3-29-2021
Working Paper Number 21003

Testing Identifying Assumptions in Bivariate Probit Models

Santiago Acerenza
Iowa State University, acerenza@iastate.edu

Otávio Bartalotti
Iowa State University, bartalot@iastate.edu

Desire Kedagni
Iowa State University, dkedagni@iastate.edu

Testing Identifying Assumptions in Bivariate Probit Models

Abstract

This paper focuses on the bivariate probit model's identifying assumptions: joint normality of errors, instrument exogeneity, and relevance conditions. First, we develop novel sharp testable equalities that can detect all possible observable violations of the assumptions. Second, we propose an easy-to-implement testing procedure for the model's validity based on feasible testable implications using existing inference methods for intersection bounds. The test achieves correct empirical size for moderately sized samples and performs well in detecting violations of the conditions in Monte Carlo simulations. Finally, we provide researchers with a road map on what to do when the bivariate probit model is rejected, including novel bounds for the average treatment effect that relax the normality assumption. Empirical examples illustrate the methodology's implementation.

Keywords

Exogeneity, bivariate probit, testable implications, moment inequalities, power, size

Disciplines

Econometrics | Statistical Models

TESTING IDENTIFYING ASSUMPTIONS IN BIVARIATE PROBIT MODELS

SANTIAGO ACERENZA, OTÁVIO BARTALOTTI AND DÉSIRÉ KÉDAGNI

Iowa State University - Department of Economics

Abstract. This paper focuses on the bivariate probit model's identifying assumptions: joint normality of errors, instrument exogeneity, and relevance conditions. First, we develop novel sharp testable equalities that can detect all possible observable violations of the assumptions. Second, we propose an easy-to-implement testing procedure for the model's validity based on feasible testable implications using existing inference methods for intersection bounds. The test achieves correct empirical size for moderately sized samples and performs well in detecting violations of the conditions in Monte Carlo simulations. Finally, we provide researchers with a road map on what to do when the bivariate probit model is rejected, including novel bounds for the average treatment effect that relax the normality assumption. Empirical examples illustrate the methodology's implementation.

Keywords: Exogeneity, bivariate probit, testable implications, moment inequalities, power, size.

JEL classification: C14, C21, C25, C26.

1. Introduction

Since the seminal work of Heckman (1978), bivariate probit models have earned a lot of attention in social sciences. The bivariate probit model provides enough structure to point-identify traditional parameters of interest such as the average treatment effect (ATE), and its counterparts for the treated (ATT) and untreated (ATU) groups. While researchers recognize the restrictive nature of this model, it remains a common approach in the literature. Influential examples include Evans and Schwab (1995), Neal (1997), and Altonji, Elder, and Taber (2005a,b) who use a bivariate probit model to estimate the relative effectiveness

Date: The present version is as of February 16, 2021. We thank Ismael Mourifié, Wendong Zhang, participants at Iowa State University, the Econometric Society 2020 Virtual World Congress, and the 2020 Southern Economic Association Annual Meeting for helpful comments. All errors are ours. Corresponding address: 260 Heady Hall, 518 Farm House Lane, Ames, IA, 50011, USA. Email addresses: acerenza@iastate.edu, bartalot@iastate.edu, dkedagni@iastate.edu.

of Catholic schools. Goldman et al. (2001) develop a bivariate probit model of insurance and mortality to explain the correlation between unobserved health and insurance status, leading to the counter-intuitive results that HIV-positive individuals receiving regular medical care using insurance have a higher probability of death. Finally, Rhine, Greene, and Toussaint-Comeau (2006) model the consumer's decision to patronize check-cashing businesses jointly with the decision to be "unbanked." However, the literature related to testing the validity of these models and its assumptions remains underdeveloped.

This paper derives testable implications for the identifying assumptions in bivariate probit models, and proposes a testing procedure that can be used to check the falsifiability of such models. The standard bivariate probit model assumes joint normality of the outcome and treatment errors. Identification of the usual parameter of interest - the coefficient on the endogenous binary regressor - comes from three main sources: (i) instrument exogeneity, that is, its exclusion from the outcome equation and its independence from the two latent variables in the triangular system of equations, (ii) the joint normality of these latent variables, and (iii) the relevance condition for the instrument. The exogeneity condition alone is not sufficient to point-identify the coefficient of interest. However, it allows partial identification of the ATE, the joint distribution of the potential outcomes, and other features of interest. In cases when the bounds are uninformative, researchers often add restrictions to the model in order to draw conclusions about these features. This paper focuses on testing these identifying restrictions.

There is a growing literature on the testability of the identifying assumptions in various econometric models. Pearl (1994) derived testable implications for instrumental variables when the endogenous regressor is discrete. Balke and Pearl (1997) provided testable inequalities for the local average treatment effect (LATE) assumptions when the outcome, treatment, and instrument are all binary. Heckman and Vytlacil (2005) generalized those results to the case where the outcome has no support restrictions and can be discrete, continuous or mixed. Kitagawa (2015) and Mourifié and Wan (2017) developed two different statistical tests for those inequalities. Huber and Mellace (2015) developed an alternative test for a version of the LATE assumptions under mean independence instead of full independence. Recently, Kédagni and Mourifié (2020) have complemented and generalized Pearl's (1994) testable inequalities to the case of discrete treatment with unrestricted outcome and instruments. Building on Pearl's (1995) conjecture, Gunsilius (2020) showed that there is no testable restriction in the continuous treatment case. Arai et al. (2018) developed a test for the identifying assumptions in the regression discontinuity design framework.

This paper has three main contributions. The first is to provide novel sharp testable equalities that can detect all possible observable violations of the bivariate probit model. However, these testable equalities are difficult to implement in practice.

Second, we propose a test for the validity of the identifying assumptions in the bivariate probit model using feasible testable implications implied by the sharp equalities. The feasible testable implications take the form of conditional moment inequalities, which can be implemented using existing inferential methods such as Chernozhukov, Lee, and Rosen (2013) or Andrews and Shi (2013). The test is extended to cover the inclusion of exogenous covariates and a broad class of bivariate distributions (see Appendix D). Monte Carlo simulations suggest that the proposed test adequately controls size in large samples, though it tends to over-reject in small samples. Furthermore, the test has power to detect violations of either the exclusion restriction and independence assumption or the joint normality, separately.

The third contribution is to provide researchers with a road map on what to do when the bivariate probit model is rejected. In particular, we provide novel bounds for the ATE that are valid even after relaxation of the normality assumption, building upon results in Machado, Shaikh, and Vytlacil (2019) and Kédagni and Mourifié (2020). Finally, we provide empirical examples to illustrate our methodology and its practical relevance.

The remainder of the paper is organized as follows. Section 2 presents the model and the identifying assumptions. Section 3 discusses identification of the model's parameters, and introduces the testable implications. Section 4 discusses the testing procedure for the feasible testable implications. Section 5 includes simulation results about the size and power of the test. Section 6 discusses how to relax the assumptions when they are rejected. Section 7 provides two empirical illustrations. Finally, Section 8 concludes.[1]

2. The baseline model

Consider the following model,

$$\begin{cases} Y = \mathbb{1}\{\beta + \alpha D - U \geqslant 0\} \\ D = \mathbb{1}\{\gamma + \delta Z - V \geqslant 0\} \end{cases} \tag{2.1}$$

[1]Additional results are discussed in the appendix. Appendix A has the proof of our main propositions. Appendix B contains some additional remarks. Appendix C contains the proof of the consistency and asymptotic behavior of the test. Appendix D generalizes our framework to (i) the inclusion of exogenous covariates, and (ii) a more general copula theory that can accommodate a broad class of bivariate distributions beyond joint normality. Appendix E has additional results for the applications.

where the vector (Y, D, Z) is the observed data, Y is a binary outcome, D is a binary treatment, $Z \in \mathcal{Z}$ is a potential instrument, (U, V) is a vector of latent variables, β, α, γ and δ are the model parameters, while α is of interest. For simplicity, we drop exogenous covariates from the model. All results derived henceforth hold conditional on covariates.

Interesting applications from the previously mentioned literature fit this setup. For example, the outcome variable (Y) could be mortality, measures of college or labor market success, such as employment status, or the decision to patronize a check-cashing business. Treatment (D) could be health insurance availability, catholic schooling attendance, having a college degree, or the decision to be "unbanked." The instrument (Z) could be eligibility thresholds of an insurance policy, being Catholic, geographic proximity to Catholic schools, owning a house, or college tuition.

Under this framework, the bivariate probit model identifying assumptions are as follows.

Assumption 1 (Random Assignment)**.** *$Z \perp\!\!\!\perp (U, V)$.*

Assumption 2 (Normality)**.** *The vector $(U, V)'$ follows the standard bivariate normal distribution with covariance ρ, i.e., $\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.*

Assumption 3 (Relevance)**.** *$\delta \neq 0$.*

Assumption 1 states that the instrument is independent of all the unobservables in the model. In the Catholic school attendance example, it requires that being Catholic is unrelated with unobserved factors that influence the decision to attend a Catholic school and student performance. Assumption 2 assumes that the vector of the unobservables in the model is jointly normally distributed. This assumption makes the model fully parametric, and eases the identification of the model parameters. Assumption 3 states that the instrument is relevant in explaining variations in the treatment variable, e.g., that being Catholic has a direct effect on attending a Catholic school.

Under Assumptions 1, 2 and 3, the parameters β, α, γ, δ and ρ are identified.[2] We can therefore identify the average treatment effect ATE, defined as $\mathbb{E}[Y_1 - Y_0]$, where $Y_1 = \mathbb{1}\{\beta + \alpha - U \geqslant 0\}$ and $Y_0 = \mathbb{1}\{\beta - U \geqslant 0\}$. Indeed, under Assumption 2 we have

[2] We briefly discuss these parameters' identification in Section 3. See Han and Vytlacil (2017), Han and Lee (2019), Mourifié and Méango (2014) for detailed identification results. Note that Li, Poskitt, and Zhao (2019) provide conditions for identification by functional form in the absence of an instrument. We focus on the testability of the standard bivariate probit model with an excluded variable, as it is the most commonly used.

$ATE = \Phi(\beta+\alpha) - \Phi(\beta)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (cdf).

3. Testable implications

In this section, we derive testable implications implied by the bivariate probit model described above. First, we heuristically discuss identification of β, α, γ, δ and ρ.

3.1. Identification.

Assumption 1 implies

$$\mathbb{P}(Y=1, D=1|Z=z) = \mathbb{P}(U \leqslant \beta+\alpha, V \leqslant \gamma+\delta z), \tag{3.1}$$

$$\mathbb{P}(Y=1, D=0|Z=z) = \mathbb{P}(U \leqslant \beta, V > \gamma+\delta z), \tag{3.2}$$

$$\mathbb{P}(Y=0, D=1|Z=z) = \mathbb{P}(U > \beta+\alpha, V \leqslant \gamma+\delta z), \tag{3.3}$$

$$\mathbb{P}(Y=0, D=0|Z=z) = \mathbb{P}(U > \beta, V > \gamma+\delta z). \tag{3.4}$$

Combining Equations (3.1) and (3.3), we have $\mathbb{P}(D=1|Z=z) = \mathbb{P}(V \leqslant \gamma+\delta z)$, which is equal to $\Phi(\gamma+\delta z)$ under Assumption 2. Then, $\Phi^{-1}(\mathbb{P}(D=1|Z)) = \gamma+\delta Z$. Hence, γ and δ are identified as follows: $\delta = \frac{Cov\left(\Phi^{-1}(\mathbb{P}(D=1|Z)),Z\right)}{Var(Z)}$, and $\gamma = \mathbb{E}\left[\Phi^{-1}(\mathbb{P}(D=1|Z))\right] - \delta\mathbb{E}[Z]$.

Under Assumption 2, the latent variables U and V are jointly normal, and we can write the linear projection $U = \rho V + e$, where $e \perp\!\!\!\perp V$ and $e \sim N(0, 1-\rho^2)$. Equation (3.1) implies

$$\mathbb{P}(Y=1|D=1,Z=z) = \mathbb{P}(U \leqslant \beta+\alpha | V \leqslant \gamma+\delta z) =$$
$$= \mathbb{P}\left(\frac{e}{\sqrt{1-\rho^2}} \leqslant a - \frac{\rho}{\sqrt{1-\rho^2}} V | V \leqslant \Phi^{-1}(\mathbb{P}(D=1|Z=z))\right),$$

where $a = \frac{\beta+\alpha}{\sqrt{1-\rho^2}}$. Since this function is strictly increasing in a, we can invert it to identify a for a particular value of ρ. Similarly, using Equation (3.2), we can identify $b = \frac{\beta}{\sqrt{1-\rho^2}}$. Then, $\beta = b\sqrt{1-\rho^2}$, and $\alpha = (a-b)\sqrt{1-\rho^2}$ can be recovered.

Let $a(\rho, z)$ and $b(\rho, z)$ describe a and b as a function of ρ and z. Under Assumption 1, ρ must satisfy $a(\rho, z) = a(\rho, z')$ and $b(\rho, z) = b(\rho, z')$ for all z, $z' \in \mathcal{Z}$. Han and Vytlacil (2017) show that ρ is uniquely determined if $\delta \neq 0$ (i.e., under Assumption 3).[3]

[3] See Section 4 in Han and Vytlacil (2017) for a more complete discussion.

3.2. **Sharp testable implications.** Denote $P(z) \equiv \mathbb{P}(D=1|Z=z)$. We have

$$\begin{aligned}
\mathbb{E}\left[YD|P(Z)=p\right] &= \mathbb{P}(Y=1, D=1|P(Z)=p), \\
&= \mathbb{P}\left(U \leqslant \beta+\alpha, V \leqslant \gamma+\delta Z|P(Z)=p\right), \\
&= \mathbb{P}\left(U \leqslant \beta+\alpha, \Phi(V) \leqslant \Phi(\gamma+\delta Z)|P(Z)=p\right), \\
&= \mathbb{P}\left(U \leqslant \beta+\alpha, \Phi(V) \leqslant p\right),
\end{aligned}$$

where the first equality holds because Y, and D are binary, the second holds by the model's definition. The third equality follows as $\Phi(\cdot)$ is an increasing function, and the last holds under Assumption 1.

For $p > p'$, we have $\mathbb{E}\left[YD|P(Z)=p\right] - \mathbb{E}\left[YD|P(Z)=p'\right] = \mathbb{P}\left(U \leqslant \beta+\alpha, p' < \Phi(V) \leqslant p\right)$. Therefore,

$$\begin{aligned}
\frac{\mathbb{E}\left[YD|P(Z)=p\right] - \mathbb{E}\left[YD|P(Z)=p'\right]}{p-p'} &= \frac{\mathbb{P}\left(U \leqslant \beta+\alpha, p' < \Phi(V) \leqslant p\right)}{p-p'}, \\
&= \frac{\mathbb{P}\left(U \leqslant \beta+\alpha, p' < \Phi(V) \leqslant p\right)}{\mathbb{P}(p' < \Phi(V) \leqslant p)}, \\
&= \mathbb{P}\left(U \leqslant \beta+\alpha | p' < \Phi(V) \leqslant p\right), \\
&= \mathbb{P}\left(U \leqslant \beta+\alpha | \Phi^{-1}(p') < V \leqslant \Phi^{-1}(p)\right),
\end{aligned}$$

where the second equality holds by assumption 2, making $\Phi(V)$ uniformly distributed. The third equality holds by the definition of conditional probability, while the fourth equality follows from $\Phi(\cdot)$ being strictly increasing.

Similarly,

$$-\frac{\mathbb{E}\left[Y(1-D)|P(Z)=p\right] - \mathbb{E}\left[Y(1-D)|P(Z)=p'\right]}{p-p'} = \mathbb{P}\left(U \leqslant \beta | \Phi^{-1}(p') < V \leqslant \Phi^{-1}(p)\right).$$

When Z is continuous, by taking the limit when p' goes to p and using the fact that $U = \rho V + e$, we have the following two testable equalities:

$$\frac{\partial \mathbb{E}\left[YD|P(Z)=p\right]}{\partial p} = \Phi\left(\frac{\beta+\alpha-\rho\Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right), \tag{3.5}$$

$$-\frac{\partial \mathbb{E}\left[Y(1-D)|P(Z)=p\right]}{\partial p} = \Phi\left(\frac{\beta-\rho\Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right). \tag{3.6}$$

These equalities are testable since the model parameters are identified. Recall that under the model assumptions the propensity score has a probit specification:

$$P(z) = \Phi(\gamma+\delta z). \tag{3.7}$$

Proposition 1 summarizes these results.

Proposition 1. *Assume that Z is continuous. Under Assumptions 1-3 in the bivariate probit model (2.1), the parameters α, β, δ and ρ are identified, and equalities (3.5), (3.6), and (3.7) must hold. Moreover, these equalities are sharp.*

Remark 1 (Sharpness)**.** *In the context of model (2.1), equalities (3.5), (3.6), and (3.7) are sharp in the sense that whenever they hold, it is possible to construct a vector of $(\tilde{Y}, \tilde{D}, \tilde{U}, \tilde{V}, Z)$ that satisfies model (2.1), Assumptions 1-3, and induces the observed distribution on the data (Y, D, Z).*

The previous equalities are sharp, but they are difficult to test in practice. For this reason, we now provide a set of inequalities implied by Equations (3.5), (3.6), and (3.7), which are easily implemented.

3.3. **Non-sharp testable implications.** In this section, we derive testable implications that can be easily implemented using the intersection bounds framework.

Equation (3.1) implies $\mathbb{P}(Y = 1, D = 1|Z = z) \leqslant \mathbb{P}(U \leqslant \beta + \alpha)$ for all z. Note that the right-hand side does not depend on Z. Under Assumption 2, we have $\mathbb{P}(U \leqslant \beta + \alpha) = \Phi(\beta + \alpha)$. Thus, the following testable implication must hold under model (2.1) and Assumptions 1-2:

$$\sup_z \mathbb{P}(Y = 1, D = 1|Z = z) \leqslant \Phi(\beta + \alpha).$$

Similarly, using equations (3.2), (3.3), and (3.4), we obtain the following testable implications

$$\begin{aligned}
\sup_z \mathbb{P}(Y = 1, D = 0|Z = z) &\leqslant \Phi(\beta), \\
\sup_z \mathbb{P}(Y = 0, D = 1|Z = z) &\leqslant 1 - \Phi(\beta + \alpha), \\
\sup_z \mathbb{P}(Y = 0, D = 0|Z = z) &\leqslant 1 - \Phi(\beta),
\end{aligned}$$

respectively. These four inequalities impose upper bounds on the joint distribution of (Y, D) conditional on the instrument Z. Note that only information about the marginal distribution of the unobserved heterogeneity Y_1 and Y_0 is used to derive these inequalities. They imply the Pearl (1994) instrumental inequalities, which are obtained by adequately combining the previous inequalities:

$$\begin{aligned}
\sup_z \mathbb{P}(Y = 1, D = 1|Z = z) + \sup_z \mathbb{P}(Y = 0, D = 1|Z = z) &\leqslant 1, \\
\sup_z \mathbb{P}(Y = 1, D = 0|Z = z) + \sup_z \mathbb{P}(Y = 0, D = 0|Z = z) &\leqslant 1.
\end{aligned}$$

Notably, the Pearl (1994) inequalities were derived in a more general potential outcome model, with no restriction on the distribution of the unobservables. In that context, the ATE is only partially identified, while it is point-identified in the current model. To derive further testable implications, we follow Kédagni and Mourifié (2020) to use restrictions that this model imposes on the joint distribution of (Y_0, Y_1). We have

$$\begin{aligned}\mathbb{P}(Y_0, Y_1) = \mathbb{P}(U \leqslant \beta+\alpha, U \leqslant \beta) &= \mathbb{P}(U \leqslant \beta+\alpha, U \leqslant \beta, D=1|Z=z) \\ &\quad +\mathbb{P}(U \leqslant \beta+\alpha, U \leqslant \beta, D=0|Z=z), \\ &\leqslant \mathbb{P}(U \leqslant \beta+\alpha, D=1|Z=z) + \mathbb{P}(U \leqslant \beta, D=0|Z=z), \\ &= \mathbb{P}(Y=1, D=1|Z=z) + \mathbb{P}(Y=1, D=0|Z=z),\end{aligned}$$

where the first equality holds from Assumption 1 and the law of total probability, the inequality follows from the monotonicity of a probability measure, and the last equality follows from the model's definition. Under Assumption 2, we have $\mathbb{P}(U \leqslant \beta+\alpha, U \leqslant \beta) = \Phi(\min(\beta+\alpha, \beta))$. Therefore, by taking the infimum over z, the following must hold under the assumptions:

$$\Phi(\min(\beta+\alpha, \beta)) \leqslant \inf_z \mathbb{P}(Y=1|Z=z).$$

Using a similar reasoning, we derive the additional testable implication in Equation (3.13). The new inequalities impose lower bounds on the marginal distribution of Y given the instrument Z. We summarize this discussion in the following proposition.

Proposition 2. *Under Assumptions 1, 2 and 3, the parameters α, β, δ and ρ are identified, and the following inequalities hold:*

$$\sup_z \mathbb{E}[YD|Z=z] \leqslant \Phi(\beta+\alpha), \tag{3.8}$$

$$\sup_z \mathbb{E}[Y(1-D)|Z=z] \leqslant \Phi(\beta), \tag{3.9}$$

$$\sup_z \mathbb{E}[(1-Y)D|Z=z] \leqslant 1-\Phi(\beta+\alpha), \tag{3.10}$$

$$\sup_z \mathbb{E}[(1-Y)(1-D)|Z=z] \leqslant 1-\Phi(\beta), \tag{3.11}$$

$$\Phi(\min(\beta+\alpha, \beta)) \leqslant \inf_z \mathbb{E}[Y|Z=z], \tag{3.12}$$

$$1-\Phi(\max(\beta+\alpha, \beta)) \leqslant \inf_z \mathbb{E}[1-Y|Z=z]. \tag{3.13}$$

Inequalities (3.8)-(3.13) are implied by the sharp equalities derived in the Section 3.2,[4] and imply the generalized instrumental inequalities of Kédagni and Mourifié (2020), since

[4] This result is collected in Remark 4 in Appendix B.

we are testing a stronger set of assumptions than those considered in their paper. We can estimate the vector of parameters $(\alpha, \beta, \delta, \gamma, \rho)$, and then use the estimates $\hat{\alpha}, \hat{\beta}$ to test these inequalities using the intersection bounds framework of Chernozhukov, Lee, and Rosen (2013) or Andrews and Shi (2013).[5]

Remark 2 (Exogenous Covariates and Generalized Bivariate Models). *The inequalities in Proposition 2 can be generalized for two empirically relevant cases, (i) the inclusion of exogenous covariates, and (ii) a more general copula theory that can accommodate a broad class of bivariate distributions (see Appendix D).*

4. Testing procedure

To test inequalities (3.8) to (3.13), we write them in the Chernozhukov, Lee, and Rosen (2013) intersection bounds framework, and use the Stata package described in Chernozhukov et al. (2015) for direct implementation. Testing inequalities (3.8) to (3.13) is equivalent to testing:

$$\begin{aligned}
&\sup_z \mathbb{E}\left[YD - \Phi(\beta+\alpha)|Z=z\right] && \leqslant 0,\\
&\sup_z \mathbb{E}[Y(1-D) - \Phi(\beta)|Z=z] && \leqslant 0,\\
&\sup_z \mathbb{E}\left[(1-Y)D - 1 + \Phi(\beta+\alpha)|Z=z\right] && \leqslant 0,\\
&\sup_z \mathbb{E}[(1-Y)(1-D) - 1 + \Phi(\beta)|Z=z] && \leqslant 0,\\
&\sup_z \mathbb{E}\left[\Phi(\min(\beta+\alpha,\beta)) - Y|Z=z\right] && \leqslant 0,\\
&\sup_z \mathbb{E}[1 - \Phi(\max(\beta+\alpha,\beta)) - (1-Y)|Z=z] && \leqslant 0.
\end{aligned}$$

To implement the test, we replace α and β by their maximum likelihood estimators (MLE), denoted $\hat{\alpha}$ and $\hat{\beta}$, respectively. In Appendix C we show the asymptotic properties of the test are unaffected by using $\hat{\alpha}$ and $\hat{\beta}$ when nonparametric estimators for the conditional expectations are used.

We now briefly describe the method. First, write the inequalities above as the null hypothesis.

$$H_0 : \theta_0 \equiv \max_{j\in\{1,\ldots,6\}} \sup_{z\in\mathcal{Z}} \theta(z,j) \leqslant 0,$$

where $\theta(z,j) \equiv \mathbb{E}[W_j|Z=z]$, and W_j represents the expression in the conditional expectation for inequality j. For example, $W_1 = YD - \Phi(\hat{\alpha}+\hat{\beta})$. The decision rule for the test is

[5] There exist two extra inequalities which are redundant and, as such, reduce the power (and possibly the size) of the test. Refer to the generalized moment selection method which selects only binding constraints (Andrews and Soares, 2010). These two inequalities are collected in Remark 3 in Appendix B.

given by Chernozhukov, Lee, and Rosen (2013), we reject H_0 if

$$\hat{\theta}_{1-\alpha} \equiv \max_{j\in\{1,\ldots,6\}} \sup_{z\in\mathcal{Z}} \left\{\hat{\theta}(z,j) - k_{1-\alpha}\hat{s}(z,j)\right\} > 0,$$

where $\hat{\theta}(z,j)$ is the local linear estimator for $\theta(z,j)$, $\hat{s}(z,j)$ its standard error, and $k_{1-\alpha}$ is a critical value at the significance level α. Details about the implementation can be found in Appendix E.

5. Monte Carlo simulations

This section presents simulation results for the size and power of the test for validity of the bivariate probit model based on the intersection bounds framework (Chernozhukov, Lee, and Rosen, 2013) exploiting the inequalities derived in Section 3.3.

5.1. **Size.** Consider the following data generating process (DGP) where Assumptions 1, 2 and 3 hold for the bivariate probit model:

$$\begin{cases} Y &= \mathbb{1}\{D - U \geqslant 0\} \\ D &= \mathbb{1}\{2Z - V \geqslant 0\} \end{cases} \tag{5.1}$$

where $\begin{pmatrix} U \\ V \end{pmatrix} \sim N(0,\Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $Z \sim \mathcal{U}_{[-3,3]}$, and $Z \perp\!\!\!\perp (U,V)$.

Table 1 shows the false rejection rates for different sample and nominal sizes. Each simulation relies on 500 replications. As expected, with reasonably large sample sizes the tests empirical and nominal sizes converge.

Table 1. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$n = 200$	48%	42%	37%
$n = 1000$	20%	16%	8%
$n = 2000$	18%	12%	5%
$n = 3000$	14%	9%	4%
$n = 6000$	11%	6%	1%

Based on 500 replications.

5.2. **Power.** The simulations below consider violations and relaxations of assumptions 1-3 to examine under which circumstances the proposed test is more powerful to detect violations of the bivariate probit model.

Three DGPs consider violations of random assignment (Assumption 1) caused by endogeneity between Z and (U, V), while the remaining assumptions hold.

$$\begin{cases} Y &= \mathbb{1}\{D - U \geqslant 0\} \\ D &= \mathbb{1}\{2Z - V \geqslant 0\} \end{cases} \tag{5.2}$$

where $\begin{pmatrix} U \\ V \\ Z \end{pmatrix} \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 & \nu \\ 0.5 & 1 & \eta \\ \nu & \eta & 1 \end{pmatrix}$.

The coefficients ν and η capture the degree and source of violation of the random assignment assumption. The parameter ν summarizes the endogeneity between the instrument and latent determinants of the outcome, while η describes the dependence of the instrument with unobservable drivers of the selection into treatment. The instrument validity holds when $\nu = \eta = 0$.

The first DGP considers the case in which $\nu = \eta \neq 0$, considering a general violation of Assumption 1. Table 2 reports the empirical power of the test, which climbs rapidly for small deviations from instrument independence, even in finite samples.

TABLE 2. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$\nu = \eta = 0.01$	39%	32%	25%
$\nu = \eta = 0.1$	85%	76%	54%
$\nu = \eta = 0.5$	100%	100%	100%
$\nu = \eta = 0.8$	100%	100%	100%

Based on 500 replications with sample size 6000.

To provide better intuition on the test's ability to detect different sources of violations of Assumption 1, the second and third DGPs examine deviations from the instrument's random assignment originating solely on the outcome or treatment equations, respectively. Hence, Table 3 reports the empirical rejection rates when the instrument is endogenous through the outcome only (DGP 2), that is, $\nu \neq 0$ and $\eta = 0$. Results are similar to the ones in Table 2.

Conversely, Table 4 reports the empirical rejection rates when the instrument is endogenous only through the treatment selection (DGP 3), that is, $\nu = 0$ and $\eta \neq 0$.

Table 3. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$\nu = 0.01$	41%	33%	24%
$\nu = 0.1$	87%	79%	55%
$\nu = 0.5$	100%	100%	100%
$\nu = 0.8$	100%	100%	100%

Based on 500 replications with sample size 6000.

Table 4. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$\eta = 0.01$	40%	34%	27%
$\eta = 0.1$	43%	38%	28%
$\eta = 0.5$	40%	35%	27%
$\eta = 0.8$	44%	36%	28%

Based on 500 replications with sample size 6000.

Combining the results in tables 2-4 indicates that, at least in this case, the feasible test implemented is more powerful in detecting violations coming from correlation from U and Z. This result seems reasonable since conditional on knowing the model parameters, the non-sharp testable inequalities are derived using only the independence between Z and U.

The remaining three DGPs consider violations of the joint normality assumption while instrument random assignment and relevance hold. It is natural to expect that the proposed tests would perform better in rejecting the model when the joint distribution of (U, V) is very distinct from a bivariate normal, as opposed to less obvious deviations. Hence, DGPs 4-6 offer evidence of the test's empirical power in three alternative scenarios with varying levels of severity in the deviations from normality.

In DGP 4, the joint distribution of the unobservables (U, V) is a convex combination between log-normal and normal distributions with weight $\lambda \in [0, 1]$. When $\lambda = 1$, the bivariate probit model holds.

$$\begin{cases} Y &= \mathbb{1}\{D - U \geqslant 0\} \\ D &= \mathbb{1}\{2Z - V \geqslant 0\} \end{cases} \tag{5.3}$$

where $U = (1-\lambda)\ln U^* + (\lambda)U^{**}$, $V = (1-\lambda)\ln V^* + (\lambda)V^{**}$, $\begin{pmatrix} U^* \\ V^* \end{pmatrix} \sim N(0,\Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\begin{pmatrix} U^{**} \\ V^{**} \end{pmatrix} \sim N(0,\Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and $Z \perp\!\!\!\perp (U^*, V^*, U^{**}, V^{**})$. Table 5 presents the empirical rejection rates for different values of λ. The test performs well, and violations of the distributional assumptions are easily detected even for reasonably small deviations from joint normality.

Table 5. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$\lambda = 0.1$	100%	100%	100%
$\lambda = 0.5$	100%	100%	100%
$\lambda = 0.8$	87%	83%	71%
$\lambda = 1$	8%	5%	2%

Based on 500 replications with sample size 6000.

In the fifth DGP, we consider a similar violation of the normality assumption, replacing the log-normal with a uniform distribution (taking values between 3 std. deviations) in the convex combination generating the vector (U, V). Table 6 indicates that violations of joint normality in this scenario are easier to detected than those arising from log-normality.

Table 6. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$\lambda = 0.5$	100%	100%	100%
$\lambda = 0.95$	100%	100%	100%
$\lambda = 0.99$	29%	18%	7%
$\lambda = 1$	11%	6%	1%

Based on 500 replications with sample size 6000.

Finally, the sixth DGP considers a case where the test is (understandably) not powerful, to highlight when it might fail. In this case the underlying joint distribution of the unobserved variables is a Student's t with v degrees of freedom. Formally, $U = T^{-1}(\Phi(U^*), v)$, $V = T^{-1}(\Phi(V^*), v)$. Where $T^{-1}(., v)$ is the inverse CDF of the standard central t-distribution with v degrees of freedom. Naturally, when $v = \infty$ this approximates the bivariate probit model. Table 7 presents the results for that case, highlighting the lack of sensitivity of the test for different values of v.

Table 7. Rejection Frequency (clrbound)

	Local linear method		
Nominal Size	10%	5%	1%
$v = 5$	13%	8%	3%
$v = 100$	14%	9%	4%
$v = 10^{12}$	14%	9%	4%

Based on 500 replications with sample size 6000.

The test struggles to reject a normal distribution from a Student's t, even for low v, which is intuitive since the t-distribution is similar to the normal distribution with fatter tails.[6]

6. What to do when the testable implications are rejected

When the test rejects inequalities (3.8) to (3.13), the researcher could relax some of the identifying assumptions in order to study identification of the treatment effect.

For example, one can relax the normality assumption and bound the ATE under the random assignment assumption following Kédagni and Mourifié (2020).

Suppose that only Assumption 1 holds. Then, by the identification results of Kédagni and Mourifié (2020, Proposition 1) can be used to obtain the following bounds on the potential outcome means $\mathbb{E}[Y_0]$ and $\mathbb{E}[Y_1]$, respectively:

$$\max\left\{\sup_z \mathbb{E}[Y(1-D)|Z=z], 1-\inf_z \mathbb{E}[1-Y|Z=z]-\inf_z \mathbb{E}[(1-Y)(1-D)+YD|Z=z]\right\}$$
$$\leqslant \mathbb{E}[Y_0] \leqslant$$
$$\min\left\{1-\sup_z \mathbb{E}[(1-Y)(1-D)|Z=z], \inf_z \mathbb{E}[Y|Z=z]+\inf_z \mathbb{E}[Y(1-D)+(1-Y)D|Z=z]\right\},$$

and

$$\max\left\{\sup_z \mathbb{E}[YD|Z=z], 1-\inf_z \mathbb{E}[1-Y|Z=z]-\inf_z \mathbb{E}[Y(1-D)+(1-Y)D|Z=z]\right\}$$
$$\leqslant \mathbb{E}[Y_1] \leqslant$$
$$\min\left\{1-\sup_z \mathbb{E}[(1-Y)D|Z=z], \inf_z \mathbb{E}[Y|Z=z]+\inf_z \mathbb{E}[(1-Y)(1-D)+YD|Z=z]\right\}.$$

Constructing confidence bounds for these potential outcome means is challenging as the identified sets involve the summation of extrema over the support of the instrument. We

[6]Similar results are observed when a using a standard Laplace distribution, which is very similar to Gaussian distributions except very close to the mean.

follow Kédagni and Mourifié (2020) to combine a sample splitting approach with the intersection bounds framework.

Suppose that we have two independent copies $(Y^{(1)}, D^{(1)}, Z^{(1)})$ and $(Y^{(2)}, D^{(2)}, Z^{(2)})$ of the data (Y, D, Z). In practice, two independent copies of the data can be obtained by randomly splitting the sample into two subsamples if the original data are independent and identically distributed.

Then, we can write these identified sets in the intersection bounds framework to perform inference. We have:[7]

$$\begin{aligned}
&\max\Bigg\{\sup_z \mathbb{E}\left[Y^{(1)}\left(1-D^{(1)}\right)|Z^{(1)}=z\right],\\
&\qquad\sup_{z,z'}\mathbb{E}\left[Y^{(1)}-\left(1-Y^{(2)}\right)\left(1-D^{(2)}\right)-Y^{(2)}D^{(2)}|Z^{(1)}=z,Z^{(2)}=z'\right]\Bigg\}\\
&\leqslant \mathbb{E}[Y_0]\leqslant\\
&\min\Bigg\{\inf_z \mathbb{E}\left[1-\left(1-Y^{(1)}\right)\left(1-D^{(1)}\right)|Z^{(1)}=z\right],\\
&\qquad\inf_{z,z'}\mathbb{E}\left[Y^{(1)}+Y^{(2)}\left(1-D^{(2)}\right)+\left(1-Y^{(2)}\right)D^{(2)}|Z^{(1)}=z,Z^{(2)}=z'\right]\Bigg\},
\end{aligned}$$

and

$$\begin{aligned}
&\max\Bigg\{\sup_z \mathbb{E}\left[Y^{(1)}D^{(1)}|Z^{(1)}=z\right],\\
&\qquad\sup_{z,z'}\mathbb{E}\left[Y^{(1)}-Y^{(2)}\left(1-D^{(2)}\right)-\left(1-Y^{(2)}\right)D^{(2)}|Z^{(1)}=z,Z^{(2)}=z'\right]\Bigg\}\\
&\leqslant \mathbb{E}[Y_1]\leqslant\\
&\min\Bigg\{\inf_z \mathbb{E}\left[1-\left(1-Y^{(1)}\right)D^{(1)}|Z^{(1)}=z\right],\\
&\qquad\inf_{z,z'}\mathbb{E}\left[Y^{(1)}+\left(1-Y^{(2)}\right)\left(1-D^{(2)}\right)+Y^{(2)}D^{(2)}|Z^{(1)}=z,Z^{(2)}=z'\right]\Bigg\}.
\end{aligned}$$

We can therefore obtain bounds on the ATE by taking the difference of the bounds on $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$. We again can use the Chernozhukov et al. (2015) Stata package to implement these bounds. If one of those bounds is empty, then Assumption 1 is rejected. In such a case, the researcher could resort to other identification strategies such as the monotone

[7]See Kédagni and Mourifié (2020) for more details on the procedure.

instrumental variable approach ($\mathbb{E}\left[Y_d|Z=z\right]$ is monotone in z for each d) developed by Manski and Pepper (2000), or some sensitivity analysis like the one developed by Altonji, Elder, and Taber (2005b).

The above bounds can be tightened by, in addition to Assumption 1, imposing monotonicity of the outcome Y in the treatment D, as proposed by Machado, Shaikh, and Vytlacil (2019). Note that this assumption is implicit in the bivariate probit specification.

Assumption 4 (Monotonicity of Y in D)**.** *Either $Y_1 \geqslant Y_0$ a.s. or $Y_1 \leqslant Y_0$ a.s..*

As pointed out by Machado, Shaikh, and Vytlacil (2019), Assumption 4 is weaker than the "monotone treatment response" considered in Manski (1997), and Manski and Pepper (2000), which assumes that the direction of the monotonicity is known *a priori*. Then, the following proposition holds.[8]

Proposition 3. *Under Assumptions 1 and 4, sharp bounds for $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$ are:*

$$\begin{cases} \sup_z \mathbb{E}[Y|Z=z] \leqslant \mathbb{E}[Y_1] \leqslant \inf_z \mathbb{E}[1-(1-Y)D|Z=z] \\ \sup_z \mathbb{E}[Y(1-D)|Z=z] \leqslant \mathbb{E}[Y_0] \leqslant \inf_z \mathbb{E}[Y|Z=z] \end{cases} \tag{6.1}$$

or

$$\begin{cases} \sup_z \mathbb{E}[YD|Z=z] \leqslant \mathbb{E}[Y_1] \leqslant \inf_z \mathbb{E}[Y|Z=z] \\ \sup_z \mathbb{E}[Y|Z=z] \leqslant \mathbb{E}[Y_0] \leqslant \inf_z \mathbb{E}[1-(1-Y)(1-D)|Z=z] \end{cases} \tag{6.2}$$

Sharp bounds on the ATE are obtained by taking the difference of the bounds on $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$. The sign of the ATE is identified only if one of the identified sets in Equations (6.1) and (6.2) is empty. Note that Balke and Pearl (1997) showed that adding monotonicity of D in Z does not improve the bounds for the ATE under Assumption 1 when the outcome, treatment and instrument are all binary. Hence, we conjecture and prove in Appendix G that imposing monotonicity of the treatment in the instrument (which is also implicit in the bivariate probit model) will not further tighten the bounds if the model is not misspecified.

If we consider the case in which only Assumption 2 holds, Mourifié and Méango (2014) showed that when $\delta = 0$ the model is generally underidentified. In such scenario α and β are identified up to the coefficient of correlation ρ, the degree of endogeneity in the model.[9] Therefore, knowing the trivial bounds $(-1,1)$ on ρ, we can partially identify α and β. In general, the bounds on the parameters α and β may be wide (and possibly uninformative)

[8]The proof of the proposition is in Appendix F. The identified set in Proposition 3 takes the form of intersection bounds, and can be implemented using Chernozhukov et al.'s (2015) Stata package.

[9]See discussion in Subsection 3.1.

without additional restrictions. Alternatively, ρ and the selection on unobservables could be bounded by the degree of selection on observables as proposed by Altonji, Elder, and Taber (2005b). Finally, one can restrict ρ and ask what values are plausible using some economic argument, as suggested by Rosenbaum and Rubin (1983), and Rosenbaum (1995).

7. Empirical illustrations

To illustrate the usefulness of the tests developed in Section 4, we apply the methodology to data sets of two policy relevant recent papers. Our first empirical example uses the data set from Zimmer (2017) and Han and Lee (2019) to analyze the effect of access to health insurance on individuals' decision to visit a doctor. The second application revisits Gao et al. (2018) to analyze how land tenure arrangements affect Chinese farmers' adoption of straw retention.

7.1. **The effect of insurance on doctor visits.** Han and Lee (2019) analyze the impact of health insurance coverage on individual's decision to visit a doctor. In this example, Y and D are indicators for whether an individual has a doctor visit, and is covered by private health insurance, respectively. The instrument, Z, is the number of employees in the firm at which the individual works. The reasoning for instrument's validity holds that larger firms are more likely to provide health insurance to their workforce.

The data comes from the 2010 wave of the Medical Expenditure Panel Survey (MEPS). As in Han and Lee (2019), we focus on all the visits that occurred in January 2010, restrict the sample to contain individuals with age between 25 and 64, and exclude those who have retained any federal or state insurance in 2010. Furthermore, individuals who are self-employed or unemployed are excluded from the analysis.

Table 8 presents the estimates and standard errors for the parameters in the model, obtained by a bivariate probit framework. The first column presents estimates for the selection into treatment, and indicates that the number of employees in a firm increases the likelihood that an individual has private health insurance coverage. The second column reports a positive effect of having private health insurance on doctor visits, as economic theory predicts.

However, when we test for the validity of the bivariate probit model as described in Section 3.3, the model is rejected at all three conventional levels 1%, 5% and 10% as

Table 8. Bivariate probit specification

	MLE	
	Private insurance	Doctor visit
Nb employees	0.3661***	
	(0.0170)	
Private insurance		0.6388***
		(0.1189)
Constant	0.4374***	−1.3166***
	(0.0154)	(0.0707)
ρ		−0.2605***
		(0.0777)
n	7,555	7,555

Standard errors (in parentheses); ***: significant at 1% level.

$\hat{\theta}_{0.99} = 3.56 > 0$, $\hat{\theta}_{0.95} = 3.59 > 0$, and $\hat{\theta}_{0.90} = 3.60 > 0$. Even after controlling for gender, race, region, and marriage status, the rejection strongly remains in most cases.[10]

Confidence Bounds on ATE under Assumption 1. Given the strong rejection of the validity of the bivariate probit model in this case, we move to construct confidence bounds on the potential outcome means under Assumption 1 only, as discussed in Section 6.

The first column of Table 9 reports estimates of the bivariate probit model for the potential outcome means ($\mathbb{E}[Y_1] = \Phi(\hat{\beta}+\hat{\alpha})$, $\mathbb{E}[Y_0] = \Phi(\hat{\beta})$), and the ATE ($\mathbb{E}[Y_1 - Y_0]$), while the second and third columns report the 95% confidence lower and upper bounds, respectively, under the random assignment assumption only.

Despite the rejection of the bivariate probit model, the point estimates for $\mathbb{E}[Y_0]$ and $\mathbb{E}[Y_1]$ lie within the confidence regions relying solely on Assumption 1. Hence, we cannot reject the null hypothesis that the ATE is equal to the value 0.155 estimated by the bivariate probit model, as it lies within the confidence region for the ATE: $[-0.496, 0.292]$.

These results suggest that, (i) the rejection of the bivariate model is due to the joint normality and/or the relevance assumptions, and (ii) that there may exist a data generating process compatible with Assumption 1 that can yield this value, but it is not the standard bivariate probit model.

Using only the information embedded in Assumption 1, does not allow us to draw a conclusion about the direction of the effect of private insurance on doctor visits as it could be positive, zero or negative, based on the confidence bounds on ATE.

[10]See Appendix E.3 for more details.

Table 9. Confidence sets for parameters

Parameters	Biprobit estimates	95% conf. LB	95% conf. UB
$\mathbb{E}[Y_0]$	0.0940	0.0495	0.6467
$\mathbb{E}[Y_1]$	0.2490	0.1507	0.3414
ATE	0.1550	-0.4960	0.2920

conf.: confidence; LB: lower bound; UB: upper bound.

Confidence Bounds on ATE under Assumptions 1 and 4. In an effort to identify the sign of the ATE, we implement the bounds in Proposition 3, by exploiting the additional information in the monotonicity assumption. The results are shown in Table 10. The confidence set for the ATE is the union of the confidence regions under Assumption 1 and the assumptions $Y_1 \geqslant Y_0$ or $Y_1 \leqslant Y_0$, respectively.[11] Since the confidence region under the assumption $Y_1 \leqslant Y_0$ is empty, the confidence set for the ATE is $[0.0218, 0.2921]$, suggesting that having a private insurance increases the chances to visit a doctor by at least 2.2%, and the effect can be as high as 29.2%.

Table 10. Confidence sets for parameters

Parameters	95% conf. LB $Y_1 \geqslant Y_0$	95% conf. UB $Y_1 \geqslant Y_0$	95% conf. LB $Y_1 \leqslant Y_0$	95% conf. UB $Y_1 \leqslant Y_0$
$\mathbb{E}[Y_0]$	0.0516	0.1777	0.1998	0.6364
$\mathbb{E}[Y_1]$	0.1995	0.3438	Empty	Empty
ATE	0.0218	0.2921	Empty	Empty

conf.: confidence; LB: lower bound; UB: upper bound.

7.2. **Do farmers adopt fewer conservation practices on rented land?** Since conservation practices are costly and benefits spread over time, farmers' adoption is uneven. Economic theory suggests that land tenants will be less likely to adopt conservation practices due to land tenure insecurity, reducing investment, especially for those with higher initial costs or longer payoff horizons. Gao et al. (2018) investigate this theory in the case of Chinese farmers' adoption of straw retention, a key conservation practice to curb air pollution from burning crop residues. The standard bivariate probit model was one of their model specifications, which we revisit here.

[11] Berger and Hsu (1996) showed that the union of the confidence regions has at least the same coverage rate as each confidence region.

The data consists of 1659 plot-level observations from 670 farmer households drawn from a rural household survey conducted by Henan Agricultural University in 2016 in Henan Province, a major grain production province in central China.[12] The dependent variable Y is an indicator for the adoption of straw retention, and the treatment variable D is a dummy for rented plot. The authors use the ratio of annual income for family's migrant workers to annual agricultural profits for the farmer household as an instrument.[13] Table 11 reports the bivariate probit model results, which suggest that families with higher ratio of migrants' income to agricultural profits are less likely to rent land. The second column seems to support the idea that farmers who rent land are less likely to adopt straw retention. However, as in the previous empirical example, our test for the bivariate probit model is rejected at all three significance levels. We then proceed again to construct confidence regions for the potential outcome means and the ATE under the random assignment assumption only.

Table 11. Bivariate probit specification

	MLE	
	Rent	Adoption
Mig-ag-ratio	−0.0109**	
	(0.0049)	
Rent		−2.0984***
		(0.2216)
Constant	−1.2556***	0.6403***
	(0.0505)	(0.0332)
ρ		0.9559*
		(0.0987)
n	1,659	1,659

Standard errors in parentheses; ***: significant at 1% level, **: 5%, *: 10%.

Confidence Bounds on ATE under Assumption 1. The first column of Table 12 reports the bivariate probit estimates for the potential outcome means and the ATE, and the second and third columns display the 95% confidence lower and upper bounds for these parameters, respectively.

Both bivariate probit estimates for $\mathbb{E}[Y_0]$ and $\mathbb{E}[Y_1]$ lie within their confidence sets. The non-emptiness of the confidence regions suggests that Assumption 1 alone is not rejected by the data. The fact that the point estimate for the ATE lies within its confidence set

[12]We thank Wendong Zhang for sharing the data with us.

[13]See Gao et al. (2018) for more details on the construction of the instrument.

suggests that the value -0.67 cannot be rejected as the effect of land leasing on straw retention adoption, but the data generating process cannot be the bivariate probit model. Once again, Assumption 1 alone does not allow to draw a conclusion on the direction of the effect, since the confidence set for the ATE contains zero.

Table 12. Confidence sets for parameters

Parameters	Biprobit estimates	95% conf. LB	95% conf. UB
$\mathbb{E}[Y_0]$	0.7390	0.6026	0.7724
$\mathbb{E}[Y_1]$	0.0724	0.0393	0.9813
ATE	-0.6666	-0.7331	0.3787

conf.: confidence; LB: lower bound; UB: upper bound.

Confidence Bounds on ATE under Assumptions 1 and 4. As in the previous empirical illustration, we tighten the bounds on the ATE by exploiting the additional information in the monotonicity assumption (Proposition 3). Table 13 displays the results. The confidence region for the ATE is $[-0.0512, 0.3549] \cup [-0.7158, 0.0512] = [-0.7158, 0.3549]$. This confidence region is slightly tighter, but remains wide. Unlike the previous application, the sign of the ATE is not identified under Assumptions 1 and 4.

Table 13. Confidence sets for parameters

Parameters	95% conf. LB $Y_1 \geqslant Y_0$	95% conf. UB $Y_1 \geqslant Y_0$	95% conf. LB $Y_1 \leqslant Y_0$	95% conf. UB $Y_1 \leqslant Y_0$
$\mathbb{E}[Y_0]$	0.6821	0.9763	0.0494	0.7334
$\mathbb{E}[Y_1]$	0.6214	0.7333	0.6822	0.7651
ATE	-0.0512	0.3549	-0.7158	0.0512

conf.: confidence; LB: lower bound; UB: upper bound.

8. Conclusion

This paper develops a falsification test for the identifying assumptions in bivariate probit models. We derive sharp testable equalities for the model assumptions, but they are difficult to implement in practice. We then propose a testing procedure based on non-sharp inequalities, which can be expressed in the form of conditional moment inequalities. This provides a novel test easily implemented using the Stata package developed by Chernozhukov et al. (2015). The test's small-sample size and power performance are studied in simulations. We find that the test tends to over-reject in small samples, but adequately controls size in large samples. We discuss partial identification approaches that can be used by researchers to relax the assumptions when they are rejected. Finally, we provide two empirical illustrations in which the bivariate model, despite its nice feature that leads to point-identication of model parameters, could be restrictive. Our proposed procedure could serve as a screening test for the validity of the bivariate probit specification.

While the test we develop in this paper can easily handle discrete covariates, it is difficult to include continuous covariates in the implementation of the procedure. We believe this question could be further explored in future research.

Appendix A. Proof of Propositions 1 and 2

A.1. **Proof of Proposition 1.**

Proof. Assume that Z is continuous and $Cov\left(\Phi^{-1}\left(\mathbb{P}(D=1|Z)\right), Z\right) \neq 0$. Let the parameters α, β, ρ, γ, δ be defined and identified as described in Subsection 3.1. Then $\delta \neq 0$ by definition. Suppose now that equalities (3.5), (3.6) and (3.7) hold. We need to show that there exists a vector $\left(\tilde{U}, \tilde{V}, \tilde{Y}, \tilde{D}, Z\right)$ such that $Z \perp\!\!\!\perp \left(\tilde{U}, \tilde{V}\right)$, $\begin{pmatrix}\tilde{U}\\ \tilde{V}\end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$, where $\mu = \begin{pmatrix}0\\0\end{pmatrix}$, and $\Sigma = \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$, and the joint distribution of $\left(\tilde{Y}, \tilde{D}, Z\right)$ is the same as that of (Y, D, Z). Define the joint density of $\left(\tilde{U}, \tilde{V}\right)$ conditional on Z as

$$f_{(\tilde{U},\tilde{V}|Z)}(u, v|z) = \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{u-\rho v}{\sqrt{1-\rho^2}}\right)\phi(v),$$

where $\phi(t) = \exp(-t^2/2)$, and define

$$\begin{cases} \tilde{Y} &= \mathbb{1}\{\beta + \alpha\tilde{D} - \tilde{U} \geqslant 0\} \\ \tilde{D} &= \mathbb{1}\{\gamma + \delta Z - \tilde{V} \geqslant 0\} \end{cases}$$

It is clear that $Z \perp\!\!\!\perp \left(\tilde{U}, \tilde{V}\right)$ since the conditional density of $\left(\tilde{U}, \tilde{V}\right)|Z = z$ does not depend on z. The joint density of $\left(\tilde{U}, \tilde{V}\right)$ is equal to

$$f_{(\tilde{U},\tilde{V})}(u, v) = \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(u\ v)\Sigma^{-1}(u\ v)'\right\}.$$

Therefore, $\begin{pmatrix}\tilde{U}\\ \tilde{V}\end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$. It remains to show that

$$\mathbb{P}\left(\tilde{Y} = y, \tilde{D} = d|Z = z\right) = \mathbb{P}\left(Y = y, D = d|Z = z\right)$$

for all y, d, and z. We have

$$
\begin{aligned}
\mathbb{P}\left(\tilde{Y}=1,\tilde{D}=1|Z=z\right) &= \mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha,\tilde{V}\leqslant\gamma+\delta z\right),\\
&= \mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha|\tilde{V}\leqslant\gamma+\delta z\right)\mathbb{P}\left(\tilde{V}\leqslant\gamma+\delta z\right),\\
&= \mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha|\Phi\left(\tilde{V}\right)\leqslant\Phi\left(\gamma+\delta z\right)\right)\Phi\left(\gamma+\delta z\right),\\
&= \mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha|\Phi\left(\tilde{V}\right)\leqslant P(z)\right)P(z),\quad\text{from Equality (3.7)},\\
&= \int_0^{P(z)}\mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha|\Phi\left(\tilde{V}\right)=p\right)\frac{f_{\Phi(\tilde{V})}(p)}{\mathbb{P}\left(\Phi\left(\tilde{V}\right)\leqslant P(z)\right)}dp*P(z),\\
&= \int_0^{P(z)}\mathbb{P}\left(\tilde{U}\leqslant\beta+\alpha|\tilde{V}=\Phi^{-1}(p)\right)dp,\ \text{as } \Phi\left(\tilde{V}\right)\sim\mathcal{U}_{[0,1]},\\
&= \int_0^{P(z)}\Phi\left(\frac{\beta+\alpha-\rho\Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)dp,\quad\text{as }\ \mathbb{P}\left(\tilde{U}\leqslant u|\tilde{V}=v\right)=\Phi\left(\frac{u-\rho v}{\sqrt{1-\rho^2}}\right),\\
&= \int_0^{P(z)}\frac{\partial\mathbb{E}[YD|P(Z)=p]}{\partial p}dp,\quad\text{from Equality (3.5)},\\
&= \mathbb{E}\left[YD|P(Z)=P(z)\right]-\mathbb{E}\left[YD|P(Z)=0\right],\\
&= \mathbb{E}\left[YD|P(Z)=P(z)\right],\quad\text{since } D=0 \text{ when } P(Z)=0,\\
&= \mathbb{E}\left[YD|Z=z\right],\quad\text{since } \{P(Z)=P(z)\}=\{Z=z\},\\
&= \mathbb{P}\left(Y=1,D=1|Z=z\right).
\end{aligned}
$$

We also have

$$
\begin{aligned}
\mathbb{P}\left(\tilde{Y}=1,\tilde{D}=0|Z=z\right) &= \mathbb{P}\left(\tilde{U}\leqslant\beta,\tilde{V}>\gamma+\delta z\right),\\
&= \mathbb{P}\left(\tilde{U}\leqslant\beta|\Phi\left(\tilde{V}\right)>\Phi\left(\gamma+\delta z\right)\right)\left(1-\Phi\left(\gamma+\delta z\right)\right),\\
&= \mathbb{P}\left(\tilde{U}\leqslant\beta|\Phi\left(\tilde{V}\right)>P(z)\right)\left(1-P(z)\right),\quad\text{from Equality (3.7)},\\
&= \int_{P(z)}^1\mathbb{P}\left(\tilde{U}\leqslant\beta|\Phi\left(\tilde{V}\right)=p\right)\frac{f_{\Phi(\tilde{V})}(p)}{\mathbb{P}\left(\Phi\left(\tilde{V}\right)>P(z)\right)}dp*\left(1-P(z)\right),\\
&= \int_{P(z)}^1\mathbb{P}\left(\tilde{U}\leqslant\beta|\tilde{V}=\Phi^{-1}(p)\right)dp,\\
&= \int_{P(z)}^1\Phi\left(\frac{\beta-\rho\Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)dp,\\
&= \int_{P(z)}^1-\frac{\partial\mathbb{E}\left[Y(1-D)|P(Z)=p\right]}{\partial p}dp,\quad\text{from Equality (3.6)},\\
&= -\mathbb{E}\left[Y(1-D)|P(Z)=1\right]+\mathbb{E}\left[Y(1-D)|P(Z)=P(z)\right],\\
&= \mathbb{E}\left[Y(1-D)|P(Z)=P(z)\right],\quad\text{since } D=1 \text{ when } P(Z)=1,\\
&= \mathbb{E}\left[Y(1-D)|Z=z\right],\quad\text{since } \{P(Z)=P(z)\}=\{Z=z\},\\
&= \mathbb{P}\left(Y=1,D=0|Z=z\right).
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
\mathbb{P}\left(\tilde{Y}=0,\tilde{D}=1|Z=z\right) &= \mathbb{P}\left(\tilde{U}>\beta+\alpha,\tilde{V}\leqslant\gamma+\delta z\right),\\
&= \mathbb{P}\left(\tilde{U}>\beta+\alpha|\Phi\left(\tilde{V}\right)\leqslant P(z)\right)P(z), \quad \text{from Equality (3.7)},\\
&= \int_0^{P(z)}\mathbb{P}\left(\tilde{U}>\beta+\alpha|\tilde{V}=\Phi^{-1}(p)\right)dp,\\
&= \int_0^{P(z)} 1-\Phi\left(\frac{\beta+\alpha-\rho\Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)dp,\\
&= \int_0^{P(z)} 1-\frac{\partial\mathbb{E}\left[YD|P(Z)=p\right]}{\partial p}dp, \quad \text{from Equality (3.5)},\\
&= \int_0^{P(z)} \frac{\partial\mathbb{E}\left[D-YD|P(Z)=p\right]}{\partial p}dp, \quad \text{since } \mathbb{E}\left[D|P(Z)=p\right]=p,\\
&= \mathbb{E}\left[(1-Y)D|P(Z)=P(z)\right]-\mathbb{E}\left[(1-Y)D|P(Z)=0\right],\\
&= \mathbb{P}\left(Y=0,D=1|Z=z\right).
\end{aligned}
$$

□

A.2. **Proof of Proposition 2.**

Proof. We begin with **inequality (3.8)**. We have

$$
\begin{aligned}
\mathbb{E}[YD|Z=z] &= \mathbb{P}(Y=1,D=1|Z=z),\\
&= \mathbb{P}(U\leqslant\alpha+\beta,V\leqslant\delta z|Z=z),\\
&\leqslant \mathbb{P}(U\leqslant\alpha+\beta|Z=z),\\
&= \mathbb{P}(U\leqslant\alpha+\beta)=\Phi(\alpha+\beta),
\end{aligned}
$$

where the last two equalities follow from Assumptions 1 and 2, respectively. Thus, by taking the supremum of the left-hand side over z, we have $\sup_z \mathbb{E}[YD|Z=z]\leqslant\Phi(\alpha+\beta)$.

Inequality (3.9)

Similar to the previous reasoning, we have

$$
\begin{aligned}
\mathbb{E}[Y(1-D)|Z=z] &= \mathbb{P}(Y=1,D=0|Z=z),\\
&= \mathbb{P}(U\leqslant\beta,V>\delta z|Z=z),\\
&\leqslant \mathbb{P}(U\leqslant\beta)=\Phi(\beta).
\end{aligned}
$$

Thus $\sup_z \mathbb{E}[Y(1-D)|Z=z]\leqslant\Phi(\beta)$.

Inequality (3.10)

$$\begin{aligned}\mathbb{E}[(1-Y)D|Z=z] &= \mathbb{P}(Y=0, D=1|Z=z),\\ &= \mathbb{P}(U>\alpha+\beta, V\leqslant \delta z),\\ &\leqslant \mathbb{P}(U>\alpha+\beta) = 1-\Phi(\alpha+\beta).\end{aligned}$$

Thus $\sup_z \mathbb{E}[(1-Y)D|Z=z] \leqslant 1-\Phi(\alpha+\beta)$.

Inequality (3.11)

$$\begin{aligned}\mathbb{E}[(1-Y)(1-D)|Z=z] &= \mathbb{P}(Y=0, D=0|Z=z),\\ &= \mathbb{P}(U>\beta, V>\delta z),\\ &\leqslant \mathbb{P}(U>\beta) = 1-\Phi(\beta).\end{aligned}$$

Thus $\sup_z \mathbb{E}[(1-Y)(1-D)|Z=z] \leqslant 1-\Phi(\beta)$.

Inequality (3.12)

$$\begin{aligned}\mathbb{P}(U\leqslant\alpha+\beta, U\leqslant\beta) &= \mathbb{P}(U\leqslant\alpha+\beta, U\leqslant\beta, D=1|Z=z)+\mathbb{P}(U\leqslant\alpha+\beta, U\leqslant\beta, D=0|Z=z),\\ &\leqslant \mathbb{P}(U\leqslant\alpha+\beta, D=1|Z=z)+\mathbb{P}(U\leqslant\beta, D=0|Z=z),\\ &= \mathbb{P}(Y=1, D=1|Z=z)+\mathbb{P}(Y=1, D=0|Z=z),\end{aligned}$$

where the first equality holds from Assumption 1 and the law of total probability. Therefore

$$\Phi(\min(\alpha+\beta,\beta)) \leqslant \inf_z \mathbb{E}[YD+Y(1-D)|Z=z]$$

since $\mathbb{P}(U\leqslant\alpha+\beta, U\leqslant\beta) = \Phi(\min(\alpha+\beta,\beta))$ under Assumption 2.

Inequality (3.13)

$$\begin{aligned}\mathbb{P}(U>\alpha+\beta, U>\beta) &= \mathbb{P}(U>\alpha+\beta, U>\beta, D=1|Z=z)+\mathbb{P}(U>\alpha+\beta, U>\beta, D=0|Z=z),\\ &\leqslant \mathbb{P}(U>\alpha+\beta, D=1|Z=z)+\mathbb{P}(U>\beta, D=0|Z=z),\\ &= \mathbb{P}(Y=0, D=1|Z=z)+\mathbb{P}(Y=0, D=0|Z=z).\end{aligned}$$

Thus $1-\Phi(\max(\alpha+\beta,\beta)) \leqslant \inf_z \mathbb{E}[(1-Y)D+(1-Y)(1-D)|Z=z]$. □

Appendix B. Additional remarks

Remark 3. *There exist two extra inequalities that are redundant. They are:*

$$\Phi(\beta)-\Phi(\min(\beta+\alpha,\beta)) \leqslant \inf_z \mathbb{E}[(1-Y)D+Y(1-D)|Z=z], \quad \text{(B.1)}$$

$$\Phi(\beta+\alpha)-\Phi(\min(\beta+\alpha,\beta)) \leqslant \inf_z \mathbb{E}[YD+(1-Y)(1-D)|Z=z]. \quad \text{(B.2)}$$

Proof. **Inequality (B.1)**

$$\begin{aligned}\mathbb{P}(U > \alpha+\beta, U \leqslant \beta) &= \mathbb{P}U > \alpha+\beta, U \leqslant \beta, D=1|Z=z) + \mathbb{P}(U > \alpha+\beta, U \leqslant \beta, D=0|Z=z), \\ &\leqslant \mathbb{P}(U > \alpha+\beta, D=1|Z=z) + \mathbb{P}(U \leqslant \beta, D=0|Z=z), \\ &= \mathbb{P}(Y=0, D=1|Z=z) + \mathbb{P}(Y=1, D=0|Z=z).\end{aligned}$$

Thus $\Phi(\beta) - \Phi(\min(\beta+\alpha, \beta)) \leqslant \inf_z \mathbb{E}[(1-Y)D + Y(1-D)|Z=z]$.

Inequality (B.2)

$$\begin{aligned}\mathbb{P}(U \leqslant \alpha+\beta, U > \beta) &= \mathbb{P}(U \leqslant \alpha+\beta, U > \beta, D=1|Z=z) + \mathbb{P}(U \leqslant \alpha+\beta, U > \beta, D=0|Z=z), \\ &\leqslant \mathbb{P}(U \leqslant \alpha+\beta, D=1|Z=z) + \mathbb{P}(U > \beta, D=0|Z=z), \\ &= \mathbb{P}(Y=1, D=1|Z=z) + \mathbb{P}(Y=0, D=0|Z=z).\end{aligned}$$

Thus $\Phi(\alpha+\beta) - \Phi(\min(\alpha+\beta, \beta)) \leqslant \inf_z \mathbb{E}[YD + (1-Y)(1-D)|Z=z]$.

We can get(B.1) by combining (3.8) and (3.11).

From (3.11) we have:

$\sup_z \mathbb{E}[1-D-Y+YD|Z=z] = \sup_z \mathbb{E}[(1-Y)(1-D)|Z=z] \leqslant 1-\Phi(\beta)$.

Thus, $\sup_z \mathbb{E}[-D-Y+YD|Z=z] \leqslant -\Phi(\beta)$ or $\inf_z \mathbb{E}[D+Y-YD|Z=z] \geqslant \Phi(\beta)$.

From (3.8) we know that $\sup_z \mathbb{E}[YD|Z=z] \leqslant \Phi(\beta+\alpha)$.

We have

$$\begin{aligned}\inf_z \mathbb{E}[(1-Y)D + Y(1-D)|Z=z] &= \inf_z \mathbb{E}[(1-Y)D + Y - YD|Z=z] \\ &\geqslant \inf_z \mathbb{E}[(1-Y)D + Y|Z=z] + \inf_z \mathbb{E}[-YD|Z=z] \\ &= \inf_z \mathbb{E}[D - YD + Y|Z=z] - \sup_z \mathbb{E}[YD|Z=z] \\ &\geqslant \Phi(\beta) - \Phi(\beta+\alpha)\end{aligned}$$

where in the first equality we just expand the interior of the expectation, in the first inequality and second equality we use the properties of inf and sup. In the third inequality we use (3.11) and (3.8).

Furthermore, we have $\inf_z \mathbb{E}[(1-Y)D + Y(1-D)|Z=z] \geqslant 0 = \Phi(\beta) - \Phi(\beta)$. Hence, (B.1) holds whenever (3.11) and (3.8) hold.

Similarly (B.2) can be obtained by combining (3.9) and (3.10). □

Remark 4. *Inequalities (3.8)-(3.13) are implied by the sharp equalities derived in Subsection 3.2.*

Proof. Consider inequality (3.8) for example. We have

$$\frac{\mathbb{E}\left[YD|P(Z)=p\right]-\mathbb{E}\left[YD|P(Z)=p'\right]}{p-p'}=\mathbb{P}\left(U\leqslant\beta+\alpha|\Phi^{-1}(p')<V\leqslant\Phi^{-1}(p)\right).$$

Set $p'=0$, this implies that, $\mathbb{E}\left[YD|P(Z)=0\right]=0$ as $D=0$ when $P(Z)=0$.

Then

$$\begin{aligned}\mathbb{E}\left[YD|P(Z)=p\right] &= p\mathbb{P}\left(U\leqslant\beta+\alpha|V\leqslant\Phi^{-1}(p)\right),\\ &= p\frac{\mathbb{P}\left(U\leqslant\beta+\alpha,\Phi(V)\leqslant p\right)}{\mathbb{P}\left(\Phi(V)\leqslant p\right)},\\ &= p\frac{\mathbb{P}\left(U\leqslant\beta+\alpha,\Phi(V)\leqslant p\right)}{p},\\ &= \mathbb{P}\left(U\leqslant\beta+\alpha,\Phi(V)\leqslant p\right),\\ &\leqslant \mathbb{P}\left(U\leqslant\beta+\alpha\right)=\Phi(\beta+\alpha).\end{aligned}$$

Now by index sufficiency (given $P(z)$ is known), we have: $\mathbb{E}\left[YD|P(Z)=P(z)\right]=\mathbb{E}\left[YD|Z=z\right]$. Thus, $\mathbb{E}\left[YD|Z=z\right]\leqslant\Phi(\beta+\alpha)$

Similar reasoning can be done for the other inequalities. □

Appendix C. Validity of the plug-in approach

The proofs for the validity of the plug-in approach for (3.8)-(3.11) are similar. So, we only show the proof for (3.8).

Proof. Define $\mathbb{E}[YD|Z=z]=g(z)$. Let $\hat{g}(z)$ be a local nonparametric estimator for $g(z)$ with convergence rate $\sqrt{nh}$ where $h\longrightarrow 0$ as $n\longrightarrow\infty$, such that:

$$\sqrt{nh}(\hat{g}(z)-g(z))\xrightarrow{d}N(0,V(z)).$$

If $\alpha,\ \beta$ are known, naturally we would have:

$$\sqrt{nh}[(\hat{g}(z)-\Phi(\alpha+\beta))-(g(z)-\Phi(\alpha+\beta))]=\sqrt{nh}[\hat{g}(z)-g(z)]\xrightarrow{d}N(0,V(z))$$

In practice, we replace $\alpha,\ \beta$ by their estimators $\hat{\alpha},\ \hat{\beta}$. We need to show that the above asymptotic distribution is not affected by this plug-in approach. We have

$$\sqrt{nh}[(\hat{g}(z)-\Phi(\hat{\alpha}+\hat{\beta}))-(g(z)-\Phi(\alpha+\beta))]=\sqrt{nh}[\hat{g}(z)-g(z)]-\sqrt{h}[\sqrt{n}(\Phi(\hat{\alpha}+\hat{\beta})-\Phi(\alpha+\beta))]$$

We know that the first part $\sqrt{nh}[\hat{g}(z)-g(z)]\xrightarrow{d}N(0,V(z))$. Since $\alpha,\ \beta$ are estimated by MLE, we know that: $\Phi(\hat{\alpha}+\hat{\beta})-\Phi(\alpha+\beta)=O_p(n^{-\frac{1}{2}})$. Then, $\sqrt{n}(\Phi(\hat{\alpha}+\hat{\beta})-\Phi(\alpha+\beta))]=O_p(1)$, and since $h=o(1)$, it is clear that $\sqrt{h}=o_p(1)$. Hence, $\sqrt{h}[\sqrt{n}(\Phi(\hat{\alpha}+\hat{\beta})-$

$\Phi(\alpha+\beta))] = o_p(1) \times O_p(1) = o_p(1)$. Therefore, by the asymptotic equivalence lemma, $\sqrt{nh}[(\hat{g}(z) - \Phi(\hat{\alpha}+\hat{\beta})) - (g(z) - \Phi(\alpha+\beta))] \xrightarrow{d} N(0, V(z))$. □

We are now going to show that the plug-in approach works for inequality (3.12). Similar reasoning holds for inequality (3.13).

Proof. Define $\theta(z) = \mathbb{E}[Y|Z=z]$ and $\hat{\theta}(z)$ is a nonparametric estimator for $\theta(z)$ such that

$$\sqrt{nh}\left\{\hat{\theta}(z) - \theta(z)\right\} \longrightarrow N(0, \Omega_1(z))$$

where $h \longrightarrow 0$ as $n \longrightarrow \infty$. We want to show that

$$\sqrt{nh}\left\{\left(\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \hat{\theta}(z)\right) - (\Phi(\min(\beta+\alpha, \beta)) - \theta(z))\right\} \longrightarrow N(0, \Omega_1(z)).$$

It suffices to show that $\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} = o_p(1)$, from the asymptotic equivalence lemma. For any $\delta > 0$, we have:

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} > \delta\right) \\
&\quad = \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} > \delta, \hat{\alpha} \geqslant 0, \alpha \geqslant 0\right) \\
&\qquad + \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} > \delta, \hat{\alpha} \leqslant 0, \alpha \leqslant 0\right) \\
&\qquad + \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} > \delta, \hat{\alpha} > 0, \alpha < 0\right) \\
&\qquad + \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\min(\hat{\beta}+\hat{\alpha}, \hat{\beta})) - \Phi(\min(\beta+\alpha, \beta))\right\} > \delta, \hat{\alpha} < 0, \alpha > 0\right), \\
&\quad \leqslant \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\hat{\beta}) - \Phi(\beta)\right\} > \delta\right) \\
&\qquad + \mathbb{P}\left(\sqrt{nh}\left\{\Phi(\hat{\beta}+\hat{\alpha}) - \Phi(\beta+\alpha)\right\} > \delta\right) \\
&\qquad + \mathbb{P}(\hat{\alpha} > 0, \alpha < 0) \\
&\qquad + \mathbb{P}(\hat{\alpha} < 0, \alpha > 0)
\end{aligned}$$

By the delta method $\sqrt{n}\left\{\Phi(\hat{\beta}) - \Phi(\beta)\right\} = O_p(1)$ and $\sqrt{n}\left\{\Phi(\hat{\beta}+\hat{\alpha}) - \Phi(\beta+\alpha)\right\} = O_p(1)$. Since $\sqrt{h} = o(1)$, we conclude using the product rule that $\sqrt{nh}\left\{\Phi(\hat{\beta}) - \Phi(\beta)\right\} = o_p(1)$ and $\sqrt{nh}\left\{\Phi(\hat{\beta}+\hat{\alpha}) - \Phi(\beta+\alpha)\right\} = o_p(1)$. Therefore, $\mathbb{P}\left(\sqrt{nh}\left\{\Phi(\hat{\beta}) - \Phi(\beta)\right\} > \delta\right) \longrightarrow 0$ as $n \longrightarrow \infty$, and $\mathbb{P}\left(\sqrt{nh}\left\{\Phi(\hat{\beta}+\hat{\alpha}) - \Phi(\beta+\alpha)\right\} > \delta\right) \longrightarrow 0$ as $n \longrightarrow \infty$. It remains to show that $\mathbb{P}(\hat{\alpha} > 0, \alpha < 0)$ and $\mathbb{P}(\hat{\alpha} < 0, \alpha > 0)$ go to zero as n goes to infinity. We have

$$\begin{aligned}
\hat{\alpha} > 0, \alpha < 0 \quad &\Longrightarrow \quad \exists\ \epsilon_1,\ \epsilon_2 :\ \hat{\alpha} \geqslant \epsilon_1 > 0,\ \alpha \leqslant \epsilon_2 < 0, \\
&\Longrightarrow \quad \hat{\alpha} - \alpha \geqslant \epsilon_1 - \epsilon_2 > 0.
\end{aligned}$$

Then $\mathbb{P}\left(\hat{\alpha}>0, \alpha<0\right) \leqslant P\left(\hat{\alpha}-\alpha \geqslant \epsilon_1-\epsilon_2\right) \longrightarrow 0$ as $n \longrightarrow \infty$, because $\hat{\alpha}$ is a consistent estimator for α. By a similar argument, $\mathbb{P}\left(\hat{\alpha}<0, \alpha>0\right) \longrightarrow 0$ as $n \longrightarrow \infty$. □

Appendix D. Further Extensions

D.1. **Adding exogenous covariates.** Suppose we have the following specification:

$$\left\{\begin{array}{lll} Y & = & \mathbb{1}\{\alpha D+\beta' X-U \geqslant 0\} \\ D & = & \mathbb{1}\{\delta Z+\lambda' X-V \geqslant 0\} \end{array}\right. \tag{D.1}$$

The testable implications in Proposition 2 become

$$\begin{array}{lll} \sup_z \mathbb{E}[YD|Z=z,X=x] & \leqslant & \Phi(\alpha+\beta' x), \\ \sup_z \mathbb{E}[Y(1-D)|Z=z,X=x] & \leqslant & \Phi(\beta' x), \\ \sup_z \mathbb{E}[(1-Y)D|Z=z,X=x] & \leqslant & 1-\Phi(\alpha+\beta' x), \\ \sup_z \mathbb{E}[(1-Y)(1-D)|Z=z,X=x] & \leqslant & 1-\Phi(\beta' x), \\ \Phi(\min(\alpha+\beta' x,\beta' x)) & \leqslant & \inf_z \mathbb{E}[YD+Y(1-D)|Z=z,X=x], \\ 1-\Phi(\max(\alpha+\beta' x,\beta' x)) & \leqslant & \inf_z \mathbb{E}[(1-Y)D+(1-Y)(1-D)|Z=z,X=x]. \end{array} \tag{D.2}$$

D.2. **Extension to generalized bivariate models.** Suppose we still have the model (2.1), but instead of Assumption 2 we have:

Assumption 5. *F_U and F_V are known marginal distributions of U and V, respectively, that are strictly increasing, are absolutely continuous with respect to Lebesgue measure, and such that $\mathbb{E}[U]=\mathbb{E}[V]=0$ and $Var(U)=Var(V)=1$.*

Assumption 6. *$(U,V) \sim F_{UV}(U,V)=C(F_U(U),F_V(V);\rho)$ where $C(.,.;\rho)$ is a copula known up to scalar parameter $\rho \in \Omega$ such that $C:(0,1)^2 \rightarrow (0,1)$ is twice differentiable in its arguments and ρ.*

We have $P(z)=F_V(\gamma+\delta z)$ and

$$\begin{array}{rcl} \mathbb{E}\left[YD|P(Z)=p\right] & = & \mathbb{P}(Y=1,D=1|P(Z)=p), \\ & = & \mathbb{P}\left(U \leqslant \beta+\alpha, V \leqslant \gamma+\delta Z|P(Z)=p\right), \\ & = & \mathbb{P}\left(F_U(U) \leqslant F_U(\beta+\alpha), F_V(V) \leqslant F_V(\gamma+\delta Z)|P(Z)=p\right), \\ & = & \mathbb{P}\left(F_U(U) \leqslant F_U(\beta+\alpha), F_V(V) \leqslant p\right), \\ & = & C(F_U(\beta+\alpha),p;\rho), \end{array}$$

where the first equality holds because Y, and D are binary, the second holds from the definition of the model, the third uses the fact that the function F_U and F_V are increasing, and the fourth holds under Assumption 1.

For $p > p'$, we have

$$\frac{\mathbb{E}\left[YD|P(Z)=p\right]-\mathbb{E}\left[YD|P(Z)=p'\right]}{p-p'} = \frac{C(F_U(\beta+\alpha),p;\rho)-C(F_U(\beta+\alpha),p';\rho)}{p-p'}.$$

By taking the limit when p goes to p', we obtain

$$\frac{\partial\mathbb{E}\left[YD|P(Z)=p\right]}{\partial p} = \frac{\partial C(F_U(\beta+\alpha),p;\rho)}{\partial p}.$$

Similarly, we have

$$-\frac{\partial\mathbb{E}\left[Y(1-D)|P(Z)=p\right]}{\partial p} = \frac{\partial C(F_U(\beta),p;\rho)}{\partial p}.$$

We can repeat the procedure in appendix A to get the following inequalities relying on the fact that the copula is known.

$$\begin{aligned}
\mathbb{E}[YD|Z=z] &= \mathbb{P}(Y=1,D=1|Z=z),\\
&= \mathbb{P}(U\leqslant\beta+\alpha,V\leqslant\gamma+\delta z|Z=z),\\
&\leqslant \mathbb{P}(U\leqslant\beta+\alpha|Z=z),\\
&= \mathbb{P}(U\leqslant\beta+\alpha)=F_U(\beta+\alpha)
\end{aligned}$$

Thus $\sup_z \mathbb{E}[YD|Z=z]\leqslant F_U(\beta+\alpha)$.

$$\begin{aligned}
\mathbb{E}[Y(1-D)|Z=z] &= \mathbb{P}(Y=1,D=0|Z=z),\\
&= \mathbb{P}(U\leqslant\beta,V>\gamma+\delta z|Z=z),\\
&\leqslant \mathbb{P}(U\leqslant\beta)=F_U(\beta).
\end{aligned}$$

Thus $\sup_z \mathbb{E}[Y(1-D)|Z=z]\leqslant F_U(\beta)$

$$\begin{aligned}
\mathbb{E}[(1-Y)D|Z=z] &= \mathbb{P}(Y=0,D=1|Z=z),\\
&= \mathbb{P}(U>\beta+\alpha,V\leqslant\gamma+\delta z),\\
&\leqslant P(U>\beta+\alpha)=1-F_U(\beta+\alpha).
\end{aligned}$$

Thus $\sup_z \mathbb{E}[(1-Y)D|Z=z]\leqslant 1-F_U(\beta+\alpha)$.

$$\begin{aligned}\mathbb{E}[(1-Y)(1-D)|Z=z] &= \mathbb{P}(Y=0,D=0|Z=z),\\ &= \mathbb{P}(U>\beta,V>\gamma+\delta z),\\ &\leqslant \mathbb{P}(U>\beta)=1-F_U(\beta).\end{aligned}$$

Thus $\sup_z \mathbb{E}[(1-Y)(1-D)|Z=z] \leqslant 1-F_U(\beta)$.

$$\begin{aligned}\mathbb{P}(U\leqslant\beta+\alpha,U\leqslant\beta) &= \mathbb{P}(U\leqslant\beta+\alpha,U\leqslant\beta,D=1|Z=z)+\mathbb{P}(U\leqslant\beta+\alpha,U\leqslant\beta,D=0|Z=z),\\ &\leqslant \mathbb{P}(U\leqslant\beta+\alpha,D=1|Z=z)+\mathbb{P}(U\leqslant\beta,D=0|Z=z),\\ &= \mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=1,D=0|Z=z).\end{aligned}$$

Thus $F_U(\min(\beta+\alpha,\beta)) \leqslant \inf_z \mathbb{E}[Y|Z=z]$.

$$\begin{aligned}\mathbb{P}(U>\beta+\alpha,U>\beta) &= \mathbb{P}(U>\beta+\alpha,U>\beta,D=1|Z=z)+\mathbb{P}(U>\beta+\alpha,U>\beta,D=0|Z=z),\\ &\leqslant \mathbb{P}(U>\beta+\alpha,D=1|Z=z)+\mathbb{P}(U>\beta,D=0|Z=z),\\ &= \mathbb{P}(Y=0,D=1|Z=z)+\mathbb{P}(Y=0,D=0|Z=z).\end{aligned}$$

Thus $1-F_U(\max(\beta+\alpha,\beta)) \leqslant \inf_z \mathbb{E}[1-Y|Z=z]$

Now that we have these more general inequalities, building upon Han and Lee (2019) we can extend our test to situations where the marginal distributions are known but the copula structure has certain conditions.

Appendix E. Additional results for the application

E.1. **Summary Statistics.** Tables 14 and 15 include additional summary statistics for the empirical examples, respectively.

Table 14. Summary Statistics for empirical example 1

	Total	
Observations	7,555	
Doctor visit	0.1820	(0.3859)
Private insurance	0.6567	(0.4749)
Nb employees (standardized)	-0.0054	(0.9970)

Average and standard deviation (in parentheses)

Table 15. Summary Statistics for empirical example 2

	Total	
Observations	1,659	
Adoption	0.7077	(0.4550)
Rent	0.0934	(0.2911)
Mig-ag-ratio	6.7027	(11.5441)

Average and standard deviation (in parentheses)

E.2. **Implementation: Chernozhukov, Lee, and Rosen (2013) conditions and commands.**

E.2.1. *Algorithm.* Let β_n denotes the parameters estimated nonparametrically. K denotes the dimension of β_n and I_k denotes the K dimensional identity matrix. Let $p_n(z) = \frac{\partial\theta_n(z,\hat{\beta}_n)}{\partial\beta_n}$. Inference is conducted in the following way:

(1) Set $\tilde{\varepsilon}_n \equiv 1-0.1/log(n)$. Recall in our procedure $\mathcal{J}$: 1..j..6. Simulate a large number R of draws denoted $Z_1, ...Z_R$ from the K-variate standard normal distribution $N(0, I_K)$.
(2) Compute $\hat{\Omega}_n$, a consistent estimator for the large sample variance of $\sqrt{n}(\hat{\beta}_n - \beta_n)$.
(3) For each $z, j \in \mathcal{Z}, \mathcal{J}$, compute $\hat{g}(z,j) = p_n(z,j)'\hat{\Omega}_n^{1/2}$ and set $\hat{s}(z,j) = ||\hat{g}(z,j)||/\sqrt{n}$.
(4) Compute $k(\tilde{\varepsilon}_n) = \tilde{\varepsilon}_n$-quantile of $\{\sup_{z,j\in\mathcal{Z},\mathcal{J}}(\hat{g}(z,j)'Z_r/||\hat{g}(z,j)||)$ for $r = 1, ...R\}$ and $\widehat{(Z,J)}_n = \{z, j \in \mathcal{Z}, \mathcal{J} : \hat{\theta}_n(z,j) \geqslant \min_{z,j\in\mathcal{Z},\mathcal{J}}[\hat{\theta}_n(z,j) - k(\tilde{\varepsilon}_n)\hat{s}(z,j)] - 2k(\tilde{\varepsilon}_n)\}$
(5) Compute $k_{1-\alpha} = 1-\alpha$ quantile of $\{\sup_{z,j\in\widehat{(Z,J)}_n}(\hat{g}(z,j); Z_r/||\hat{g}(z,j))||$ for $r = 1, ..., R\}$ Then set: $\hat{\theta}_{1-\alpha} = \sup_{z,j\in\mathcal{Z},\mathcal{J}}[\hat{\theta}(z,j) - k_{1-\alpha}\hat{s}(z,j)]$.

E.2.2. *Commands.* Find below the code for the implementation of the test in Stata.

```
use ‘‘HLData.DTA’’, clear
biprobit (Doctorvisit=Privateins) (Privateins= Stdnbemp)
matrix define C=e(b)
gen Y=Doctorvisit
gen D=Privateins
gen ldepen1=Y*D−normal(C[1,1]+C[1,2])
gen ldepen2=Y*(1−D)−normal(C[1,2])
gen ldepen3=(1−Y)*D−1−normal(C[1,1]+C[1,2])
gen ldepen4=(1−Y)*(1−D)−1−normal(C[1,2])
```

```
gen ldepen5=normal(min(C[1,1]+C[1,2],C[1,2]))-Y
gen ldepen6=1-normal(max(C[1,1]+C[1,2],C[1,2]))-(1-Y)

*** Define the range for the instrument
gen Z=Stdnbemp
centile(Z), centile(1 99)
scalar LBZ=r(c_1)
scalar UBZ=r(c_2)
sum Z
gen RZ =LBZ + _n*(UBZ-LBZ)/200
replace RZ=. if _n>200

clrbound (ldepen1 Z RZ) (ldepen2 Z RZ) (ldepen3 Z RZ) (ldepen4 Z RZ)
(ldepen5 Z RZ) (ldepen6 Z RZ), low met(``local'') level(0.5 0.9 0.95 0.99)
norseed rnd(20000)
```

E.3. **Additional results for the first empirical illustration.** Table 16 below summarizes additional results for the bivariate probit specification in the estimation of the effect of insurance on doctor visits. In most cases, the bivariate probit model is rejected.

E.4. **Additional results for the second empirical illustration.** Table 17 below summarizes additional results for the bivariate probit specification in the estimation of the effect of land tenancy arrangements on the adoption of conservation practices. The bivariate probit model is rejected in all cases.

Appendix F. Proof of Proposition 3

We show the validity of the bounds in Proposition 3, and propose a joint distribution on (Y_0, Y_1) that achieves the lower bound on Y_1 and the upper bound on Y_0, and vice versa. First, we combine the condition $Y_1 \geqslant Y_0$ with Assumption 1. Results for condition $Y_1 \leqslant Y_0$ are obtained symmetrically by defining $\check{Y} = 1 - Y$, $\check{Y}_1 = 1 - Y_1$, and $\check{Y}_0 = 1 - Y_0$. Define

Table 16. The effect of insurance on doctor visits: test results for the bivariate probit specification by gender, race, region, and marriage status.

	W=0, M=0, =0, EM=0	W=0, M=0, S=0, EM=1	W=0, M=0, S=1, EM=1
Sig. levels			
10%	NR	NR	R
5%	NR	NR	R
1%	NR	NR	R
N	41	74	671
	W=0, M=1, S=1, EM=0	W=0, M=0, S=1, EM=0	W=0, M=1, S=1, EM=1
Sig. levels			
10%	R	R	R
5%	R	R	R
1%	R	R	R
N	293	582	59
	W=0, M=1, S=0, EM=1	W=1, M=0, S=1, EM=1	W=1, M=0, S=0, EM=1
Sig. levels			
10%	R	R	R
5%	R	R	R
1%	R	R	R
N	234	1,902	360
	W=1, M=1, S=0, EM=0	W=1, M=1, S=0, EM=1	W=1, M=1, S=1, EM=0
Sig. levels			
10%	R	R	R
5%	R	R	R
1%	R	R	R
N	62	339	561
	W=1, M=1, S=1, EM=1	W=1, M=0, S=0, EM=0	W=1, M=0, S=1, EM=0
Sig. levels			
10%	R	NR	R
5%	R	NR	R
1%	R	NR	R
N	1,924	44	393

The partitions do not add up to the total sample size because the group of white=0, male=1, SMSA=0, ever married=0 does not converge and has 16 observations. R=Reject, NR=No reject, Sig.=significance, N=sample size. W is indicator for white, M is indicator for male, EM is indicator for ever married person, and S is indicator for living in an SMSA. Rejection regions are corrected by the family-wise error rate.

the correspondence G between the unobservables (Y_0, Y_1) and the observables (Y, D):

$$G\{(0,0)\} = \{(0,0),(0,1)\},$$

$$G\{(0,1)\} = \{(0,0),(1,1)\},$$

$$G\{(1,1)\} = \{(1,0),(1,1)\}.$$

TABLE 17. Do farmers adopt fewer conservation practices on rented land?: test results for the bivariate probit specification by education, insurance and number of laborers.

	E=0, I=1, L=1	E=1, I=0, L=0
Sig. levels		
10%	R	R
5%	R	R
1%	R	R
N	38	397
	E=1, I=0, L=1	
Sig. levels		
10%	R	
5%	R	
1%	R	
N	207	

The partitions do not add up to the total sample size because E=0, I=0, L=0; E=1, I=1, L=1 and E=0, I=0, L=1 are excluded because there are no observations in that range; E=0, I=1, L=0 does not converge and has 15 observations; E=1, I=1, L=0 does not converge and has 8 observations. R=Reject, NR=No reject, Sig.=significance, N=sample size. E is indicator for having 13 years of education or more, I is indicator for having insurance, and L is indicator for having more than 2 workers in the farm. Rejection regions are corrected by the family-wise error rate.

Let $p_{ij} \equiv \mathbb{P}(Y_0 = i, Y_1 = j)$. By Galichon and Henry (2011, Theorem 1), we have that all restrictions on the unconditional joint distribution of (Y_0, Y_1) and the marginals of Y_0 and Y_1 are given by: for all $A \subset \{(0,0),(0,1),(1,0),(1,1)\}$,

$$\begin{aligned}\mathbb{P}((Y,D) \in A | Z = z) &\leqslant \mathbb{P}(G(Y_0,Y_1) \cap A \neq \varnothing | Z = z) \\ &= \mathbb{P}(G(Y_0,Y_1) \cap A \neq \varnothing), \ \forall\, z \in \mathcal{Z},\end{aligned}$$

that is:
for singletons,

$$\mathbb{P}(Y = 1, D = 1 | Z = z) \leqslant p_{01} + p_{11}, \tag{F.1}$$

$$\mathbb{P}(Y = 0, D = 1 | Z = z) \leqslant p_{00}, \tag{F.2}$$

$$\mathbb{P}(Y = 1, D = 0 | Z = z) \leqslant p_{11}, \tag{F.3}$$

$$\mathbb{P}(Y = 0, D = 0 | Z = z) \leqslant p_{00} + p_{01}; \tag{F.4}$$

for pairs,

$$\begin{aligned}
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=1,D=0|Z=z) &\leqslant p_{11}+p_{01}, &\text{(F.5)}\\
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=0,D=1|Z=z) &\leqslant 1, &\text{(F.6)}\\
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant 1, &\text{(F.7)}\\
\mathbb{P}(Y=1,D=0|Z=z)+\mathbb{P}(Y=0,D=1|Z=z) &\leqslant p_{00}+p_{11}, &\text{(F.8)}\\
\mathbb{P}(Y=1,D=0|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant 1, &\text{(F.9)}\\
\mathbb{P}(Y=0,D=1|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant p_{00}+p_{01}; &\text{(F.10)}
\end{aligned}$$

for triplets,

$$\begin{aligned}
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=1,D=0|Z=z)+\mathbb{P}(Y=0,D=1|Z=z) &\leqslant 1,\\
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=1,D=0|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant 1,\\
\mathbb{P}(Y=1,D=1|Z=z)+\mathbb{P}(Y=0,D=1|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant 1,\\
\mathbb{P}(Y=1,D=0|Z=z)+\mathbb{P}(Y=0,D=1|Z=z)+\mathbb{P}(Y=0,D=0|Z=z) &\leqslant 1.
\end{aligned}$$

Inequalities (F.6)-(F.7), (F.9), and all inequalities for the triplets are redundant. Inequality (F.5) implies (F.1), inequality (F.10) implies (F.4), and inequalities (F.2)-(F.3) imply (F.8). Hence, the set of non-redundant inequalities are inequalities (F.2), (F.3), (F.5) and (F.10). Since $p_{01}+p_{11}=\mathbb{E}[Y_1]$, (F.5) implies

$$\sup_z \mathbb{E}[Y|Z=z]\leqslant \mathbb{E}[Y_1].$$

Since $p_{00}=1-(p_{01}+p_{11})=1-\mathbb{E}[Y_1]$, inequality (F.2) imples

$$\mathbb{E}[Y_1]\leqslant \inf_z \mathbb{E}[1-(1-Y)D|Z=z].$$

Since $p_{00}+p_{01}=\mathbb{P}(Y_0=0)=1-\mathbb{E}[Y_0]$, inequality (F.10) implies

$$\mathbb{E}[Y_0]\leqslant \inf_z \mathbb{E}[Y|Z=z].$$

Finally, since $p_{11}=1-(p_{00}+p_{01})=\mathbb{E}[Y_0]$, inequality (F.3) implies

$$\sup_z \mathbb{E}[Y(1-D)|Z=z]\leqslant \mathbb{E}[Y_0].$$

Therefore, we obtain the bounds in Equation (6.1).

To show sharpness, suppose that $\sup_z \mathbb{E}[Y(1-D)|Z=z]\leqslant \inf_z \mathbb{E}[Y|Z=z]$, and $\sup_z \mathbb{E}[Y|Z=z]\leqslant \inf_z \mathbb{E}[1-(1-Y)D|Z=z]$. Define $\tilde{p}_{11}=\inf_z \mathbb{E}[Y|Z=z]$, $\tilde{p}_{00}=\inf_z \mathbb{E}[1-Y|Z=z]$, and $\tilde{p}_{01}=1-\inf_z \mathbb{E}[1-Y|Z=z]-\inf_z \mathbb{E}[Y|Z=z]$. Those quantities are well-defined probabilities. Indeed, $\tilde{p}_{11}+\tilde{p}_{01}+\tilde{p}_{00}=1$, $\tilde{p}_{11}\geqslant 0$, $\tilde{p}_{00}\geqslant 0$, and $\tilde{p}_{01}=$

$\sup_z \mathbb{E}[Y|Z=z] - \inf_z \mathbb{E}[Y|Z=z] \geqslant 0$. Since $\tilde{p}_{10}=0$ by definition, we have $Y_1 \geqslant Y_0$ a.s.. Define $\mathbb{P}(\tilde{Y}_1=i,\tilde{Y}_0=j|Z=z)=\tilde{p}_{ij}$. Then, Assumption 1 holds. Moreover, we have $\mathbb{E}[\tilde{Y}_1]=\tilde{p}_{11}+\tilde{p}_{01}=\sup_z \mathbb{E}[Y|Z=z]$, and $\mathbb{E}[\tilde{Y}_0]=\tilde{p}_{11}+\tilde{p}_{10}=\inf_z \mathbb{E}[Y|Z=z]$. Therefore, the lower bound for $\mathbb{E}[Y_1]$ in Equation (6.1) is achieved, while the upper bound for $\mathbb{E}[Y_0]$ is achieved by the same joint distribution.

On the other hand, define $\tilde{p}_{11}=\sup_z \mathbb{E}[Y(1-D)|Z=z]$, $\tilde{p}_{00}=\sup_z \mathbb{E}[(1-Y)D|Z=z]$, and $\tilde{p}_{01}=1-\sup_z \mathbb{E}[Y(1-D)|Z=z]-\sup_z \mathbb{E}[(1-Y)D|Z=z]$. As before, we can check that those quantities are well-defined probabilities. Indeed, $\tilde{p}_{11}+\tilde{p}_{01}+\tilde{p}_{00}=1$, $\tilde{p}_{11}\geqslant 0$, $\tilde{p}_{00}\geqslant 0$, and

$$\tilde{p}_{01}=\inf_z \mathbb{E}[1-(1-Y)D|Z=z]-\sup_z \mathbb{E}[Y(1-D)|Z=z] \geqslant \sup_z \mathbb{E}[Y|Z=z]-\inf_z \mathbb{E}[Y|Z=z] \geqslant 0,$$

where the second inequality follows from the conditions

$$\sup_z \mathbb{E}[Y(1-D)|Z=z] \leqslant \inf_z \mathbb{E}[Y|Z=z], \text{ and } \sup_z \mathbb{E}[Y|Z=z] \leqslant \inf_z \mathbb{E}[1-(1-Y)D|Z=z].$$

Since $\tilde{p}_{10}=0$ by definition, we have $Y_1 \geqslant Y_0$ a.s.. Define $\mathbb{P}(\tilde{Y}_1=i,\tilde{Y}_0=j|Z=z)=\tilde{p}_{ij}$. Then, Assumption 1 holds. Moreover, we have $\mathbb{E}[\tilde{Y}_1]=\tilde{p}_{11}+\tilde{p}_{01}=\inf_z \mathbb{E}[1-(1-Y)D|Z=z]$, and $\mathbb{E}[\tilde{Y}_0]=\tilde{p}_{11}+\tilde{p}_{10}=\sup_z \mathbb{E}[Y(1-D)|Z=z]$. Therefore, the upper bound for $\mathbb{E}[Y_1]$ in Equation (6.1) is achieved, while the lower bound for $\mathbb{E}[Y_0]$ is achieved by the same joint distribution.

Now, suppose $Y_1 \leqslant Y_0$ a.s., and define $\check{Y}=1-Y$, $\check{Y}_1=1-Y_1$, and $\check{Y}_0=1-Y_0$. Then, $\check{Y}_1 \geqslant \check{Y}_0$, and $\check{Y}=\check{Y}_1 D+\check{Y}_0(1-D)$. Using the previous results, we have the following sharp bounds for $\check{Y}_1$ and $\check{Y}_0$:

$$\sup_z \mathbb{E}[\check{Y}|Z=z] \leqslant \mathbb{E}[\check{Y}_1] \leqslant \inf_z \mathbb{E}[1-(1-\check{Y})D|Z=z],$$
$$\sup_z \mathbb{E}[\check{Y}(1-D)|Z=z] \leqslant \mathbb{E}[\check{Y}_0] \leqslant \inf_z \mathbb{E}[\check{Y}|Z=z].$$

After rewriting these inequalities in terms of Y, Y_1, and Y_0, we obtain the bounds in (6.2).

Finally, we propose the following joint distribution on $(\tilde{Y}_0, \tilde{Y}_1, D)$ given Z, which is compatible with the data (Y, D, Z), and satisfies Assumptions 1 and 4:

$$
\begin{aligned}
\mathbb{P}(\tilde{Y}_0 = 0, \tilde{Y}_1 = 0, D = 1|Z = z) &= \mathbb{P}(Y = 0, D = 1|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 0, \tilde{Y}_1 = 0, D = 0|Z = z) &= \tilde{p}_{00} - \mathbb{P}(Y = 0, D = 1|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 0, \tilde{Y}_1 = 1, D = 0|Z = z) &= \mathbb{P}(Y = 0, D = 0|Z = z) - \tilde{p}_{00} + \mathbb{P}(Y = 0, D = 1|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 1, \tilde{Y}_1 = 1, D = 0|Z = z) &= \mathbb{P}(Y = 1, D = 0|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 0, \tilde{Y}_1 = 1, D = 1|Z = z) &= \tilde{p}_{01} + \tilde{p}_{00} - \mathbb{P}(Y = 0|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 1, \tilde{Y}_1 = 1, D = 1|Z = z) &= \mathbb{P}(Y = 1, D = 1|Z = z) - \tilde{p}_{01} - \tilde{p}_{00} + \mathbb{P}(Y = 0|Z = z), \\
\mathbb{P}(\tilde{Y}_0 = 1, \tilde{Y}_1 = 0, D = 0|Z = z) &= 0, \\
\mathbb{P}(\tilde{Y}_0 = 1, \tilde{Y}_1 = 0, D = 1|Z = z) &= 0.
\end{aligned}
$$

Appendix G. Proof that adding monotonicity of the treatment in the instrument does not improve bounds in Proposition 3

Consider the following model:

$$
\begin{cases} Y = Y_1 D + Y_0(1-D) \\ D = \mathbb{1}\{V^* \leqslant P(Z)\} \end{cases} \tag{G.1}
$$

where the vector (Y, D, Z) represents the observed data, while the vector (Y_1, Y_0, V^*) is latent. The selection equation is equivalent to imposing monotonicity of the treatment D in the instrument Z as shown by Vytlacil (2002) under Assumption 1.

Assumption 7 (Absolute continuity of V)**.** *The latent variable V^* is absolutely continuous. Without loss of generality, the unconditional distribution of V^* is uniform over $[0, 1]$, and the support of the function $P(z)$ is included in $[0, 1]$.*

This assumption 7 holds in the bivariate probit specification, where $V^* = \Phi(V)$ and $P(z) = \Phi(\delta z) \in [0, 1]$.

Assumption 8 (Support condition for $P(Z)$)**.** *The support of the random variable $P(Z)$ is an interval: $\mathcal{P} = [\underline{p}, \overline{p}]$.*

This assumption holds if for example, the instrument Z is continuous. The result below holds when the instrument Z is discrete. We use Assumption 8 to ease the exposition.

Under Assumptions 1 and 7, the function $P(z)$ is identified as the propensity score: $P(z) = \mathbb{P}(D = 1|Z = z)$. For any $p \in \text{int } \mathcal{P}$, we have

$$\begin{aligned}
\mathbb{E}\left[YD|P(Z) = p\right] &= \mathbb{E}\left[Y_1 \mathbb{1}\left\{V \leqslant p\right\}|P(Z) = p\right], \\
&= \mathbb{E}\left[Y_1 \mathbb{1}\left\{V \leqslant p\right\}\right], \\
&= \mathbb{E}\left[Y_1|V \leqslant p\right]\mathbb{P}\left[V \leqslant p\right], \\
&= \left(\int_0^p \mathbb{E}\left[Y_1|V = v\right]\frac{f_V(v)}{\mathbb{P}\left(V \leqslant p\right)}dv\right)\cdot \mathbb{P}\left(V \leqslant p\right), \\
&= \int_0^p \mathbb{E}\left[Y_1|V = v\right]dv, \qquad \text{(G.2)}
\end{aligned}$$

where the second equality follows from Assumption 1, the third and fourth equalities follow from the law of iterated expectations, and the last equality follows from Assumption 7. Differentiating each side with respect to p point-identifies $\mathbb{E}\left[Y_1|V = p\right]$

$$\mathbb{E}\left[Y_1|V = p\right] = \frac{\partial \mathbb{E}\left[YD|P(Z) = p\right]}{\partial p}. \qquad \text{(G.3)}$$

Similarly, we have

$$\mathbb{E}\left[Y(1-D)|P(Z) = p\right] = \int_p^1 \mathbb{E}\left[Y_0|V = v\right]dv. \qquad \text{(G.4)}$$

Then,

$$\mathbb{E}\left[Y_0|V = p\right] = -\frac{\partial \mathbb{E}\left[Y(1-D)|P(Z) = p\right]}{\partial p}. \qquad \text{(G.5)}$$

Therefore, we identify the marginal treatment effect (MTE) for every $p \in \mathcal{P}$ as follows:

$$MTE(p) \equiv \mathbb{E}\left[Y_1 - Y_0|V = p\right] = \frac{\partial \mathbb{E}\left[Y|P(Z) = p\right]}{\partial p}.$$

This result has been shown by Heckman and Vytlacil (2005). They also showed that Equations (G.3) and (G.5) have the following testable implications:

$$0 \leqslant \frac{\partial \mathbb{E}\left[YD|P(Z) = p\right]}{\partial p} \leqslant 1, \qquad \text{(G.6)}$$

$$0 \leqslant -\frac{\partial \mathbb{E}\left[Y(1-D)|P(Z) = p\right]}{\partial p} \leqslant 1, \qquad \text{(G.7)}$$

and the model has these index sufficiency implications

$$\mathbb{E}\left[YD|P(Z) = P(z)\right] = \mathbb{E}\left[YD|Z = z\right], \qquad \text{(G.8)}$$

$$\mathbb{E}\left[Y(1-D)|P(Z) = P(z)\right] = \mathbb{E}\left[Y(1-D)|Z = z\right]. \qquad \text{(G.9)}$$

We build upon these results to derive sharp bounds on the ATE under 1, 4, 7, and 8. We have

$$\begin{aligned}\mathbb{E}\left[Y_1\right] &= \int_0^1 \mathbb{E}\left[Y_1|V=v\right]dv,\\ &= \int_0^{\overline{p}} \mathbb{E}\left[Y_1|V=v\right]dv + \int_{\overline{p}}^1 \mathbb{E}\left[Y_1|V=v\right]dv,\\ &= \mathbb{E}\left[YD|P(Z)=\overline{p}\right] + \int_{\overline{p}}^1 \mathbb{E}\left[Y_1|V=v\right]dv,\end{aligned}$$

where the last equality holds from Equation (G.2). Since $Y_1 \in \{0,1\}$, the integral $\int_{\overline{p}}^1 \mathbb{E}\left[Y_1|V=v\right]dv$ is bounded between $[0, 1-\overline{p}]$. Therefore, we obtain the following bounds on $\mathbb{E}\left[Y_1\right]$:

$$\mathbb{E}\left[YD|P(Z)=\overline{p}\right] \leqslant \mathbb{E}\left[Y_1\right] \leqslant \mathbb{E}\left[YD|P(Z)=\overline{p}\right] + 1 - \overline{p}. \tag{G.10}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}\left[Y_0\right] &= \int_0^1 \mathbb{E}\left[Y_0|V=v\right]dv,\\ &= \int_0^{\underline{p}} \mathbb{E}\left[Y_0|V=v\right]dv + \int_{\underline{p}}^1 \mathbb{E}\left[Y_0|V=v\right]dv,\\ &= \int_0^{\underline{p}} \mathbb{E}\left[Y_0|V=v\right]dv + \mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right],\end{aligned}$$

where the last equality holds from Equation (G.4). Since $Y_0 \in \{0,1\}$, we have $\mathbb{E}\left[Y_0|V=v\right]$ bounded between $[0, \underline{p}]$. Hence, the following bounds:

$$\mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right] \leqslant \mathbb{E}\left[Y_0\right] \leqslant \mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right] + \underline{p}. \tag{G.11}$$

Suppose that the first part of Assumption 4 holds, i.e., $Y_1 \geqslant Y_0$. Then, we have

$$\begin{aligned}\int_{\overline{p}}^1 \mathbb{E}\left[Y_1|V=v\right]dv &\geqslant \int_{\overline{p}}^1 \mathbb{E}\left[Y_0|V=v\right]dv = \mathbb{E}\left[Y(1-D)|P(Z)=\overline{p}\right],\\ \int_0^{\underline{p}} \mathbb{E}\left[Y_0|V=v\right]dv &\leqslant \int_0^{\underline{p}} \mathbb{E}\left[Y_1|V=v\right]dv = \mathbb{E}\left[YD|P(Z)=\underline{p}\right].\end{aligned}$$

Therefore, we obtain a tighter lower for $\mathbb{E}[Y_1]$, and a tighter upper bound for $\mathbb{E}[Y_]$:

$$\mathbb{E}\left[Y|P(Z)=\overline{p}\right] \leqslant \mathbb{E}\left[Y_1\right] \leqslant \mathbb{E}\left[YD|P(Z)=\overline{p}\right] + 1 - \overline{p}, \tag{G.12}$$

$$\mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right] \leqslant \mathbb{E}\left[Y_0\right] \leqslant \mathbb{E}\left[Y|P(Z)=\underline{p}\right]. \tag{G.13}$$

Similarly, if $Y_1 \leqslant Y_0$, we have

$$\mathbb{E}\left[YD|P(Z)=\overline{p}\right] \leqslant \mathbb{E}\left[Y_1\right] \leqslant \mathbb{E}\left[Y|P(Z)=\overline{p}\right], \tag{G.14}$$

$$\mathbb{E}\left[Y|P(Z)=\underline{p}\right] \leqslant \mathbb{E}\left[Y_0\right] \leqslant \mathbb{E}\left[Y_0\right] \leqslant \mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right]+\underline{p}. \tag{G.15}$$

Proposition 4. *Under Assumptions 1, 4, 7, and 8, sharp bounds for the average structural functions $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$ are given by equations (G.12) and (G.13), or (G.14) and (G.15), respectively; and sharp bounds for the ATE are the following:*

$$\begin{aligned}&\mathbb{E}\left[Y|P(Z)=\overline{p}\right]-\mathbb{E}\left[Y|P(Z)=\underline{p}\right]\\ &\quad\leqslant ATE \leqslant \mathbb{E}\left[YD|P(Z)=\overline{p}\right]+1-\overline{p}-\mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right].\end{aligned} \tag{G.16}$$

or

$$\begin{aligned}&\mathbb{E}\left[YD|P(Z)=\overline{p}\right]-\mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right]-\underline{p}\\ &\quad\leqslant ATE \leqslant \mathbb{E}\left[Y|P(Z)=\overline{p}\right]-\mathbb{E}\left[Y|P(Z)=\underline{p}\right].\end{aligned} \tag{G.17}$$

Comments. The bounds in Proposition 4 are identical to the bounds in Proposition 3 if conditions (G.6)-(G.9) hold, and $0 \leqslant \frac{\partial\mathbb{E}[Y|P(Z)=p]}{\partial p} \leqslant 1$ or $-1 \leqslant \frac{\partial\mathbb{E}[Y|P(Z)=p]}{\partial p} \leqslant 0$. To see this, notice that if $Y_1 \geqslant Y_0$, then $MTE(p) \in [0,1]$, which implies under the assumptions that $0 \leqslant \frac{\partial\mathbb{E}[Y|P(Z)=p]}{\partial p} \leqslant 1$. This condition combined with conditions (G.6)-(G.9) imply the following:

$$\begin{aligned}
\mathbb{E}\left[Y|P(Z)=\overline{p}\right] &= \sup_{p\in\mathcal{P}}\mathbb{E}\left[Y|P(Z)=p\right] = \sup_{P(z)\in\mathcal{P}}\mathbb{E}\left[Y|P(Z)=P(z)\right] = \sup_{z\in\mathcal{Z}}\mathbb{E}\left[Y|Z=z\right],\\
\mathbb{E}\left[Y|P(Z)=\underline{p}\right] &= \inf_{p\in\mathcal{P}}\mathbb{E}\left[Y|P(Z)=p\right] = \inf_{P(z)\in\mathcal{P}}\mathbb{E}\left[Y|P(Z)=P(z)\right] = \inf_{z\in\mathcal{Z}}\mathbb{E}\left[Y|Z=z\right],\\
\mathbb{E}\left[Y(1-D)|P(Z)=\underline{p}\right] &= \sup_{p\in\mathcal{P}}\mathbb{E}\left[Y(1-D)|P(Z)=p\right] = \sup_{P(z)\in\mathcal{P}}\mathbb{E}\left[Y(1-D)|P(Z)=P(z)\right],\\
&= \sup_{z\in\mathcal{Z}}\mathbb{E}\left[Y(1-D)|Z=z\right],\\
\mathbb{E}\left[YD|P(Z)=\overline{p}\right]+1-\overline{p} &= 1-\mathbb{E}\left[(1-Y)D|P(Z)=\overline{p}\right] = 1-\sup_{p\in\mathcal{P}}\mathbb{E}\left[(1-Y)D|P(Z)=p\right],\\
&= \inf_{p\in\mathcal{P}}\mathbb{E}\left[1-(1-Y)D|P(Z)=p\right] = \inf_{P(z)\in\mathcal{P}}\mathbb{E}\left[1-(1-Y)D|P(Z)=P(z)\right],\\
&= \inf_{z\in\mathcal{Z}}\mathbb{E}\left[1-(1-Y)D|Z=z\right].
\end{aligned}$$

Similar results hold in the case where $-1 \leqslant \frac{\partial\mathbb{E}[Y|P(Z)=p]}{\partial p} \leqslant 0$, and this will correspond to the scenario $Y_1 \leqslant Y_0$.

Proof of Proposition 4. In this section, we assume that the testable implications (G.6), (G.7), (G.8), and (G.9) hold. We show the proof for the case $Y_1 \geqslant Y_0$, since that for the case $Y_0 \geqslant Y_1$ is similar.

The lower bound for the ATE is attained. Define

$$\begin{cases} \tilde{Y} &= \tilde{Y}_1 \tilde{D} + \tilde{Y}_0(1-\tilde{D}) \\ \tilde{D} &= \mathbb{1}\left\{\tilde{V}^* \leqslant P(Z)\right\} \end{cases} \tag{G.18}$$

such that $\tilde{V}^* \sim \mathcal{U}_{[0,1]}$, $P(z) = \mathbb{P}(D=1|Z=z)$, $Z \perp\!\!\!\perp \tilde{V}^*$,

$$\mathbb{P}(\tilde{Y}_1 = 1, \tilde{Y}_0 = 0 | \tilde{V}^* = p) = \begin{cases} \frac{\partial \mathbb{E}[Y|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ 0 \text{ if } p \leqslant \underline{p} \\ 0 \text{ if } p \geqslant \overline{p} \end{cases}$$

$$\mathbb{P}(\tilde{Y}_1 = 1, \tilde{Y}_0 = 1 | \tilde{V}^* = p) = \begin{cases} -\frac{\partial \mathbb{E}[Y(1-D)|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ \frac{\mathbb{E}\left[YD|P(Z)=\underline{p}\right]}{\underline{p}} & \text{if } p \leqslant \underline{p} \\ \frac{\mathbb{E}[Y(1-D)|P(Z)=\overline{p}]}{1-\overline{p}} & \text{if } p \geqslant \overline{p} \end{cases}$$

$$\mathbb{P}(\tilde{Y}_1 = 0, \tilde{Y}_0 = 0 | \tilde{V}^* = p) = \begin{cases} 1 - \frac{\partial \mathbb{E}[YD|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ 1 - \frac{\mathbb{E}\left[YD|P(Z)=\underline{p}\right]}{\underline{p}} & \text{if } p \leqslant \underline{p} \\ 1 - \frac{\mathbb{E}[Y(1-D)|P(Z)=\overline{p}]}{1-\overline{p}} & \text{if } p \geqslant \overline{p} \end{cases}$$

and $\mathbb{P}(\tilde{Y}_1 = y_1, \tilde{Y}_0 = y_0 | \tilde{V}^* = p, Z = z) = \mathbb{P}(\tilde{Y}_1 = y_1, \tilde{Y}_0 = y_0 | \tilde{V}^* = p)$, y_1, $y_0 \in \{0,1\}$. It is easy to check that this joint distribution is well-defined, satisfies Assumptions 1, 4, 7, and 8. We are going to show that it is compatible with the data and its ATE is equal to the lower bound. We have

$$
\begin{aligned}
\mathbb{P}\left(\tilde{Y}=1,\tilde{D}=1|Z=z\right) &= \mathbb{P}\left(\tilde{Y}_1=1,\tilde{V}^*\leqslant P(z)|Z=z\right),\\
&= \mathbb{P}\left(\tilde{Y}_1=1,\tilde{V}^*\leqslant P(z)\right),\\
&= \mathbb{E}\left[\tilde{Y}_1|\tilde{V}^*\leqslant P(z)\right]\mathbb{P}\left(\tilde{V}^*\leqslant P(z)\right),\\
&= \left(\int_0^{P(z)}\mathbb{E}\left[\tilde{Y}_1|\tilde{V}^*=p\right]\frac{f_{\tilde{V}^*}(p)}{\mathbb{P}\left(\tilde{V}^*\leqslant P(z)\right)}dp\right)\cdot\mathbb{P}\left(\tilde{V}^*\leqslant P(z)\right),\\
&= \int_0^{P(z)}\mathbb{E}\left[\tilde{Y}_1|\tilde{V}^*=p\right]dp,\\
&= \mathbb{E}\left[YD|P(Z)=\underline{p}\right]+\mathbb{E}\left[YD|P(Z)=P(z)\right]-\mathbb{E}\left[YD|P(Z)=\underline{p}\right],\\
&= \mathbb{E}\left[YD|P(Z)=P(z)\right],\\
&= \mathbb{E}\left[YD|Z=z\right]=\mathbb{P}\left(Y=1,D=1|Z=z\right).
\end{aligned}
$$

Similarly, we show that

$$
\begin{aligned}
\mathbb{P}\left(\tilde{Y}=0,\tilde{D}=1|Z=z\right) &= \mathbb{P}\left(Y=0,D=1|Z=z\right),\\
\mathbb{P}\left(\tilde{Y}=1,\tilde{D}=0|Z=z\right) &= \mathbb{P}\left(Y=1,D=0|Z=z\right),\\
\mathbb{P}\left(\tilde{Y}=0,\tilde{D}=0|Z=z\right) &= \mathbb{P}\left(Y=0,D=0|Z=z\right).
\end{aligned}
$$

Finally, we have

$$
A\tilde{T}E=\mathbb{E}\left[\tilde{Y}_1-\tilde{Y}_0\right]=\mathbb{E}\left[\tilde{Y}_1\right]-\mathbb{E}\left[\tilde{Y}_0\right]=\mathbb{P}(\tilde{Y}_1=1,\tilde{Y}_0=0),
$$

where

$$
\begin{aligned}
\mathbb{P}(\tilde{Y}_1=1,\tilde{Y}_0=0) &= \int_0^1\mathbb{P}(\tilde{Y}_1=1,\tilde{Y}_0=0|\tilde{V}^*=p)dp,\\
&= \int_{\underline{p}}^{\overline{p}}\frac{\partial\mathbb{E}\left[Y|P(Z)=p\right]}{\partial p}dp,\\
&= \mathbb{E}\left[Y|P(Z)=\overline{p}\right]-\mathbb{E}\left[Y|P(Z)=\underline{p}\right].
\end{aligned}
$$

The upper bound for the ATE is attained. Similarly, the following joint distribution achieves the upper bound for the ATE. Define

$$\mathbb{P}(\tilde{Y}_1 = 1, \tilde{Y}_0 = 0 | \tilde{V}^* = p) = \begin{cases} \frac{\partial \mathbb{E}[Y|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ \frac{\mathbb{E}[(1-Y)(1-D)|P(Z)=\overline{p}]}{1-\overline{p}} & \text{if } p \geqslant \overline{p} \\ \frac{\mathbb{E}\left[YD|P(Z)=\underline{p}\right]}{\underline{p}} & \text{if } p \leqslant \underline{p} \end{cases}$$

$$\mathbb{P}(\tilde{Y}_1 = 1, \tilde{Y}_0 = 1 | \tilde{V}^* = p) = \begin{cases} -\frac{\partial \mathbb{E}[Y(1-D)|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ 1 - \frac{\mathbb{E}[(1-Y)(1-D)|P(Z)=\overline{p}]}{1-\overline{p}} & \text{if } p \geqslant \overline{p} \\ 1 - \frac{\mathbb{E}\left[YD|P(Z)=\underline{p}\right]}{\underline{p}} & \text{if } p \leqslant \underline{p} \end{cases}$$

$$\mathbb{P}(\tilde{Y}_1 = 0, \tilde{Y}_0 = 0 | \tilde{V}^* = p) = \begin{cases} 1 - \frac{\partial \mathbb{E}[YD|P(Z)=p]}{\partial p} & \text{if } \underline{p} < p < \overline{p} \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbb{P}(\tilde{Y}_1 = y_1, \tilde{Y}_0 = y_0 | \tilde{V}^* = p, Z = z) = \mathbb{P}(\tilde{Y}_1 = y_1, \tilde{Y}_0 = y_0 | \tilde{V}^* = p)$, $y_1, y_0 \in \{0, 1\}$.

References

Altonji, J. G., T. E. Elder, and C. R. Taber. 2005a. "An evaluation of instrumental variable strategies for estimating the effects of catholic schooling." *Journal of Human Resources* 40 (4):791–821.

Altonji, Joseph, Todd Elder, and Christopher Taber. 2005b. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1):151–184.

Andrews, D. W. K. and X. Shi. 2013. "Inference Based on Conditional Moment Inequalities." *Econometrica* 81:609–666.

Andrews, D. W. K. and G. Soares. 2010. "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection." *Econometrica* 78 (1):119–157.

Arai, Y., Y-C. Hsu, T. Kitagawa, I. Mourifé, and Y. Wan. 2018. "Testing identifying assumptions in fuzzy regression discontinuity designs." *Cemmap Working Paper* CWP50/18.

Balke, Alexander and Judea Pearl. 1997. "Bounds on Treatment Effects from Studies with Imperfect Compliance." *Journal of the American Statistical Association* 92 (439):1171–1176.

Berger, R. L. and J. C. Hsu. 1996. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets." *Statistical Science* 11 (4):283–319.

Chernozhukov, Victor, Wooyoung Kim, Sokbae Lee, and Adam M. Rosen. 2015. "Implementing Intersection Bounds in Stata." *Stata Journal* 15 (1):21–44.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica* 81 (2):667–737. URL `http://dx.doi.org/10.3982/ECTA8718`.

Evans, W. N. and R. M. Schwab. 1995. "Finishing high school and starting college: Do Catholic schools make a difference?" *Quarterly Journal of Economics* 110 (4):941–974.

Galichon, Alfred and Marc Henry. 2011. "Set identification in models with multiple equilibria." *The Review of Economic Studies* 78 (4):1264–1298.

Gao, Li, Wendong Zhang, Yingdan Mei, Abdoul G. Sam, Yu Song, and Shuqin Jin. 2018. "Do farmers adopt fewer conservation practices on rented land? Evidence from straw retention in China." *Land Use Policy* 79:609–621.

Goldman, D., j. Bhattacharya, D. Mccaffrey, N. Duan, A. Leibowitz, G. Joyce, and S. Morton. 2001. "Effect of Insurance on Mortality in an HIV-Positive Population in Care." *Journal of American Statistical Association* 96 (455):883–894.

Gunsilius, F. 2020. "Non-testability of instrument validity under continuous treatments." *Biometrika (forthcoming)* .

Han, S. and S. Lee. 2019. "Estimation in a generalization of bivariate probit models with dummy endogenous regressors." *Journal of Applied Econometrics* :1–22.

Han, S. and E. J. Vytlacil. 2017. "Identification in a generalization of bivariate probit models with dummy endogenous regressors." *Journal of Econometrics* 199:63–73.

Heckman, James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46 (4):931–959.

Heckman, James J and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3):669–738.

Huber, Martin and Giovanni Mellace. 2015. "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints." *The Review of Economics and Statistics* 97 (2):398–411.

Kédagni, D. and I. Mourifié. 2020. "Generalized Instrumental Inequalities: Testing the Instrumental Variable Independence Assumption." *Biometrika* 107 (3):661–675.

Kitagawa, Toru. 2015. "A Test for Instrument Validity." *Econometrica* 83:2043–2063.

Li, Chuhui, D.S. Poskitt, and Xueyan Zhao. 2019. "The bivariate probit model, maximum likelihood estimation, pseudo true parameters and partial identification." *Journal of Econometrics* 209:94–113.

Machado, C., A. Shaikh, and E. Vytlacil. 2019. "Instrumental Variables and the Sign of the Average Treatment Effect." *Journal of Econometrics* 212:522–555.

Manski, C. F. and J. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68:997–1010.

Manski, Charles F. 1997. "Monotone Treatment Response." *Econometrica* 65 (6):1311–1334.

Mourifié, I. and R. Méango. 2014. "A note on the identification in two equations probit model with dummy endogenous regressor." *Economics Letters* 125:360–363.

Mourifié, I. and Y. Wan. 2017. "Testing Local Average Treatment Effect Assumptions." *The Review of Economics and Statistics* 99 (2):305–313.

Neal, D. A. 1997. "The effects of catholic secondary schooling on educational achievement." *Journal of Labor Economics* 15 (1, Part 1):98–123.

Pearl, J. 1995. "Causal inference from indirect experiments." *Artificial Intelligence in Medicine* 7:561–582.

Pearl, Judea. 1994. "On the Testability of Causal Models with Latent and Instrumental Variables." *Uncertainty in Artificial Intelligence* 11:435–443.

Rhine, S. L., W. H. Greene, and M. Toussaint-Comeau. 2006. "The importance of check-cashing businesses to the unbanked: Racial/ethnic differences." *Review of Economics and Statistics* 88 (1):146–157.

Rosenbaum, P. R. 1995. *Observational Studies*. New York: Springer-Verlag.

Rosenbaum, P. R. and D. B. Rubin. 1983. "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome." *Journal of the Royal Statistical Society. Series B (Methodological)* 45 (2):212–218.

Vytlacil, Edward. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70 (1):331–341.

Zimmer, David. 2017. "Using copulas to estimate the coefficient of a binary endogenous regressor in a Poisson regression: Application to the effect of insurance on doctor visits." *Health Economics* 27:545–456.