve the association parameters [also see discussion in Imbens and Rubin (1997)]. Inference for sample-average causal estimands for the units in the study, on the other hand, does generally involve the association parameters. In our application inference for sample-average causal estimands is drawn under the assumption that for each unit i, potential outcomes, $Y_i(0)$ and $Y_i(1)$, are independent conditional on $\mathbf{X}_i$ and $\boldsymbol{\theta}$.

Posterior inference of $\boldsymbol{\theta}_{Y,G}$ can be obtained using Gibbs sampling with a data augmentation step to impute the missing A_i^{mis}, iteratively drawing from the two posterior predictive distributions, $p(\boldsymbol{\theta}_{Y,G}|\mathbf{Y}^{\text{obs}}, \mathbf{A}^{\text{obs}}, \mathbf{A}^{\text{mis}}, \mathbf{X}, \mathbf{S}; \mathcal{U}_{s_0})$ and $p(\mathbf{A}^{\text{mis}}|\mathbf{Y}^{\text{obs}}, \mathbf{A}^{\text{obs}}, \mathbf{X}, \mathbf{S}, \boldsymbol{\theta}_{Y,S}; \mathcal{U}_{s_0})$.

Specification of $\pi_{i,g}$, $f_{i,gz}$ and the corresponding prior to posterior computation depends on the specific application. Details of the models and computation in our application will be provided in Section 5. As a general guideline, we recommend to specify $\pi_{i,g}$ and $f_{i,gz}$ conditional on both covariates $\mathbf{X}$ and the forcing variable S, even though equation (6) suggests conditioning on S is not required. Indeed, if the true subpopulations $\mathcal{U}_{s_0}$ were known, in theory, we would not need to adjust for S, because local randomization guarantees that for units in $\mathcal{U}_{s_0}$ values of the forcing variable falling above or below the threshold are independent of the potential outcomes. However, in practice, the true subpopulations $\mathcal{U}_{s_0}$ are usually unknown and once a subpopulation has been selected, that is, once a value for h, say h^*, has been chosen, there may be some units with a realized value of S between $s_0 - h^*$ and $s_0 + h^*$ who do not belong to $\mathcal{U}_{s_0}$. For these units there may be a relationship between the forcing variable and potential outcomes, and these potential dependences need to be modeled. Specifically, systematic differences in the forcing variable S that, by definition, occur between eligible and ineligible units may affect inference in the presence of students who do not belong to $\mathcal{U}_{s_0}$.

5. Evaluation of Italian university grants.

5.1. *Data*. We apply the proposed method to the data from the cohort of first-year students enrolled in 2004 to 2006 at University of Pisa and University of Florence. For each student, information on grant application status (A_i^{obs}), grant receipt status (W_i^{obs}) at the beginning of the academic year, dropout status at the end of the academic year, and covariates ($\mathbf{X}_i$) is obtained from the ministry of education and university administrative records. The forcing variable S is a combined economic measure of each student, calculated from one's income tax return and property adjusted for family size based on a formula that is typically not fully known to the students. In all three years, the threshold of eligibility is $s_0 = 15{,}000$ euros. Thus, the eligibility status (Z) is also observed. Typically, students need support from fiscal experts to compute their value of S, and the income revenue authority conducts random inspections to verify that the official tax returns are reported. These factors make it extremely difficult, if not impossible, for students

TABLE 1
Summary statistics of the first-year students enrolled in 2004–2006 *at Universities of Pisa and Florence, for the students with* $S_i \in (14{,}000, 16{,}000)$ *euros* (*i.e.*, $h = 1000, s_0 = 15{,}000$)

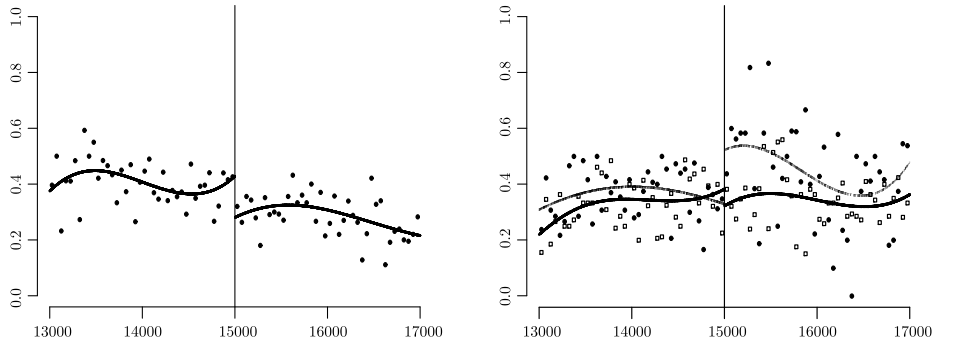
	$Z = 0$		$Z = 1$	
Variable	$A^{\text{obs}} = 0$	$A^{\text{obs}} = 1$	$A^{\text{obs}} = 0$	$A^{\text{obs}} = 1$
Sample Size	657	304	703	444
Dropout	0.36	0.50	0.35	0.36
S (euros)	15,495	15,509	14,504	14,499
Female	0.59	0.61	0.60	0.55
HS grade	80.80	84.35	80.17	84.47
University (Pisa)	0.37	0.51	0.37	0.51

to manipulate the value of S in order to end up on the right side of the threshold. Therefore, we argue that the local randomization assumption is reasonable here. Ineligible students apply either because they are not fully aware of their eligibility status or because they hope that their application will be still considered because of extra funding or other considerations.

Covariates include sex, high school grade, high school type (4 categories), major in university (6 categories), indicator of year of enrollment (2004, 2005, 2006) and indicator of university (Pisa vs. Florence). The data only include students who had a high school grade of at least 70/100 and applied either for a grant or for a reduction of tuition fee. Summary statistics of important variables for the students with S within 1000 euros of the threshold are given in Table 1. An unadjusted comparison would suggest that the applicants have higher high-school grades, which is an important indicator of a student's academic performance, but also higher dropout rate regardless of their eligibility status.

Application rate and dropout rate as a function of S among the students are given in Figure 1. The overall dropout rate is high, consistently between 30% to 50% regardless of the economic measure. From the fitted lines using local logistic polynomial models of order 3 on the two sides of the threshold, discontinuity is clearly visible in both application rate and dropout rate at the threshold. As the economic measure increases, application rate steadily decreases, while the trend in dropout rate has a concave change at the threshold, increasing on the left of the threshold and decreasing on the right.

5.2. *Selection of the subpopulation*. We apply the approach discussed in Section 3.3 to find subpopulations of units where our RD assumptions hold. Specifically, we use a hierarchical Bayesian model for assessing the balance of the covariates between eligibility groups. We specify probit models for binary variables, conditional probit models for categorical variables, and Gaussian models for continuous variables. Formally, we assume that $X_j \sim N(\gamma_{0j} + \gamma_{1j} Z_i, \sigma_j^2)$ if



(a) Application rate (b) Dropout rate.

FIG. 1. *Application rate* (a) *and dropout rate* (b) *as a function of the forcing variable for the first-year students in Universities of Florence and Pisa in* 2004–2006. *The smoothed lines are estimated using polynomial logistic regression models* (*of order* 3) *on each side of the threshold separately*; *each point is calculated from the units within a binwidth of* 50 *euros*.

X_j is continuous, and $\Pr(X_{ij} = 1) = \Pr(X_{ij}^* > 0)$ with $X_{ij}^* \sim N(\gamma_{0j} + \gamma_{1j} Z_i, 1)$ if X_j is binary. If X_j is a categorical variable taking on K values, we assume that $\Pr(X_{ij} = 1) = \Pr(X_{ij}^{*(1)} \le 0)$, and $\Pr(X_{ij} = k) = \Pr(\bigcap_{\ell=1}^{k-1} \{X_{ij}^{*(\ell)} > 0\} \cap X_{ij}^{*(k)} \le 0)$ for $k = 2, \ldots, K-1$, where $X_{ij}^{*(k)} \sim N(\gamma_{0j}^{(k)} + \gamma_{1j}^{(k)} Z_i, 1)$, $k = 1, \ldots, K-1$, independently. Let $\boldsymbol{\gamma}_{0j} = (\gamma_{0j}^{(1)}, \ldots, \gamma_{0j}^{(K-1)})'$ and $\boldsymbol{\gamma}_{1j} = (\gamma_{1j}^{(1)}, \ldots, \gamma_{1j}^{(K-1)})'$.

We specify the following prior distributions for the model parameters. The variances of the continuous variables have an inverse-Gamma distribution: $\sigma_j^2 \sim \mathrm{IG}(\underline{a}, \underline{b})$. The γ_0's have Gaussian prior distributions: for continuous and binary variables, $\gamma_{0j} \sim N(\mu_{\gamma_0}, \sigma_{\gamma_0}^2)$, and for categorical variables, $\boldsymbol{\gamma}_{0j} \sim N(\mu_{\gamma_0} \mathbf{u}_{K-1}, \sigma_{\gamma_0}^2 \mathbf{I}_{K-1})$ with $\mathbf{u}_{K-1}$ and $\mathbf{I}_{K-1}$ being the $K-1$-dimensional vector of ones and the identity matrix of order $K-1$, respectively. Further, for continuous and binary variables, parameters γ_{1j} are the difference between means/proportions for eligible and ineligible units. If $\gamma_{1j} = 0$, then X_j has the same distribution for eligible and ineligible units. For a categorical variable taking on K values, the proportion of units in each category is the same for eligible and ineligible units if and only if $\gamma_{1j}^{(k)} = 0$ for each $k = 1, \ldots, K-1$. We assign positive probability to these possibilities using the following mixture prior distributions:

$$\gamma_{1j} \sim \pi_{\gamma_1} \delta_0(\gamma_{1j}) + (1 - \pi_{\gamma_1}) N\big(\mu_{\gamma_1}, \sigma_{\gamma_1}^2\big)$$

and

$$\boldsymbol{\gamma}_{1j} \sim \prod_{k=1}^{K-1} \big[\pi_{\gamma_1} \delta_0\big(\gamma_{1j}^{(k)}\big) + (1 - \pi_{\gamma_1}) N\big(\mu_{\gamma_1}, \sigma_{\gamma_1}^2\big)\big],$$

where $\delta_0(\cdot)$ is the Dirac delta distribution.

For the hyperparameters, we assign the following prior distributions: $\mu_{\gamma_0} \sim N(\underline{\mu}_{\gamma_0}, \underline{\sigma}^2_{\gamma_0})$; $\sigma^2_{\gamma_0} \sim \text{IG}(\underline{a}_{\gamma_0}, \underline{b}_{\gamma_0})$; $\mu_{\gamma_1} \sim N(\underline{\mu}_{\gamma_1}, \underline{\sigma}^2_{\gamma_1})$; $\sigma^2_{\gamma_1} \sim \text{IG}(\underline{a}_{\gamma_1}, \underline{b}_{\gamma_1})$; and $\pi_{\gamma_1} \sim \text{Beta}(\underline{a}_{\pi}, \underline{b}_{\pi})$.

We implement the Bayesian model for assessing the balance of covariates on the two sides of the threshold for various subpopulations defined by different values of h. Details of the Monte Carlo Markov Chain (MCMC) for the posterior computation are relegated to the supplementary article [Li, Mattei and Mealli (2015)]. Table 2 shows the posterior probabilities that the covariates have the same distribution between eligible and ineligible students for the subpopulations defined by $h = 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 4000, 5000$. These values show that the probability of the pre-assignment variables being well balanced is high for subpopulations defined by values of h strictly lower than 1500: the vast majority of these probabilities are larger than or close to 0.8. Note that the probabilities are in general lower among the covariates of "major in university," suggesting these covariates may not be as balanced as other covariates. Nonetheless, nearly all these probabilities are still higher 0.6 with a single lowest probability being 0.565 (Tech major in university). For larger subpopulations, some covariates, such as the "indicator of university," are clearly unbalanced.

Given that the risk that a chosen interval around the threshold defines a subpopulation that includes units not belonging to the target subpopulation, $\mathcal{U}_{s_0}$, is not zero, in order to account for the presence of these units, we conduct the subsequent analyses conditioning on both covariates and the realized values on the forcing variable. Also, we evaluate the robustness of our results conducting analyses using various values of h ($h = 500, 1000, 1500$).

5.3. *Parametric models*. For the units within the selected subpopulation $\mathcal{U}_{s_0}$, we assume parametric models for the outcome (f_{gz}) and principal strata (π_g). Alternative models, such as Student-t models [Chib and Jacobi (2015)] and Bayesian nonparametric models [Schwartz, Li and Mealli (2011)], can be considered. Note that although we are using parametric models, identification does not rely on parametric assumptions. The model for the principal strata of application consists of two conditional probit models:

$$\begin{aligned}
\pi_{i,\text{AA}} &= \Pr\bigl(G_i^*(\text{AA}) \le 0\bigr),\\
\pi_{i,\text{NA}} &= \Pr\bigl(G_i^*(\text{AA}) > 0 \text{ and } G_i^*(\text{NA}) \le 0\bigr),\\
\pi_{i,\text{CA}} &= 1 - \pi_{i,\text{AA}} - \pi_{i,\text{NA}},
\end{aligned}$$

where

$$\begin{aligned}
G_i^*(\text{AA}) &= \alpha_{\text{AA},0} + \alpha_{\text{AA}}^{(S)} S_i^* + \mathbf{X}_i' \boldsymbol{\alpha}_{\text{AA}}^{(X)} + \varepsilon_{\text{AA},i},\\
G_i^*(\text{NA}) &= \alpha_{\text{NA},0} + \alpha_{\text{NA}}^{(S)} S_i^* + \mathbf{X}_i' \boldsymbol{\alpha}_{\text{NA}}^{(X)} + \varepsilon_{\text{NA},i},
\end{aligned}$$

with $\varepsilon_{\text{AA},i} \sim N(0,1)$, $\varepsilon_{\text{NA},i} \sim N(0,1)$ independently, and $S_i^* = (S_i - s_0)/1000$.

TABLE 2

Posterior probabilities that the covariates have the same distribution between eligible and ineligible students for various subpopulation

Variable	$h = 250$ ($n = 528$)	$h = 500$ ($n = 1042$)	$h = 750$ ($n = 1577$)	$h = 1000$ ($n = 2108$)	$h = 1500$ ($n = 3166$)	$h = 2000$ ($n = 4197$)	$h = 2500$ ($n = 5159$)	$h = 3000$ ($n = 6113$)	$h = 4000$ ($n = 8061$)	$h = 5000$ ($n = 9846$)
Sex	0.955	0.950	0.960	0.962	0.977	0.970	0.991	0.960	0.968	0.797
High school type (Baseline: Other)										
Humanity	0.951	0.952	0.949	0.955	0.979	0.970	0.965	0.986	0.953	0.962
Science	0.894	0.905	0.926	0.927	0.951	0.889	0.916	0.926	0.045	0.000
Tech	0.790	0.807	0.790	0.808	0.819	0.619	0.751	0.793	0.003	0.000
HS grade	0.955	0.958	0.972	0.978	0.971	0.981	0.987	0.990	0.984	0.986
Year (Baseline: 2004)										
2005	0.932	0.964	0.954	0.926	0.973	0.977	0.976	0.983	0.861	0.918
2006	0.883	0.918	0.914	0.909	0.959	0.934	0.952	0.970	0.807	0.884
University (Pisa)	0.950	0.916	0.971	0.983	0.686	0.097	0.225	0.300	0.082	0.000
Major in university (Baseline: Other)										
Humanity	0.946	0.899	0.689	0.797	0.798	0.932	0.958	0.990	0.964	0.946
Science	0.894	0.857	0.660	0.751	0.783	0.901	0.929	0.966	0.911	0.913
Social science	0.798	0.821	0.624	0.713	0.758	0.864	0.913	0.953	0.878	0.858
Bio-Med	0.728	0.776	0.604	0.677	0.736	0.837	0.889	0.926	0.839	0.832
Tech	0.632	0.634	0.565	0.624	0.699	0.794	0.863	0.876	0.719	0.453

Dropout, the primary outcome in our application, is binary. Therefore, we assume the following generalized linear outcome model with a probit link [Albert and Chib (1993)]:

$$\Pr(Y_i(z)=1|G_i=g,S_i,\mathbf{X}_i)=\Phi(\beta_{0,g,z}+\beta_{g,z}^{(S)}S_i^*+\mathbf{X}_i'\boldsymbol{\beta}_{g,z}^{(X)}).$$

We impose prior equality of the slope coefficients in the outcome regressions: $\boldsymbol{\beta}_{g,z}^{(X)}\equiv\boldsymbol{\beta}^{(X)}$ for $g=\text{AA, CA, NA}$ and $z=0,1$.

Define $\boldsymbol{\alpha}_g=[\alpha_{g0},\alpha_g^{(S)},\boldsymbol{\alpha}_g^{(X)}]'$, $g=\text{AA, NA}$, and $\boldsymbol{\beta}_{g,z}=[\beta_{0,g,z},\beta_{g,z}^{(S)}]'$, $g=\text{AA, CA, NA}$; $z=0,1$. By Assumption 6, $\boldsymbol{\beta}_{\text{NA},0}=\boldsymbol{\beta}_{\text{NA},1}$. We assume that parameters are a priori independent and use multivariate normal prior distributions:

$$\boldsymbol{\alpha}_g\sim N(\underline{\boldsymbol{\mu}}_{\alpha_g};\underline{\sigma}^2_{\alpha_g}\mathbf{I}),\qquad \boldsymbol{\beta}_{g,z}\sim N(\underline{\boldsymbol{\mu}}_{\beta_{g,z}};\underline{\sigma}^2_{\beta_{g,z}}\mathbf{I}),\qquad \boldsymbol{\beta}^{(X)}\sim N(\underline{\boldsymbol{\mu}}_{\beta};\underline{\sigma}^2_{\beta}\mathbf{I}),$$

where $\mathbf{I}$ is the identity matrix. We specify flat priors setting the hyperparameters as follows: setting $\underline{\boldsymbol{\mu}}_{\alpha_g}$, $\underline{\boldsymbol{\mu}}_{\beta_{g,z}}$, $\underline{\boldsymbol{\mu}}_{\beta}$ to be null vectors; and setting large prior variances $\underline{\sigma}^2_{\alpha_g}=10$, $\underline{\sigma}^2_{\beta_{g,z}}=10$, $\underline{\sigma}^2_{\beta}=10$ for $g=\text{AA, CA, NA}$; $z=0,1$.

5.4. *Posterior computation*. Upon obtaining the posterior draws of the parameters, we calculate three estimates for each causal estimand: population-average effect *within* $\mathcal{U}_{s_0}$ and *at* s_0, and sample-average effect within $\mathcal{U}_{s_0}$. The population-average effects within $\mathcal{U}_{s_0}$ are calculated averaging the model-based dropout proportions over the empirical distribution of the pre-assignment variables and the forcing variable:

$$\frac{\sum_{i\in\mathcal{U}_{s_0}}\pi_{i,g}\Phi(\beta_{0,g,1}+\beta_{g,1}^{(S)}S_i^*+\mathbf{X}_i'\boldsymbol{\beta}^{(X)})}{\sum_{i\in\mathcal{U}_{s_0}}\pi_{i,g}}$$
$$-\frac{\sum_{i\in\mathcal{U}_{s_0}}\pi_{i,g}\Phi(\beta_{0,g,0}+\beta_{g,0}^{(S)}S_i^*+\mathbf{X}_i'\boldsymbol{\beta}^{(X)})}{\sum_{i\in\mathcal{U}_{s_0}}\pi_{i,g}},$$

for $g=\text{AA, CA, \{AA, CA\}}$. The population-average effects at s_0 are calculated in a similar way, setting $S_i^*=0$ (i.e., $S_i=s_0$) for each i. To obtain the sample-average estimates, we compute the posterior predictive distributions of the potential outcomes for each student i in $\mathcal{U}_{s_0}$.

5.5. *Results*. We conducted Bayesian analysis using $h=500$, 1000, 1500. Posterior inference is based on 5000 draws from the posterior distributions simulated using single chains, which were run for 125,000 iterations. To assess convergence of iterative simulation methods, we calculated the Cramer–von-Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution and visually inspected the trace-plots of the causal parameters (functions of model parameters). We also run multiple MCMC chains from different starting values from each h to evaluate the mixing of the chains using the

TABLE 3

Posterior median and 95% credible intervals of principal strata proportion and super-population and finite-sample causal effects on dropout for always-applicants (τ_{AA,s_0}), compliant-applicants (τ_{CA,s_0}) and their union (τ_{s_0}), for the subpopulation within different bandwidths h around the threshold

	Population-average		**Sample-average**		**Population-average at s_0**	
h	**Median**	**95% CI**	**Median**	**95% CI**	**Median**	**95% CI**
			$h = 500$			
$\Pr(G_i = \text{AA})$	0.323	(0.294; 0.355)	0.322	(0.309; 0.336)	0.320	(0.291; 0.352)
$\Pr(G_i = \text{CA})$	0.060	(0.031; 0.105)	0.041	(0.021; 0.090)	0.058	(0.030; 0.094)
$\Pr(G_i = \text{NA})$	0.616	(0.570; 0.650)	0.637	(0.590; 0.651)	0.621	(0.583; 0.654)
τ_{AA,s_0}	−0.153	(−0.313; −0.030)	−0.152	(−0.307; −0.038)	−0.154	(−0.298; −0.030)
τ_{CA,s_0}	0.045	(−0.170; 0.497)	0.074	(−0.256; 0.545)	0.039	(−0.169; 0.474)
τ_{s_0}	−0.116	(−0.253; −0.005)	−0.120	(−0.265; −0.009)	−0.120	(−0.245; −0.012)
			$h = 1000$			
$\Pr(G_i = \text{AA})$	0.336	(0.312; 0.365)	0.333	(0.318; 0.354)	0.335	(0.311; 0.363)
$\Pr(G_i = \text{CA})$	0.043	(0.002; 0.086)	0.027	(0.002; 0.075)	0.043	(0.001; 0.075)
$\Pr(G_i = \text{NA})$	0.623	(0.584; 0.652)	0.640	(0.599; 0.645)	0.625	(0.594; 0.656)
τ_{AA,s_0}	−0.161	(−0.273; −0.052)	−0.161	(−0.270; −0.057)	−0.154	(−0.259; −0.052)
τ_{CA,s_0}	0.028	(−0.745; 0.828)	0.031	(−0.778; 0.871)	0.010	(−0.918; 0.933)
τ_{s_0}	−0.132	(−0.242; −0.021)	−0.139	(−0.247; −0.034)	−0.128	(−0.229; −0.020)
			$h = 1500$			
$\Pr(G_i = \text{AA})$	0.332	(0.315; 0.349)	0.332	(0.326; 0.337)	0.329	(0.312; 0.346)
$\Pr(G_i = \text{CA})$	0.042	(0.035; 0.077)	0.027	(0.020; 0.066)	0.042	(0.036; 0.062)
$\Pr(G_i = \text{NA})$	0.625	(0.591; 0.642)	0.642	(0.605; 0.644)	0.628	(0.606; 0.646)
τ_{AA,s_0}	−0.183	(−0.286; −0.077)	−0.187	(−0.291; −0.085)	−0.153	(−0.247; −0.063)
τ_{CA,s_0}	0.010	(−0.304; 0.797)	0.011	(−0.207; 0.928)	0.000	(−0.154; 0.951)
τ_{s_0}	−0.153	(−0.256; −0.040)	−0.165	(−0.266; −0.057)	−0.130	(−0.217; −0.019)

Gelman–Rubin statistic [Gelman and Rubin (1992)]. The results provided no evidence against convergence.[4]

Table 3 shows posterior medians and 95% credible intervals for the principal strata proportions under monotonicity and for the causal parameters τ_{AA,s_0}, τ_{CA,s_0}, τ_{s_0}, for bandwidths ranging from 500 to 1500 euros. The results are robust across different bandwidths. The estimated proportions of the principal strata are very similar across different h: there are more than 61% never-applicants, more than

[4]We also conducted Bayesian analysis using alternative models with different order polynomials in S as well as models conditioning only on S (without using the pretreatment variables) and null models, conditioning on neither S nor the pretreatment covariates. Consistent with results found in Mealli and Rampichini (2012), higher order polynomials do not lead to substantial inferential benefits, and posterior distributions of the causal effects of interest did not substantially change with the alternative models, so here we only show the results based on models conditional on both S and the pretreatment covariates.

32% always-applicants and less than 6.5% compliant-applicants. The three estimates for the same causal parameter are also similar. The posterior distributions of the causal effect for always-applicants, τ_{AA,s_0}, and the union of always-applicants and compliant-applicants, τ_{s_0}, are centered on negative values, and the 95% credible intervals do not cover 0, irrespective of the choice of the bandwidth.

For instance, consider the finite-sample causal effects for the subpopulation within $h = 1000$ euros around the threshold (middle block of columns in Table 3). The estimated τ_{s_0} suggests a 13.9% [95% CI: (3.4%; 24.7%)] reduction in dropout rate for the students who receive the grants. The estimated τ_{AA,s_0} suggests an even stronger positive effect among the always-applicants: a 16.1% [95% CI: (5%; 27%)] reduction in dropout rate. In fact, τ_{s_0}, which is a weighted average of the effects for always-applicants and compliant-applicants, appears to be diluted by the somewhat surprising small and negative effect among the compliant-applicant. However, the data do not seem to contain much information on compliant-applicants (the estimated proportion of compliant-applicants is very small, less than 5%), and the effects were estimated with large uncertainties.

These results suggest that the current Italian university grants are effective in reducing dropout from universities among students from families with annual economic measure around 15,000 euros, and also reveal that the effects are heterogeneous among always-applicants and compliant-applicants. From a cost-effective perspective, our results suggest that it might be beneficial for the education administration to weaken the eligibility criteria (i.e., increase the threshold) to allow more applicants and thus always-applicants to get the grant. The combination of a low percentage of compliant-applicants and a high percentage of always-applicants suggests that most students with the economic measure being around the threshold who intend to apply for the grants would apply irrespective of their eligibility. From a policy perspective, this implies that the educational administration should better explain the rule of eligibility to potential applicants to discourage ineligible students from applying, and thus reduce unnecessary efforts from these students and the administration for processing these applications.

5.6. *Posterior predictive model checking.* Assessing the plausibility of model assumptions is critical in model-based approaches. Model checking here is not as crucial as in other model-based approaches, thanks to the randomization assumption, but it is still prudent to check the model fit since there are uncertainties in the selection of $\mathcal{U}_{s_0}$. We adopt Bayesian posterior predictive checks [Gelman, Meng and Stern (1996)] to assess goodness of fit of our models in the application. Posterior predictive checks evaluate goodness of fit of models by measuring the discrepancy between the observed data and replicated data simulated from its posterior predictive distribution. The particular procedure adopted here is similar to that in Mattei, Li and Mealli (2013), Section 6. Specifically, we consider three discrepancy measures aimed at assessing whether the model can preserve broad

TABLE 4
Bayesian p-values of signal, noise and SNR under different h for the model used in the application to Italian university grants

h	Principal strata	Signal	Noise	SNR
500	{CA}	0.095	0.630	0.094
	{AA}	0.254	0.325	0.254
	{AA, CA}	0.338	0.273	0.370
1000	{CA}	0.411	0.425	0.419
	{AA}	0.400	0.444	0.473
	{AA, CA}	0.493	0.335	0.518
	{CA}	0.208	0.444	0.210
1500	{AA}	0.372	0.400	0.261
	{AA, CA}	0.455	0.337	0.470

features of signal, noise and signal-to-noise ratio (SNR) in the dropout status distribution for compliant-applicants, always-applicants and the union of these two principal strata, and calculate posterior predictive p-values (PPPVs) to summarize discrepancies between the observed data and replicated data. Extreme (close to 0 or 1) PPPVs can be interpreted as evidence of lack of fit of the model in, at least some aspects of, the observed data. Further details of the procedure are relegated to the supplementary article [Li, Mattei and Mealli (2015)].

Table 4 shows the PPPVs for the model fit to the subpopulation with bandwidth of 500, 1000 and 1500 euros, respectively. The PPPVs suggest good model fit for all bandwidths, except for a slight under-fit for always-applicants in the subpopulation with $h = 500$, which is possibly due to the small sample size. We have also calculated the less conservative sampled posterior predictive p-values [Gosselin (2011), Johnson (2007)] and obtained similar conclusions.

6. Discussion. Motivated by the evaluation of Italian university grants, we propose a probabilistic formulation of the assignment mechanism for regression discontinuity designs and develop a full Bayesian approach to draw causal inference within the framework of principal stratification. In particular, we illustrate how to utilize information on application status to gain additional insights in program evaluation. Applying the method to the data from two Italian universities, we find that university grants reduce dropping out of higher education for students from low-income families and the effect size is especially pronounced for motivated students (always-applicants). Because the amount of grants is the same for all awardees in the data, we were not able to study the dose-response relation between the amount of grants and dropout, which is a topic of potentially important policy implications that merits further investigation.

The core of our approach is the local randomization assumption, which is intrinsically nontestable. Therefore, it may be worthwhile to conduct sensitivity analyses

to assess the robustness of the results with respect to violations of local randomization. To this end, we conduct further analyses deriving the posterior distributions of the causal effects under three additional model specifications: (i) a model where we specify the model for principal strata, $\pi_{i,g}$, and the conditional model for potential outcomes given principal strata, $f_{i,gz}$, conditioning on neither the forcing variable nor the pretreatment variables; (ii) a model where we specify $\pi_{i,g}$ and $f_{i,gz}$ conditioning only on the forcing variable, without including the pretreatment variables; and (iii) a model where we specify $\pi_{i,g}$ and $f_{i,gz}$ conditioning only on the pretreatment variables, without including the forcing variable. Under local randomization, adjusting inference for either the forcing variable or the pretreatment variables should not be required, therefore, we expect that results are similar across different model specifications. Indeed, results, shown in the supplementary article [Li, Mattei and Mealli (2015)], are robust across different models, suggesting that causal inference under the local randomization assumption is robust and credible.

A fundamental distinction between our approach and the previous local-regression based RD approaches lies in the role of the forcing variable in the analysis. Specifically, previous approaches generally view the forcing variable as a pre-assignment covariate rather than a random variable as in our approach. As a consequence, the standard overlap assumption, which requires that there are both treated and control units for all values of the covariates including the forcing variable, is violated. Violation of the overlap assumption implies that the conditional independence assumption, which trivially holds in RD settings, cannot be exploited directly. Instead some kind of extrapolation is required and, in order to avoid that estimates heavily rely on extrapolation, previous analyses focus on causal effects of the treatment for units *at* the threshold. Smoothness assumptions, for example, continuity of conditional regression functions of potential outcomes given the forcing variable, are usually assumed to draw inference on those causal effects. Local randomization and continuity are different assumptions, leading to different causal estimands: under continuity assumptions, units with a realized value of the forcing variable around the threshold are used to draw inference on causal effects for units *at* the threshold, whereas under local randomization, inference is drawn on causal effects for units *around* the threshold.

In the evaluation of Italian university grants, student's academic performance (measured by total credits taken or passing rate of exams) is also of great interest in policy. As illustrated by Mattei, Li and Mealli (2013) and Mercatanti, Li and Mealli (2015), jointly modeling two outcomes, dropout and academic performance in this case, would be worthwhile for both practical and inferential purposes.

After the first year, the Italian university grant assignment rule combines sequential and RD designs [Cellini, Ferreira and Rothstein (2010)]: grants are allocated both on the basis of students' family economic indicator and on the ground of their academic performance (exam scores above a certain threshold). Such complex assignment mechanisms pose challenges to causal inference, requiring new structures and assumptions; meanwhile, they also present great opportunities for

extending the existing framework to more general RD settings. One specific direction of our future research is to develop methods that combine Bayesian tools for RDs and dynamic treatment regimes [Murphy (2003), Zajonc (2012)] in the presence of multiple forcing variables [Imbens and Zajonc (2011)].

Acknowledgments. The authors are grateful to the Associate Editor and two reviewers for constructive comments, to Quanli Wang for computing support, and to Guido Imbens, Michael Hudgens and Sid Chib for helpful discussions.

SUPPLEMENTARY MATERIAL

Supplement to "Evaluating the causal effect of university grants on student dropout: Evidence from a regression discontinuity design using principal stratification" (DOI: 10.1214/15-AOAS881SUPP; .pdf). We describe in detail the Bayesian approach we used to select the subpopulations, the Markov Chain Monte Carlo (MCMC) methods used to simulate the posterior distributions of the parameters of the models, the posterior predictive checks, and the sensitivity analysis regarding local randomization described in Section 6.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *J. Amer. Statist. Assoc.* **91** 444–472.

BATTISTIN, E. and RETTORE, E. (2008). Ineligibles and eligible non-participants as a double comparison group in regression discontinuity designs. *J. Econometrics* **142** 715–730.

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BERK, R. A. and DE LEUUW, J. (1999). An evaluation of California's inmate classification system using a generalized regression discontinuity design. *J. Amer. Statist. Assoc.* **94** 1045–1052.

BERRY, S. M. and BERRY, D. A. (2004). Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model. *Biometrics* **60** 418–426. MR2066276

CATTANEO, M. D., FRANDSEN, B. and TITIUNIK, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. Senate. *Journal of Causal Inference* **3** 1–24.

CELLINI, S. R., FERREIRA, F. and ROTHSTEIN, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *Q. J. Econ.* **125** 215–261.

CHIB, S. and GREENBERG, E. (2014). Nonparametric Bayes analysis of the sharp and fuzzy regression discontinuity designs. Technical report, Washington Univ. St Louis, Olin School of Business.

CHIB, S. and JACOBI, L. (2015). Bayesian fuzzy regression discontinuity analysis and returns to compulsory schooling. *J. Appl. Econometrics*. Published online in Wiley Online Library (wileyonlinelibrary.com), DOI:10.1002/jae.2481.

COOK, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *J. Econometrics* **142** 636–654. MR2416822

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* **7** 1–68.

DINARDO, J. and LEE, D. S. (2011). Program evaluation and research designs. In *Handbook of Labor Economics* **4A** 463–536. Elsevier, Philadelphia, PA.

ELLIOTT, M. R., RAGHUNATHAN, T. E. and LI, Y. (2010). Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics* **11** 353–372.

FRANGAKIS, C. E. and RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58** 21–29. MR1891039

FRUMENTO, P., MEALLI, F., PACINI, B. and RUBIN, D. B. (2012). Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. *J. Amer. Statist. Assoc.* **107** 450–466. MR2980057

GARIBALDI, P., GIAVAZZI, F., ICHINO, A. and RETTORE, E. (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. *Rev. Econ. Stat.* **94** 699–711.

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. MR1422404

GELMAN, A. E. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GOSSELIN, F. (2011). A new calibrated Bayesian internal goodness-of-fit method: Sampled posterior p-values as simple and general p-values that allow double use of the data. *PLoS ONE* **6** 1–10.

HAHN, J., TODD, P. E. and VAN DER KLAAUW, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* **69** 201–209.

IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics* **86** 4–29.

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–476.

IMBENS, G. and KALYANARAMAN, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Rev. Econ. Stud.* **79** 933–959. MR2986387

IMBENS, G. W. and LEMIEUX, T. (2008). Regression discontinuity designs: A guide to practice. *J. Econometrics* **142** 615–635. MR2416821

IMBENS, G. W. and RUBIN, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann. Statist.* **25** 305–327. MR1429927

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951

IMBENS, G. W. and ZAJONC, T. (2011). Regression discontinuity design with multiple forcing variables. Technical report, Harvard Univ., Dept. Economics.

JOHNSON, V. E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Anal.* **2** 719–733. MR2361972

LEE, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *J. Econometrics* **142** 675–697. MR2416824

LEE, D. S. and LEMIEUX, T. (2010). Regression discontinuity designs in economics. *J. Econ. Lit.* **485** 281–355.

LI, F., MATTEI, A. and MEALLI, F. (2015). Supplement to "Evaluating the causal effect of university grants on student dropout: Evidence from a regression discontinuity design using principal stratification." DOI:10.1214/15-AOAS881SUPP.

LUDWIG, J. and MILLER, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Q. J. Econ.* **122** 15981–208.

MATTEI, A., LI, F. and MEALLI, F. (2013). Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program. *Ann. Appl. Stat.* **7** 2336–2360. MR3161725

MEALLI, F. and PACINI, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *J. Amer. Statist. Assoc.* **108** 1120–1131. MR3174688

MEALLI, F. and RAMPICHINI, C. (2012). Evaluating the effects of university grants by using regression discontinuity designs. *J. Roy. Statist. Soc. Ser. A* **175** 775–798. MR2948374

MEALLI, F. and RUBIN, B. D. (2002). Assumptions when analyzing randomized experiments with noncompliance and missing outcomes. *Health Serv. Outcomes Res. Methodol.* **3** 225–232.

MERCATANTI, A. (2013). A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with noncompliance. *Aust. N. Z. J. Stat.* **55** 129–153. MR3079024

MERCATANTI, A., LI, F. and MEALLI, F. (2015). Improving inference of Gaussian mixtures using auxiliary variables. *Stat. Anal. Data Min.* **8** 34–48. MR3315979

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 331–366. MR1983752

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psycology* **66** 688–701.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

RUBIN, D. B. (1980). Discussion of "Randomization analysis of experimental data: The Fisher randomization test" by D. Basu. *J. Amer. Statist. Assoc.* **75** 591–593.

SALES, A. and HANSEN, B. (2014). Limitless regression discontinuity: Causal inference for a population surrounding a threshold. Available at arXiv:1403.5478.

SCHWARTZ, S. L., LI, F. and MEALLI, F. (2011). A Bayesian semiparametric approach to intermediate variables in causal inference. *J. Amer. Statist. Assoc.* **31** 949–962.

SCOTT, J. G. and BERGER, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inference* **136** 2144–2162. MR2235051

THISTLETHWAITE, D. and CAMPBELL, D. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *J. Educ. Psychol.* **51** 309–317.

VAN DER KLAAUW, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *Internat. Econom. Rev.* **43** 1249–1287.

VAN DER KLAAUW, W. (2008). Regression-discontinuity analysis: A survey of recent development in economics. *Labour* **22** 219–245.

ZAJONC, T. (2012). Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *J. Amer. Statist. Assoc.* **107** 80–92. MR2949343

F. LI
DEPARTMENT OF STATISTICAL SCIENCE
DUKE UNIVERSITY
122 OLD CHEMISTRY BUILDING
DURHAM, NORTH CAROLINA 27708-0251
USA
E-MAIL: fli@stat.duke.edu

A. MATTEI
F. MEALLI
DEPARTMENT OF STATISTICS,
INFORMATICS, APPLICATIONS
UNIVERSITY OF FLORENCE
VIALE MORGAGNI, 59
50134 FLORENCE
ITALY
E-MAIL: mattei@disia.unifi.it
mealli@disia.unifi.it

Submitted to the Annals of Applied Statistics

SUPPLEMENTARY MATERIAL FOR EVALUATING THE CAUSAL EFFECT OF UNIVERSITY GRANTS ON STUDENT DROPOUT: EVIDENCE FROM A REGRESSION DISCONTINUITY DESIGN USING PRINCIPAL STRATIFICATION

By Fan Li[*], Alessandra Mattei[†] and Fabrizia Mealli[†]

Duke University[*] and University of Florence[†]

Details of Calculation.

Bayesian Selection of the Subpopulations. We used a hierarchical Bayesian model for assessing the balance of the covariates between eligibility groups. The posterior distributions of the parameters are obtained from Markov chain Monte Carlo (MCMC) methods.

Let X_1, X_2, X_3, X_4, X_5, and X_6 denote the six observed covariates with X_1 = Sex; X_2 = High school type; X_3 = High school grade; X_4 = year of enrollment; X_5 = University; and X_6 = Major in university. Therefore X_1 and X_5 are binary variables, X_3 is a continuous variable, and X_2, X_4, and X_6 are categorical variables with 4, 3 and 6 categories, respectively.

The joint posterior distribution of the parameters is

$$
\begin{aligned}
&p(\sigma^2_{\gamma_0}, \mu_{\gamma_0}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \pi_{\gamma_1}, \sigma^2_3, \gamma_{01}, \boldsymbol{\gamma}_{02}, \gamma_{03}, \boldsymbol{\gamma}_{04}, \gamma_{05}, \boldsymbol{\gamma}_{06}, \gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}, \boldsymbol{\gamma}_{16} \mid \mathbf{X}, \mathbf{Z}) \propto \\
&\quad p(\sigma^2_{\gamma_0}, \mu_{\gamma_0}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \pi_{\gamma_1}) \times \\
&\quad p(\sigma^2_3, \gamma_{01}, \boldsymbol{\gamma}_{02}, \gamma_{03}, \boldsymbol{\gamma}_{04}, \gamma_{05}, \boldsymbol{\gamma}_{06}, \gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}, \boldsymbol{\gamma}_{16} \mid \sigma^2_{\gamma_0}, \mu_{\gamma_0}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \pi_{\gamma_1}) \times \\
&\quad \mathcal{L}(\sigma^2_3, \gamma_{01}, \boldsymbol{\gamma}_{02}, \gamma_{03}, \boldsymbol{\gamma}_{04}, \gamma_{05}, \boldsymbol{\gamma}_{06}, \gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}, \boldsymbol{\gamma}_{16}; \mathbf{X}, \mathbf{Z})
\end{aligned}
$$

where

$$
\begin{aligned}
&p(\sigma^2_{\gamma_0}, \mu_{\gamma_0}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \pi_{\gamma_1}) = \\
&\quad \frac{\underline{b}_{\gamma_0}^{\underline{a}_{\gamma_0}}}{\Gamma(\underline{a}_{\gamma_0})} (\sigma^2_{\gamma_0})^{-(\underline{a}_{\gamma_0}-1)} \exp\left\{-\frac{\underline{b}_{\gamma_0}}{\sigma^2_{\gamma_0}}\right\} \times \frac{1}{\sqrt{2\pi\underline{\sigma}^2_{\gamma_0}}} \exp\left\{-\frac{1}{2\underline{\sigma}^2_{\gamma_0}}\left(\mu_{\gamma_0} - \underline{\mu}_{\gamma_0}\right)^2\right\} \times \\
&\quad \frac{\underline{b}_{\gamma_1}^{\underline{a}_{\gamma_1}}}{\Gamma(\underline{a}_{\gamma_1})} (\sigma^2_{\gamma_1})^{-(\underline{a}_{\gamma_1}-1)} \exp\left\{-\frac{\underline{b}_{\gamma_1}}{\sigma^2_{\gamma_1}}\right\} \times \frac{1}{\sqrt{2\pi\underline{\sigma}^2_{\gamma_1}}} \exp\left\{-\frac{1}{2\underline{\sigma}^2_{\gamma_1}}\left(\mu_{\gamma_1} - \underline{\mu}_{\gamma_1}\right)^2\right\} \times \\
&\quad \frac{\Gamma(\underline{a}_\pi + \underline{b}_\pi)}{\Gamma(\underline{a}_\pi)\Gamma(\underline{b}_\pi)} \pi_{\gamma_1}^{\underline{a}_\pi - 1} \left(1 - \pi_{\gamma_1}\right)^{\underline{b}_\pi - 1}
\end{aligned}
$$

$$
\begin{aligned}
&p(\sigma_3^2, \gamma_{01}, \boldsymbol{\gamma}_{02}, \gamma_{03}, \boldsymbol{\gamma}_{04}, \gamma_{05}, \boldsymbol{\gamma}_{06}, \gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}, \boldsymbol{\gamma}_{16} \mid \sigma_{\gamma_0}^2, \mu_{\gamma_0}, \sigma_{\gamma_1}^2, \mu_{\gamma_1}, \pi_{\gamma_1}) = \\
&\quad \frac{b^a}{\Gamma(a)} (\sigma_3^2)^{-(a-1)} \exp\left\{-\frac{b}{\sigma_3^2}\right\} \times \prod_{j=1,3,5} \frac{1}{\sqrt{2\pi\sigma_{\gamma_0}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_0}^2}\left(\gamma_{0j} - \mu_{\gamma_0}\right)^2\right\} \times \\
&\quad \prod_{k=1}^{3} \frac{1}{\sqrt{2\pi\sigma_{\gamma_0}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_0}^2}\left(\gamma_{02}^{(k)} - \mu_{\gamma_0}\right)^2\right\} \times \prod_{k=1}^{2} \frac{1}{\sqrt{2\pi\sigma_{\gamma_0}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_0}^2}\left(\gamma_{04}^{(k)} - \mu_{\gamma_0}\right)^2\right\} \times \\
&\quad \prod_{k=1}^{5} \frac{1}{\sqrt{2\pi\sigma_{\gamma_0}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_0}^2}\left(\gamma_{06}^{(k)} - \mu_{\gamma_0}\right)^2\right\} \times \\
&\quad \prod_{j=1,3,5} \left[\pi_{\gamma_1} \mathbf{1}(\{\gamma_{1j} = 1\}) + (1 - \pi_{\gamma_1}) \frac{1}{\sqrt{2\pi\sigma_{\gamma_1}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_1}^2}\left(\gamma_{1j} - \mu_{\gamma_1}\right)^2\right\}\right] \times \\
&\quad \prod_{k=1}^{3} \left[\pi_{\gamma_1} \mathbf{1}(\{\gamma_{12}^{(k)} = 1\}) + (1 - \pi_{\gamma_1}) \frac{1}{\sqrt{2\pi\sigma_{\gamma_1}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_1}^2}\left(\gamma_{12}^{(k)} - \mu_{\gamma_1}\right)^2\right\}\right] \times \\
&\quad \prod_{k=1}^{2} \left[\pi_{\gamma_1} \mathbf{1}(\{\gamma_{14}^{(k)} = 1\}) + (1 - \pi_{\gamma_1}) \frac{1}{\sqrt{2\pi\sigma_{\gamma_1}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_1}^2}\left(\gamma_{14}^{(k)} - \mu_{\gamma_1}\right)^2\right\}\right] \times \\
&\quad \prod_{k=1}^{5} \left[\pi_{\gamma_1} \mathbf{1}(\{\gamma_{16}^{(k)} = 1\}) + (1 - \pi_{\gamma_1}) \frac{1}{\sqrt{2\pi\sigma_{\gamma_1}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_1}^2}\left(\gamma_{16}^{(k)} - \mu_{\gamma_1}\right)^2\right\}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathcal{L}(\sigma_3^2, \gamma_{01}, \boldsymbol{\gamma}_{02}, \gamma_{03}, \boldsymbol{\gamma}_{04}, \gamma_{05}, \boldsymbol{\gamma}_{06}, \gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}, \boldsymbol{\gamma}_{16}; \mathbf{X}, \mathbf{Z}) = \\
&\quad \prod_i \prod_{j=1,5} \Phi\left((-1)^{\mathbf{1}\{X_{ij}=0\}}(\gamma_{0j} + \gamma_{1j} Z_i)\right) \times \frac{1}{\sqrt{2\pi\sigma_3^2}} \exp\left\{\frac{1}{\sigma_3^2}\left(X_{i3} - \gamma_{03} - \gamma_{13} Z_i\right)^2\right\} \times \\
&\quad \left[1 - \Phi\left(\gamma_{02}^{(1)} + \gamma_{12}^{(1)} Z_i\right)\right]^{\mathbf{1}\{X_{i2}=1\}} \left[\Phi\left(\gamma_{02}^{(1)} + \gamma_{12}^{(1)} Z_i\right)\left(1 - \Phi\left(\gamma_{02}^{(2)} + \gamma_{12}^{(2)} Z_i\right)\right)\right]^{\mathbf{1}\{X_{i2}=2\}} \\
&\quad \left[\Phi\left(\gamma_{02}^{(1)} + \gamma_{12}^{(1)} Z_i\right) \Phi\left(\gamma_{02}^{(2)} + \gamma_{12}^{(2)} Z_i\right)\left(1 - \Phi\left(\gamma_{02}^{(3)} + \gamma_{12}^{(3)} Z_i\right)\right)\right]^{\mathbf{1}\{X_{i2}=3\}} \\
&\quad \left[\Phi\left(\gamma_{02}^{(1)} + \gamma_{12}^{(1)} Z_i\right) \Phi\left(\gamma_{02}^{(2)} + \gamma_{12}^{(2)} Z_i\right) \Phi\left(\gamma_{02}^{(3)} + \gamma_{12}^{(3)} Z_i\right)\right]^{\mathbf{1}\{X_{i2}=4\}} \times
\end{aligned}
$$

$$\left[1-\Phi\left(\gamma_{04}^{(1)}+\gamma_{14}^{(1)}Z_i\right)\right]^{\mathbf{1}\{X_{i4}=1\}}\left[\Phi\left(\gamma_{04}^{(1)}+\gamma_{14}^{(1)}Z_i\right)\left(1-\Phi\left(\gamma_{04}^{(2)}+\gamma_{14}^{(2)}Z_i\right)\right)\right]^{\mathbf{1}\{X_{i4}=2\}}$$
$$\left[\Phi\left(\gamma_{04}^{(1)}+\gamma_{14}^{(1)}Z_i\right)\Phi\left(\gamma_{04}^{(2)}+\gamma_{14}^{(2)}Z_i\right)\right]^{\mathbf{1}\{X_{i4}=3\}}\times$$
$$\left[1-\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\right]^{\mathbf{1}\{X_{i6}=1\}}\left[\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\left(1-\Phi\left(\gamma_{06}^{(2)}+\gamma_{16}^{(2)}Z_i\right)\right)\right]^{\mathbf{1}\{X_{i6}=2\}}$$
$$\left[\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\Phi\left(\gamma_{06}^{(2)}+\gamma_{16}^{(2)}Z_i\right)\left(1-\Phi\left(\gamma_{06}^{(3)}+\gamma_{16}^{(3)}Z_i\right)\right)\right]^{\mathbf{1}\{X_{i6}=3\}}$$
$$\left[\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\Phi\left(\gamma_{06}^{(2)}+\gamma_{16}^{(2)}Z_i\right)\Phi\left(\gamma_{06}^{(3)}+\gamma_{16}^{(3)}Z_i\right)\left(1-\Phi\left(\gamma_{06}^{(4)}+\gamma_{16}^{(4)}Z_i\right)\right)\right]^{\mathbf{1}\{X_{i6}=4\}}$$
$$\left[\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\Phi\left(\gamma_{06}^{(2)}+\gamma_{16}^{(2)}Z_i\right)\Phi\left(\gamma_{06}^{(3)}+\gamma_{16}^{(3)}Z_i\right)\Phi\left(\gamma_{06}^{(4)}+\gamma_{16}^{(4)}Z_i\right)\right.$$
$$\left.\left(1-\Phi\left(\gamma_{06}^{(5)}+\gamma_{16}^{(5)}Z_i\right)\right)\right]^{\mathbf{1}\{X_{i6}=5\}}\left[\Phi\left(\gamma_{06}^{(1)}+\gamma_{16}^{(1)}Z_i\right)\Phi\left(\gamma_{06}^{(2)}+\gamma_{16}^{(2)}Z_i\right)\right.$$
$$\left.\Phi\left(\gamma_{06}^{(3)}+\gamma_{16}^{(3)}Z_i\right)\Phi\left(\gamma_{06}^{(4)}+\gamma_{16}^{(4)}Z_i\right)\Phi\left(\gamma_{06}^{(5)}+\gamma_{16}^{(5)}Z_i\right)\right]^{\mathbf{1}\{X_{i6}=6\}}$$

The full conditional distributions for the hyper-parameters are

$$\begin{aligned}\sigma^2_{\gamma_0} &\mid \mu_{\gamma_0},\gamma_{01},\boldsymbol{\gamma}_{02},\gamma_{03},\boldsymbol{\gamma}_{04},\gamma_{05},\boldsymbol{\gamma}_{06}\sim IG\left(\overline{a}_{\gamma_0},\overline{b}_{\gamma_0}\right)\\ \mu_{\gamma_0} &\mid \sigma^2_{\gamma_0},\gamma_{01},\boldsymbol{\gamma}_{02},\gamma_{03},\boldsymbol{\gamma}_{04},\gamma_{05},\boldsymbol{\gamma}_{06}\sim N(\overline{\mu}_{\gamma_0},\overline{\sigma}^2_{\gamma_0})\end{aligned}$$

where

$$\begin{aligned}\overline{a}_{\gamma_0} &= \underline{a}_{\gamma_0}+\frac{N_{\gamma_0}}{2}\\ \overline{b}_{\gamma_0} &= \underline{b}_{\gamma_0}+\frac{1}{2}\left[\sum_{j=1,3,5}(\gamma_{0j}-\mu_{\gamma_0})^2+\sum_{k=1}^{3}(\gamma_{02}^{(k)}-\mu_{\gamma_0})^2+\sum_{k=1}^{2}(\gamma_{04}^{(k)}-\mu_{\gamma_0})^2+\sum_{k=1}^{5}(\gamma_{06}^{(k)}-\mu_{\gamma_0})^2\right]\\ \overline{\sigma}^2_{\gamma_0} &= \left(\frac{1}{\underline{\sigma}^2_{\gamma_0}}+\frac{N_{\gamma_0}}{\sigma^2_{\gamma_0}}\right)^{-1}\\ \overline{\mu}_{\gamma_0} &= \overline{\sigma}^2_{\gamma_0}\left[\frac{\underline{\mu}_{\gamma_0}}{\underline{\sigma}^2_{\gamma_0}}+\frac{1}{\sigma^2_{\gamma_0}}\left(\sum_{j=1,3,5}\gamma_{0j}+\sum_{k=1}^{3}\gamma_{02}^{(k)}+\sum_{k=1}^{2}\gamma_{04}^{(k)}+\sum_{k=1}^{5}\gamma_{06}^{(k)}\right)\right];\end{aligned}$$

$$\begin{aligned}\sigma^2_{\gamma_1} &\mid \mu_{\gamma_1},\gamma_{11},\boldsymbol{\gamma}_{12},\gamma_{13},\boldsymbol{\gamma}_{14},\gamma_{15},\boldsymbol{\gamma}_{16}\sim IG\left(\overline{a}_{\gamma_1},\overline{b}_{\gamma_1}\right)\\ \mu_{\gamma_1} &\mid \sigma^2_{\gamma_1},\gamma_{11},\boldsymbol{\gamma}_{12},\gamma_{13},\boldsymbol{\gamma}_{14},\gamma_{15},\boldsymbol{\gamma}_{16}\sim N(\overline{\mu}_{\gamma_1},\overline{\sigma}^2_{\gamma_1}),\end{aligned}$$

where

$$\begin{aligned}
\overline{a}_{\gamma_1} &= \underline{a}_{\gamma_1} + \frac{N_{\gamma_1}}{2} \\
\overline{b}_{\gamma_1} &= \underline{b}_{\gamma_1} + \frac{1}{2}\left[\sum_{j=1,3,5}(\gamma_{1j}-\mu_{\gamma_1})^2 + \sum_{k=1}^{3}(\gamma_{12}^{(k)}-\mu_{\gamma_1})^2 + \sum_{k=1}^{2}(\gamma_{14}^{(k)}-\mu_{\gamma_1})^2 + \sum_{k=1}^{5}(\gamma_{16}^{(k)}-\mu_{\gamma_1})^2\right] \\
\overline{\sigma}^2_{\gamma_1} &= \left(\frac{1}{\underline{\sigma}^2_{\gamma_1}} + \frac{N_{\gamma_1}}{\sigma^2_{\gamma_1}}\right)^{-1} \\
\overline{\mu}_{\gamma_1} &= \overline{\sigma}^2_{\gamma_1}\left[\frac{\underline{\mu}_{\gamma_1}}{\underline{\sigma}^2_{\gamma_1}} + \frac{1}{\sigma^2_{\gamma_1}}\left(\sum_{j=1,3,5}\gamma_{1j} + \sum_{k=1}^{3}\gamma_{12}^{(k)} + \sum_{k=1}^{2}\gamma_{14}^{(k)} + \sum_{k=1}^{5}\gamma_{16}^{(k)}\right)\right];
\end{aligned}$$

and

$$\pi_{\gamma_1} \mid \gamma_{11}, \boldsymbol{\gamma}_{12}, \gamma_{13}, \boldsymbol{\gamma}_{14}, \gamma_{15}, \boldsymbol{\gamma}_{16} \sim Beta(\overline{a}_\pi, \overline{b}_\pi),$$

where

$$\begin{aligned}
\overline{a}_\pi &= \underline{a}_\pi + \left(\sum_{j=1,3,5}\mathbf{1}\{\gamma_{1j}=0\} + \sum_{k=1}^{3}\mathbf{1}\{\gamma_{12}^{(k)}=0\} + \sum_{k=1}^{2}\mathbf{1}\{\gamma_{14}^{(k)}=0\} + \sum_{k=1}^{5}\mathbf{1}\{\gamma_{16}^{(k)}=0\}\right) \\
\overline{b}_\pi &= \underline{b}_\pi + N_{\gamma_1} - \left(\sum_{j=1,3,5}\mathbf{1}\{\gamma_{1j}=0\} + \sum_{k=1}^{3}\mathbf{1}\{\gamma_{12}^{(k)}=0\} + \sum_{k=1}^{2}\mathbf{1}\{\gamma_{14}^{(k)}=0\} + \sum_{k=1}^{5}\mathbf{1}\{\gamma_{16}^{(k)}=0\}\right)
\end{aligned}$$

with $N_{\gamma_0} = 13$ and $N_{\gamma_1} = 13$ are the number of $\gamma_0's$ and $\gamma_1's$.

The full conditional distributions for the parameters are as follows:
Parameters of the distribution of the continuous variable, X_3 (High school grade):

$$\sigma^2_3 \mid \gamma_{03}, \gamma_{13}, \mathbf{X}, \mathbf{Z} \sim IG\left(\overline{a}, \overline{b}\right) \qquad \gamma_{03} \mid \sigma^2_3, \gamma_{13}, \sigma^2_{\gamma_0}, \mu_{\gamma_0}, \mathbf{X}, \mathbf{Z} \sim N\left(\overline{\mu}_{\gamma_{03}}, \overline{\sigma}^2_{\gamma_{03}}\right)$$

where

$$\overline{a} = \underline{a} + \frac{N}{2} \qquad \overline{b} = \underline{b} + \frac{1}{2}\sum_i (X_{i3} - \gamma_{03} - \gamma_{13}Z_i)^2$$

$$\overline{\sigma}^2_{\gamma_{03}} = \left(\frac{1}{\sigma^2_{\gamma_0}} + \frac{N}{\sigma^2_3}\right)^{-1} \qquad \overline{\mu}_{\gamma_{03}} = \overline{\sigma}^2_{\gamma_{03}}\left[\frac{\mu_{\gamma_0}}{\sigma^2_{\gamma_0}} + \frac{\sum_i (X_{i3} - \gamma_{13}Z_i)}{\sigma^2_3}\right]$$

and

$$p\left(\gamma_{13} \mid \sigma^2_3, \gamma_{03}, \pi_{\gamma_1}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \mathbf{X}, \mathbf{Z}\right) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2_3}}\right)^N \exp\left\{-\frac{1}{2\sigma^2_3}\sum_i (X_{i3} - \gamma_{03} - \gamma_{13}Z_i)^2\right\} \times$$
$$\left[\pi_{\gamma_1}\mathbf{1}(\{\gamma_{13}=0\}) + (1-\pi_{\gamma_1})\frac{1}{\sqrt{2\pi\sigma^2_{\gamma_1}}}\exp\left\{-\frac{1}{2\sigma^2_{\gamma_1}}\left(\gamma_{13}-\mu_{\gamma_1}\right)^2\right\}\right]$$

Parameters of the distribution of the binary variables, X_1 (Gender) and X_5 (University): Let $X^*_{ij} \sim N\left(\gamma_{0j} + \gamma_{1j}Z_i, 1\right)$, $j = 1, 5$ a latent variable such that $Pr(X_{ij} = 1) = Pr(X^*_{ij} > 0)$. The distribution of $X^*_{ij} \mid \gamma_{0j}, \gamma_{1j}, \mathbf{X}, \mathbf{Z}$ is $N(\gamma_{0j} + \gamma_{1j}Z_i, 1)$ truncated to the left of zero if $X_{ij} = 1$ and truncated to the right of zero if $X_{ij} = 0$. We have

$$\gamma_{0j} \mid \gamma_{1j}, \sigma^2_{\gamma_0}, \mu_{\gamma_0}, \mathbf{X}, \mathbf{Z}, \mathbf{X}^*_j \sim N\left(\overline{\mu}_{\gamma_{0j}}, \overline{\sigma}^2_{\gamma_{0j}}\right)$$

where

$$\overline{\sigma}^2_{\gamma_{0j}} = \left(\frac{1}{\sigma^2_{\gamma_0}} + N\right)^{-1} \qquad \overline{\mu}_{\gamma_{0j}} = \overline{\sigma}^2_{\gamma_{0j}}\left[\frac{\mu_{\gamma_0}}{\sigma^2_{\gamma_0}} + \sum_i (X^*_{ij} - \gamma_{1j}Z_i)\right]$$

and

$$p\left(\gamma_{1j} \mid \gamma_{0j}, \pi_{\gamma_1}, \sigma^2_{\gamma_1}, \mu_{\gamma_1}, \mathbf{X}, \mathbf{Z}\right) \propto \Phi\left((-1)^{\mathbf{1}\{X_{ij}=0\}}(\gamma_{0j} + \gamma_{1j}Z_i)\right) \times \left[\pi_{\gamma_1}\mathbf{1}(\{\gamma_{1j} = 0\}) + (1 - \pi_{\gamma_1})\frac{1}{\sqrt{2\pi\sigma^2_{\gamma_1}}}\exp\left\{-\frac{1}{2\sigma^2_{\gamma_1}}\left(\gamma_{1j} - \mu_{\gamma_1}\right)^2\right\}\right]$$

Parameters of the distribution of the categorical variables, X_2 (High school type) and X_4 (Enrollment year), X_5 (Major in university): Let K be the number of values/levels for X_j and let $X^{*(k)}_{ij} \sim N\left(\gamma_{0j} + \gamma_{1j}Z_i, 1\right)$, $K = 1, \ldots, K-1$ be independent latent variables such that $Pr(X_{ij} = 1) = Pr\left(X^{*(1)}_{ij} \leq 0\right)$, and $Pr(X_{ij} = k) = Pr\left(\cap_{\ell=1}^{k-1}\{X^{*(\ell)}_{ij} > 0\} \cap X^{*(k)}_{ij} \leq 0\right)$ for $k = 2, \ldots, K-1$. The distribution of $X^{*(k)}_{ij}$ given $\boldsymbol{\gamma}_{0j}, \boldsymbol{\gamma}_{1j}$ and the data, $\mathbf{X}, \mathbf{Z}$, is $N\left(\gamma^{(k)}_{0j} + \gamma^{(k)}_{1j}Z_i, 1\right)$ truncated either to the left or to the right of zero depending on X_{ij}. We have

$$\gamma^{(k)}_{0j} \mid \gamma_{1j}, \sigma^2_{\gamma_0}, \mu_{\gamma_0}, \mathbf{X}, \mathbf{Z}, \mathbf{X}^*_j \sim N\left(\overline{\mu}^{(k)}_{\gamma_{0j}}, \overline{\sigma}^{2(k)}_{\gamma_{0j}}\right)$$

where $\mathbf{X}^*_j = \left(X^{*(1)}_j, \ldots, X^{*(K-1)}_j\right)$ and

$$\overline{\sigma}^{2(1)}_{\gamma_{0j}} = \left(\frac{1}{\sigma^2_{\gamma_0}} + N\right)^{-1} \qquad \overline{\mu}^{(1)}_{\gamma_{0j}} = \overline{\sigma}^2_{\gamma_{0j}}\left[\frac{\mu_{\gamma_0}}{\sigma^2_{\gamma_0}} + \textstyle\sum_i (X^{*(k)}_{ij} - \gamma_{1j}Z_i)\right]$$

$$\overline{\sigma}^{2(k)}_{\gamma_{0j}} = \left(\frac{1}{\sigma^2_{\gamma_0}} + \textstyle\sum_i \mathbf{1}(X_{ij} \notin \{1, \ldots, k-1\})\right)^{-1}$$

$$\overline{\mu}^{(k)}_{\gamma_{0j}} = \overline{\sigma}^2_{\gamma_{0j}}\left[\frac{\mu_{\gamma_0}}{\sigma^2_{\gamma_0}} + \textstyle\sum_{i:X_{ij}\notin\{1,\ldots,k-1\}} (X^{*(k)}_{ij} - \gamma_{1j}Z_i)\right]$$

and

$$p\left(\gamma_{1j}^{(k)} \mid \boldsymbol{\gamma}_{0j}, \pi_{\gamma_1}, \sigma_{\gamma_1}^2, \mu_{\gamma_1}, \mathbf{X}, \mathbf{Z}\right) \propto \left[1 - \Phi\left(\gamma_{0j}^{(k)} + \gamma_{1j}^{(k)} Z_i\right)\right]^{\mathbf{1}\{X_{ij}=k\}} \prod_{\ell=k+1,\ldots,K} \left[\Phi\left(\gamma_{0j}^{(k)} + \gamma_{1j}^{(k)} Z_i\right)^{\mathbf{1}\{X_{ij}=\ell\}}\right] \times \left[\pi_{\gamma_1} \mathbf{1}(\{\gamma_{1j}^{(k)} = 0\}) + (1 - \pi_{\gamma_1}) \frac{1}{\sqrt{2\pi\sigma_{\gamma_1}^2}} \exp\left\{-\frac{1}{2\sigma_{\gamma_1}^2}\left(\gamma_{1j}^{(k)} - \mu_{\gamma_1}\right)^2\right\}\right]$$

The MCMC algorithm we use to sample from the posterior distributions of the parameters simulates successively from the above full conditional distributions. The data augmentation method is used to impute at each step the latent variables underlying binary and categorical covariates. Simulations from the mixture distributions for the values of $\gamma_{11}, \boldsymbol{\gamma}_{12}, \boldsymbol{\gamma}_{14}, \gamma_{13}, \gamma_{15}$, and $\boldsymbol{\gamma}_{16}$ use a mixture Metropolis-Hastings step (Berry and Berry, 2004). We draw from a point mass at zero with probability 0.5 and a Normal distribution centered on the current value of the parameter. Let γ_{1j}^{cand} be the candidate value and $\gamma_{1j}^{(t)}$ the current value at iteration t. For simplicity we omit superscript (k) for $j = 2, 4, 6$. Define

$$r_j = \frac{p\left(\gamma_{1j}^{cand} \mid \boldsymbol{\gamma}_{0j}, \pi_{\gamma_1}, \sigma_{\gamma_1}^2, \mu_{\gamma_1}, \mathbf{X}, \mathbf{Z}\right)}{p\left(\gamma_{1j}^{(t)} \mid \boldsymbol{\gamma}_{0j}, \pi_{\gamma_1}, \sigma_{\gamma_1}^2, \mu_{\gamma_1}, \mathbf{X}, \mathbf{Z}\right)},$$

where $\boldsymbol{\gamma}_{0j} = \gamma_{0j}$ for $j = 1, 3, 5$. The candidate draw is accepted with probability

$$r = \begin{cases} r_j & \text{if } \gamma_{1j}^{cand} = \gamma_{1j}^{(t)} = 0 \\ \dfrac{1}{\sqrt{2\pi\sigma_{MH}^2}} \exp\left\{-\dfrac{(\gamma_{1j}^{(t)})^2}{2\sigma_{MH}^2}\right\} r_j & \text{if } \gamma_{1j}^{cand} = 0 \text{ and } \gamma_{1j}^{(t)} \neq 0 \\ \dfrac{r_j}{\dfrac{1}{\sqrt{2\pi\sigma_{MH}^2}} \exp\left\{-\dfrac{(\gamma_{1j}^{(cand)})^2}{2\sigma_{MH}^2}\right\}} & \text{if } \gamma_{1j}^{cand} \neq 0 \text{ and } \gamma_{1j}^{(t)} = 0 \\ r_j & \text{if } \gamma_{1j}^{cand} \neq 0 \text{ and } \gamma_{1j}^{(t)} \neq 0 \end{cases}$$

where σ_{MH}^2 is the variance of the Normal proposal distribution.

Markov chain Monte Carlo Algorithm. The posterior distributions of the parameters are obtained from Markov chain Monte Carlo (MCMC) methods. The Markov chain algorithm that we adopt is based on Gibbs sampler and uses the data augmentation method to impute at each step the missing principal stratum indicators and to exploit the complete principal strata-data posterior distribution to update the parameters.

Let $(\mathbf{G}^{(t)}, \boldsymbol{\theta}^{(t)})$, where $\boldsymbol{\theta}^{(t)}=\Big(\boldsymbol{\alpha}_{AA}^{(t)}, \boldsymbol{\alpha}_{NA}^{(t)}, \beta_{0,AA,0}^{(t)}, \beta_{AA,0}^{(S,t)}, \beta_{0,AA,1}^{(t)}, \beta_{AA,1}^{(S,t)}, \beta_{0,CA,0}^{(t)}, \beta_{CA,0}^{(S,t)}, \beta_{0,CA,1}^{(t)}, \beta_{CA,1}^{(S,t)}, \beta_{0,NA}^{(t)}, \beta_{NA}^{(S,t)}, \boldsymbol{\beta}^{(X,t)}\Big)$, denote the state of the chain at time t. The state of the chain at time $t+1$ follows from applying the following steps, where $Z_i = \mathbf{1}\{S_i \leq s_0\}$.

1. Sample $\mathbf{G}^{(t+1)}$ according to $p(G|\boldsymbol{\theta}^{(t)}, \mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{X}, \mathbf{S}; \mathcal{U}_{s_0})$:

$$Pr\Big(G_i = AA|\boldsymbol{\theta}^{(t)}, Y_i^{obs}, A_i^{obs} = 1, \mathbf{X}_i, S_i, S_i > s_0; i \in \mathcal{U}_{s_0}\Big) = 1$$
$$Pr\Big(G_i = NA|\boldsymbol{\theta}^{(t)}, Y_i^{obs}, A_i^{obs} = 0, \mathbf{X}_i, S_i, S_i \leq s_0; i \in \mathcal{U}_{s_0}\Big) = 1$$

For observations $i \in \mathcal{U}_{s_0}$ with $S_i > s_0$ and $A_i^{obs} = 0$

$$\begin{aligned} Pr\Big(G_i = CA|Y_i^{obs}, A_i^{obs} = 0, \mathbf{X}_i, S_i, S_i > s_0; i \in \mathcal{U}_{s_0}\Big) \propto \\ Pr(G_i = CA) \cdot \Phi\Big((-1)^{1\{Y_i=0\}}(\beta_{0,CA,0} + \beta_{CA,0}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big) \times \\ \Big[Pr(G_i = CA) \cdot \Phi\Big((-1)^{1\{Y_i=0\}}(\beta_{0,CA,0} + \beta_{CA,0}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big) + \\ Pr(G_i = NA) \cdot \Phi\Big((-1)^{1\{Y_i=0\}}(\beta_{0,NA} + \beta_{NA}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big)\Big]^{(-1)} \end{aligned}$$

For observations $i \in \mathcal{U}_{s_0}$ with $S_i \leq s_0$ and $A_i^{obs} = 1$

$$\begin{aligned} Pr\Big(G_i = CA|Y_i^{obs}, A_i^{obs} = 1, \mathbf{X}_i, S_i, S_i \leq s_0; i \in \mathcal{U}_{s_0}\Big) \propto \\ Pr(G_i = CA) \cdot \Phi\Big((-1)^{\mathbf{1}\{Y_i=0\}}(\beta_{0,CA,1} + \beta_{CA,1}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big) \times \\ \Big[Pr(G_i = CA) \cdot \Phi\Big((-1)^{\mathbf{1}\{Y_i=0\}}(\beta_{0,CA,1} + \beta_{CA,1}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big) + \\ Pr(G_i = AA) \cdot \Phi\Big((-1)^{\mathbf{1}\{Y_i=0\}}(\beta_{0,AA,1} + \beta_{AA,1}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\beta}^{(X)})\Big)\Big]^{(-1)} \end{aligned}$$

2. Sample the latent variables $G_i^*(AA)$ and $G_i^*(NA)$:
   (a) Sample the latent variable $G_i^*(AA)$ from $N(\alpha_{0,AA} + \alpha_{AA}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\alpha}_{AA}^{(X)}, 1)$ truncated to $(-\infty, 0)$ if $G_i = AA$ and to $(0, \infty)$ if $G_i \neq AA$.
   (b) Sample the latent variable $G_i^*(NA)$ from$N(\alpha_{0,NA} + \alpha_{NA}^{(S)} S_i^* + \mathbf{X}_i'\boldsymbol{\alpha}_{NA}^{(X)}, 1)$ truncated to $(-\infty, 0)$ if $G_i = NA$ and to $(0, \infty)$ if $G_i \neq NA$.
3. Sample the coefficients $\boldsymbol{\alpha}_{AA}$ and $\boldsymbol{\alpha}_{NA}$ given the following prior distributions:

$$\boldsymbol{\alpha}_{AA} \sim N\Big(\underline{\mu}_{\alpha_{AA}}, \underline{\sigma}_{\alpha_{AA}}^2 \mathbf{I}\Big) \qquad \boldsymbol{\alpha}_{NA} \sim N\Big(\underline{\mu}_{\alpha_{NA}}, \underline{\sigma}_{\alpha_{NA}}^2 \mathbf{I}\Big)$$

Let $\boldsymbol{SX} = [\mathbf{S}|\mathbf{X}]$ denote the $N \times (p+1)$ matrix with ith row equals to $(S_i, X_{i1}, \ldots, X_{ip})$.
   (a) Sample $\boldsymbol{\alpha}_{AA}$ from $N(\mu_{\alpha_{AA}}, \Sigma_{\alpha_{AA}})$ where

$$\mu_{\alpha_{AA}} = \Sigma_{\alpha_{AA}}\left(\frac{1}{\underline{\sigma}_{\alpha_{AA}}^2}\underline{\mu}_{\alpha_{AA}} + \boldsymbol{SX}'\boldsymbol{G}^*(AA)\right) \quad \text{and} \quad \Sigma_{\alpha_{AA}} = \left(\frac{1}{\underline{\sigma}_{\alpha_{AA}}^2} + \boldsymbol{SX}'\boldsymbol{SX}\right)^{-1}$$

(b) Let $\boldsymbol{SX}_{CA,NA}$ denote the sub-matrix of $\boldsymbol{SX}$ for units with $G_i^{(t+1)} = CA$ or $G_i^{(t+1)} = NA$ and let $\mathbf{G}^*_{CA,NA}(NA)$ be the sub-vector of $\mathbf{G}^*(NA)$ for units with $G_i^{(t+1)} = CA$ or $G_i^{(t+1)} = NA$. Sample α_{NA} from $N(\mu_{\alpha_{NA}}, \Sigma_{\alpha_{NA}})$ where

$$\mu_{\alpha_{NA}} = \Sigma_{\alpha_{NA}} \left(\frac{1}{\underline{\sigma}^2_{\alpha_{NA}}} \underline{\mu}_{\alpha_{NA}} + \boldsymbol{SX}'_{CA,NA} \boldsymbol{G}^*_{CA,NA}(NA) \right)$$

and

$$\Sigma_{\alpha_{NA}} = \left(\frac{1}{\underline{\sigma}^2_{\alpha_{NA}}} + \boldsymbol{SX}'_{CA,NA} \boldsymbol{SX}_{CA,NA} \right)^{-1}$$

4. For each $i \in \mathcal{U}_{s_0}$ with $G_i = g$ and $Z_i \equiv \mathbf{1}\{S_i \leq s_0\} = z$, sample the latent outcome $Y_i^*(z)$ from the normal distribution $N\left(\beta_{0,g,z} + \beta^{(S)}_{g,z} S_i + \mathbf{X}_i' \boldsymbol{\beta}^{(X)}, 1\right)$ truncated to $(0, \infty)$ if $Y_i^{obs} = 1$ and to $(-\infty, 0)$ if $Y_i^{obs} = 0$
5. Define $\mathbf{1} = (1, \ldots, 1)'$ and let $\mathbf{Y}^{obs}_{g,z}$ and $\mathbf{S}_{g,z}$ denote the sub-vectors of $\mathbf{Y}^{obs}$ and $\mathbf{S}$ for units with $G_i^{(t+1)} = g$ and $Z_i = z$. Also let $\mathbf{X}_{g,z}$ and $[\mathbf{1}|\mathbf{S}]_{g,z}$ denote the sub-matrices of $\mathbf{X}$ and $[\mathbf{1}|\mathbf{S}]$ for units with $G_i^{(t+1)} = g$ and $Z_i = z$.
6. Sample the coefficients $\boldsymbol{\beta}^{(X)} = (\beta_1, \ldots, \beta_p)'$ given their joint Normal prior distribution, $\boldsymbol{\beta}^{(X)} \sim N\left(\underline{\mu}_\beta, \underline{\sigma}^2_\beta \mathbf{I}\right)$, from the multivariate normal distribution $N(\mu_\beta, \Sigma_\beta)$, where

$$\mu_\beta = \Sigma_\beta \left(\frac{1}{\underline{\sigma}^2_\beta} \underline{\mu}_\beta + \sum_{g=AA,CA,NA} \sum_{z=0,1} \mathbf{X}'_{g,z} \left(\mathbf{Y}^{obs}_{g,z} - \mathbf{1}\beta_{0,g,z} - \beta^{(S)}_{g,z} \mathbf{S}^*_{g,z} \right) \right)$$

and

$$\Sigma_\beta = \left(\frac{1}{\underline{\sigma}^2_\beta} + \mathbf{X}'\mathbf{X} \right)^{-1}.$$

7. Sample the coefficients $\beta_{g,z} = (\beta_{0,g,z}, \beta^{(S)}_{g,z})'$ given their joint Normal prior distribution, $\boldsymbol{\beta}_{g,z} \sim N\left(\underline{\mu}_{\beta_{g,z}}, \underline{\sigma}^2_{\beta_{g,z}} \mathbf{I}\right)$, from the multivariate normal distribution $N(\mu_{\beta_{g,z}}, \Sigma_{\beta_{g,z}})$, where

$$\mu_{\beta_{g,z}} = \Sigma_{\beta_{g,z}} \left(\underline{\Sigma}^{-1}_{\beta_{g,z}} \underline{\mu}_{\beta_{g,z}} + [\mathbf{1}|\mathbf{S}]'_{g,z} (\mathbf{Y}^{obs}_{g,z} - \mathbf{X}_{g,z} \boldsymbol{\beta}^{(X)}) \right)$$

and

$$\Sigma_{\beta_{g,z}} = \left(\underline{\Sigma}^{-1}_{\beta_{g,z}} + [\mathbf{1}|\mathbf{S}]'_{g,z} [\mathbf{1}|\mathbf{S}]_{g,z} \right)^{-1}.$$

Posterior Predective Checks. A posterior predictive check generally consists of three steps: (*a*) generating replicated data from the posterior predictive distribution, (*b*) choosing and computing a discrepancy measure, and (*c*) computing a Bayesian p-value that summarizes the discrepancy. Here we provide some details on these steps.

Replicated Data. Given a set of draws θ^t ($t = 1, \ldots, T$) of the parameter vector $\boldsymbol{\theta}$ from its posterior distribution, for each t we first draw the compliance status $G_i^{obs,t}$ using a data augmentation step of the MCMC algorithm. Then we draw a simulated replicated data set, $(Y_i^{rep,t}, G_i^{rep,t}, A_i^{rep,t}, Z_i^{rep,t})$ from their respective posterior predictive distributions as follows:

1. Calculate the principal stratum probabilities

$$\begin{aligned}
\pi_{i,AA}^t &= 1 - \Phi\left(\alpha_{AA,0}^t + \alpha_{AA}^{(S,t)} S_i + \mathbf{X}_i' \alpha_{AA}^{(X,t)}\right) \\
\pi_{i,NA}^t &= \Phi\left(\alpha_{AA,0}^t + \alpha_{AA}^{(S,t)} S_i + \mathbf{X}_i' \alpha_{AA}^{(X,t)}\right) \cdot \left(1 - \Phi\left(\alpha_{NA,0}^t + \alpha_{NA}^{(S,t)} S_i + \mathbf{X}_i' \alpha_{NA}^{(X,t)}\right)\right) \\
\pi_{i,CA}^t &= 1 - \pi_{i,AA}^t - \pi_{i,NA}^t
\end{aligned}$$

and the outcome probabilities

$$\mu_{i,NA}^t = \Phi(\beta_{0,NA}^t + S_i \beta_{NA}^{(S,t)} + \mathbf{X}_i' \beta^{(X,t)})$$
$$\mu_{i,AA,z}^t = \Phi(\beta_{0,AA,z}^t + S_i \beta_{AA,z}^{(S,t)} + \mathbf{X}_i' \beta^{(X,t)}) \quad \text{and} \quad \mu_{i.CA,z}^t = \Phi(\beta_{0,CA,z}^t + S_i \beta_{CA,z}^{(S,t)} + \mathbf{X}_i' \beta^{(X,t)})$$

for $z = 0, 1$

2. We impose that $Z_i^{rep,t} = Z_i$ for each $t = 1, \ldots, T$ and $i = 1. \ldots, N$
3. Generate the compliance status $G_i^{rep,t}$, $i =, \ldots, N$
   (a) Generate N random numbers, $G_1^{draw}, \ldots, G_N^{draw}$, from a uniform distribution between 0 and 1;
   (b) Define

$$G_i^{rep,t} = \begin{cases} AA & \text{if } G_i^{draw} \leq \pi_{i,AA}^t\} \\ CA & \text{if } G_i^{draw} > \pi_{i,AA}^t \text{ and } G_i^{draw} \leq \pi_{i,AA}^t + \pi_{i,CA}^t \\ NA & \text{if } G_i^{draw} > \pi_{i,AA}^t + \pi_{i,CA}^t \end{cases}$$

4. Assign applicant status:

$$\begin{aligned}
A_i^{rep,t} = Z_i \cdot \left[1 \cdot \mathbf{I}\{G_i^{rep,t} = AA \text{ o } G_i^{rep,t} = CA\} + 0 \cdot \mathbf{I}\{G_i^{rep,t} = NA\}\right] + \\
(1 - Z_i) \cdot \left[1 \cdot \mathbf{I}\{G_i^{rep,t} = AA\} + 0 \cdot \mathbf{I}\{G_i^{rep,t} = CA \text{ o } G_i^{rep,t} = NA\}\right]
\end{aligned}$$

5. Simulate the outcome data $Y_i^{rep,t}$ as follows:
   (a) Generate N random numbers, $Y_1^{draw}, \ldots, Y_N^{draw}$, from a uniform distribution between 0 and 1:
   (b) Define

$$\begin{aligned}
Y_i^{rep,t} = \mathbf{I}\{Y_i^{draw} \leq \mu_{i,NA}^t\} \cdot \mathbf{I}\{G_i^{rep,t} = NA\} + \\
\left[\mathbf{I}\{Y_i^{draw} \leq \mu_{i,AA,0}^t\} \cdot (1 - Z_i) + \mathbf{I}\{Y_i^{draw} \leq \mu_{i,AA,1}^t\} \cdot Z_i\right] \cdot \mathbf{I}\{G_i^{rep,t} = AA\} + \\
\left[\mathbf{I}\{Y_i^{draw} \leq \mu_{i,CA,0}^t\} \cdot (1 - Z_i) + \mathbf{I}\{Y_i^{draw} \leq \mu_{i,CA,1}^t\} \cdot Z_i\right] \cdot \mathbf{I}\{G_i^{rep,t} = CA\}
\end{aligned}$$

Discrepancy Measures. Define $\mathcal{D}^{study}_{\mathcal{G},z} = \left\{i : G^{study}_i \in \mathcal{G}, Z^{study}_i = z\right\}$ where $\mathcal{G} \in \{\{AA\}, \{CA\}, \{AA, CA\}\}$, and $study = obs$ for the observed data and $study = rep$ for a replicated data set. Let $N^{study}_{\mathcal{G},z}$ be the number of units in the *study* data belonging to the $\mathcal{D}^{study}_{\mathcal{G},z}$ group, and $p^{study}_{\mathcal{G},z}$ be the proportion of units belonging to the $\mathcal{D}^{study}_{\mathcal{G},z}$ group with $Y^{study}_i = 1$: $p^{study}_{\mathcal{G},z} = \sum_{i \in \mathcal{D}^{study}_{\mathcal{G},z}} Y^{study}_i / N^{study}_{\mathcal{G},z}$.

Following Barnard et al. (2003), we use the following discrepancy measures:

$$D^{study}_{Signal,\mathcal{G}}(\theta) = \left| p^{study}_{\mathcal{G},1} - p^{study}_{\mathcal{G},0} \right| \text{ and } D^{study}_{Noise,\mathcal{G}}(\theta) = \sqrt{\frac{p^{study}_{\mathcal{G},0}(1 - p^{study}_{\mathcal{G},0})}{N^{study}_{\mathcal{G},0}} + \frac{p^{study}_{\mathcal{G},1}(1 - p^{study}_{\mathcal{G},1})}{N^{study}_{\mathcal{G},1}}},$$

and their ratio $D^{study}_{SNR,\mathcal{G}}(\theta) = \dfrac{D^{study}_{Signal,\mathcal{G}}(\theta)}{D^{study}_{Noise,\mathcal{G}}(\theta)}$.

Bayesian P-Value. Bayesian $p-$values are posterior probability statements on discrepancy between the observed data and replicated data conditional on the observed data and the posited model. Extreme (close to 0 or 1) $p-$values can be interpreted as evidence of lack-of-fit of the model in, at least some aspects of, the observed data. We consider the posterior predictive $p-$value (PPPV) (Gelman, Meng and Stern, 1996), which is the probability over the posterior predictive distribution of the principal strata membership and the parameters $\boldsymbol{\theta}$ that a discrepancy measure in a replicated data drawn with the same $\boldsymbol{\theta}$ as in the observed data, would be as or more extreme than the *realized* value of that discrepancy measure in the observed study. After generating the replicated data t ($t = 1, \ldots, T$), one can calculate $D^{obs,t}_{M,\mathcal{G}}(\theta^t)$ and $D^{rep}_{M,\mathcal{G}}(\theta^t)$ for $M = Signal, Noise, SNR$, and estimate the PPPV by the proportion of the T pairs $\left\{D^{obs,t}_{M,\mathcal{G}}(\theta^t), D^{rep}_{M,\mathcal{G}}(\theta^t)\right\}$, $t = 1, \ldots, T$, for which $D^{obs,t}_{M,\mathcal{G}}(\theta^t)$ exceeds $D^{rep}_{M,\mathcal{G}}(\theta^t)$. Formally

$$\hat{p}_M = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\left\{D^{obs}_{M,\mathcal{G}}(\theta^t) > D^{rep}_{M,\mathcal{G}}(\theta^t)\right\}.$$

Sensitivity Analyses. In order to assessing the robustness of the results with respect to violations of the local randomization assumption, we conduct further analyses deriving the posterior distributions of the causal estimands of interest under three additional model specifications: (1) a model where we specify the model for principal strata, $\pi_{i,g}$, and the conditional model for potential outcomes given principal strata, $f_{i,gz}$, conditioning on neither the forcing variable nor the pre-treatment variables (M_0- specification); (2) a model where we specify $\pi_{i,g}$ and $f_{i,gz}$ conditioning only on the forcing variable, without including the pre-treatment variables (M_1- specification); and (3) a model where we specify $\pi_{i,g}$

and $f_{i,gz}$ conditioning only on the pre-treatment variables, without including the forcing variable (M_X- specification).

Under local randomization, adjusting inference for either the forcing variable, S, or the pre-treatment variables, $\mathbf{X}$, should not be required, therefore we expect that results are similar across different model specifications.

Table A1 shows posterior medians and 95% posterior credible intervals for the principal strata proportions and the causal effects of interest for the subpopulation within 1 000 Euro around the threshold, under these three additional model specifications and the model specification we use in the paper, here named $M_{1X}-$*specification*.

As we can see in Table A1, results are robust across different model specifications. The posterior distributions for principal strata proportions and for average causal effects for always applicants, τ_{AA,s_0} and compliers, τ_{s_0}, are similar across different model specifications. The posterior medians of the average causal effect for compliant applicants, τ_{CA,s_0}, are more variable across model specifications, but their posterior distributions are always characterized by a large posterior variability: the posterior 95% credible intervals always cover zero and are evenly spread around zero with a large span. The robustness of the results across different model specifications suggests that causal inference under the local randomization assumption is credible and defensible.

In our study we prefer to specify the model for principal strata and the conditional model for potential outcomes given principal strata conditioning on both the pre-treatment variables, $\mathbf{X}$, and the forcing variable, S, for various reasons. First, as we highlight in the manuscript (see Section 4 and Section 5.2), we strongly recommend to adjust inference for the forcing variable S because systematic differences in the forcing variable S that, by definition, occur between eligible and ineligible units, may affect inference in the presence of units who do not belong to any subpopulation $\mathcal{U}_{s_0}$, and the risk that a selected sub-population includes units not belonging to any subpopulation $\mathcal{U}_{s_0}$ is not zero (see Table 2 in the main text). Second, adjusting for the pre-treatment variables $\mathbf{X}$ allows us to control for residual imbalance in the covariate distribution that may always occur in finite samples, even in completely randomized experiments. Finally, as it is well known in the causal inference literature, in general, and in the literature on principal stratification, in particular, adjusting for pre-treatment variables, even if they do not enter the assignment mechanism (as in the case of randomized experiments), may improve inference, because pre-treatment variables can improve prediction of both principal stratum membership and missing potential outcomes for the primary endpoint (dropout in our study).

References.

Barnard, J., Frangakis, C. F., Hill, J. L. and Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City (with discussion). *Journal of the American Statistical Association* **98** 299-323.

Berry, S. M. and Berry, D. A. (2004). Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model. *Biometrics* **60** 418-426.

TABLE A1

Posterior median and 95% credible intervals of principal strata proportion and super-population and finite-sample causal effects on dropout for always-applicants (τ_{AA,s_0}), compliant-applicants (τ_{CA,s_0}), and their union (τ_{s_0}), for the subpopulation within 1 000 *Euro around the threshold under model specifications M_0, M_1, M_X and M_{1X}*

	Population-average		Sample-average		Population-average at s_0	
h = 1 000	Median	95% CI	Median	95% CI	Median	95% CI
M_0−specification						
$\Pr(G_i = AA)$	.322	(.293; .354)	.322	(.299; .346)	–	–
$\Pr(G_i = CA)$	.059	(.015; .102)	.059	(.015; .100)	–	–
$\Pr(G_i = NA)$	.619	(.590; .647)	.619	(.599; .639)	–	–
τ_{AA,s_0}	−.151	(−.347; .002)	−.150	(−.341; −.008)	–	–
τ_{CA,s_0}	.009	(−.985; .996)	.000	(−.987; 1.000)	–	–
τ_{s_0}	−.120	(−.240; −.008)	−.120	(−.225; −.019)	–	–
M_1−specification						
$\Pr(G_i = AA)$	.294	(.253; .363)	.291	(.259; .354)	.295	(.255; .360)
$\Pr(G_i = CA)$	.097	(.005; .153)	.098	(.006; .150)	.087	(.003; .152)
$\Pr(G_i = NA)$	.613	(.560; .651)	.613	(.566; .644)	.619	(.566; .664)
τ_{AA,s_0}	−.297	(−.602; −.045)	−.296	(−.599; −.053)	−.351	(−.628; −.045)
τ_{CA,s_0}	.446	(−.882; .997)	.444	(−.886; 1.000)	.546	(−.972; 1.000)
τ_{s_0}	−.146	(−.369; .127)	−.148	(−.367; .124)	−.158	(−.391; .030)
M_X−specification						
$\Pr(G_i = AA)$	.333	(.312; .357)	.332	(.324; .352)	–	–
$\Pr(G_i = CA)$	.043	(.007; .057)	.026	(.006; .047)	–	–
$\Pr(G_i = NA)$	.625	(.604; .647)	.643	(.625; .645)	–	–
τ_{AA,s_0}	−.101	(−.152; −.051)	−.101	(−.138; −.064)	–	–
τ_{CA,s_0}	.000	(−.162; .958)	.000	(−.250; .957)	–	–
τ_{s_0}	−.086	(−.137; .019)	−.091	(−.132; −.025)	–	–
M_{1X}−specification						
$\Pr(G_i = AA)$	.336	(.312; .365)	.333	(.318; .354)	.335	(.311; .363)
$\Pr(G_i = CA)$	.043	(.002; .086)	.027	(.002; .075)	.043	(.001; .075)
$\Pr(G_i = NA)$	.623	(.584; .652)	.640	(.599; .645)	.625	(.594; .656)
τ_{AA,s_0}	−.161	(−.273; −.052)	−.161	(−.270; −.057)	−.154	(−.259; −.052)
τ_{CA,s_0}	.028	(−.745; .828)	.031	(−.778; .871)	.010	(−.918; .933)
τ_{s_0}	−.132	(−.242; −.021)	−.139	(−.247; −.034)	−.128	(−.229; −.020)

GELMAN, A. E., MENG, X. L. and STERN, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* **6** 733-807.

ALESSANDRA MATTEI AND FABRIZIA MEALLI
UNIVERSITY OF FLORENCE
DEPARTMENT OF STATISTICS, INFORMATICS, APPLICATIONS
VIALE MORGAGNI, 59
50134 FLORENCE, ITALY
E-MAIL: mattei@disia.unifi.it
mealli@disia.unifi.it

FAN LI
DUKE UNIVERSITY
DEPARTMENT OF STATISTICAL SCIENCE
122 OLD CHEMISTRY BUILDING
DURHAM, NORTH CAROLINA 27708-0251, U.S.A.
E-MAIL: fli@stat.duke.edu