Specification tests for generalized propensity scores using double projections[*]

Pedro H. C. Sant'Anna[†]   Xiaojun Song[‡]

April 15, 2023

Abstract

This paper proposes a new class of nonparametric tests for the correct specification of models based on conditional moment restrictions, paying particular attention to generalized propensity score models. The test procedure is based on two different projection arguments, leading to test statistics that are suitable to setups with many covariates, and are (asymptotically) invariant to the estimation method used to estimate the nuisance parameters. We show that our proposed tests are able to detect a broad class of local alternatives converging to the null at the usual parametric rate and illustrate its attractive power properties via simulations. We also extend our proposal to test parametric or semiparametric single-index-type models.

Keywords: Double projections; generalized propensity scores; multiplier bootstrap; single/multiple-index models; specification tests

1 Introduction

One of the primary goals of many scientific fields is to quantify the effect of an exposure, policy, or treatment on outcomes of interest. When the assignment to treatment is not randomized, groups with different levels of the treatment variable usually differ in important ways other than the observed treatment. Because these differences are many times associated with the outcome variable, ascertaining the causal effect of the treatment requires more sophisticated statistical tools than a simple comparison of means. It is in this setting that the propensity score and its multi-valued generalizations have been shown to be among the most widely used tools for causal inference; see, e.g., Imbens and Rubin (2015), Linden et al. (2016) and Lopez and Gutman (2017).

Although statistical procedures that build on the propensity score and its generalizations (henceforth GPS) are popular, a main concern of these methods is that the GPS is usually unknown, and therefore has to be estimated. Given the availability of many pre-treatment covariates and limited sample size, researchers usually adopt parametric models for the GPS. Such a common practice raises the important issue of model misspecifications. Indeed, as illustrated by Busso et al. (2014) and Linden et al. (2016), model misspecifications can lead to misleading treatment effect estimates. Thus, in practice, assessing if your parametric putative model for the GPS is correctly specified is recommended.

In this paper, we propose new goodness-of-fit tests for models based on conditional moment restrictions. Although our tests can be used more generally, we pay particular attention to testing whether a GPS model is correctly specified. The main distinguishing feature of our tests is that they combine two different projections. The first projection is a dimension-reduction device that allows us to handle situations with many covariates (Escanciano, 2006). The second projection is used to make the test asymptotically invariant to the estimation method used to estimate the nuisance parameters (Neyman, 1959; Bickel et al., 2006; Escanciano and Goh, 2014; Sant'Anna and Song, 2019). As a result of this second projection, we can consider a wider range of estimators for nuisance parameters and implement a fairly simple bootstrap procedure that works for discrete responses, continuous responses, or a mixture of both.

This paper is the first to combine the two aforementioned projections. Escanciano (2006)'s test, although robust to dimensionality, is not robust to the estimation of nuisance parameters. A potential drawback of the test of Escanciano (2006) is that its implementation relies on the wild bootstrap, and it is not directly applicable when the response variables are discrete or mixed, which is the case in GPS models. On the other hand, the test of Escanciano and Goh (2014), although robust to the estimation of nuisance parameters, is not robust to the presence of many covariates. The test of Sant'Anna and Song (2019), although also constructed to be robust to many covariates and invari-

ant to the estimation of nuisance parameters, only tests for an *implication* of the null hypothesis, not for the null hypothesis itself. Furthermore, it is only justified for binary choice models. By contrast, our proposed tests are based on testing the null hypothesis itself, are robust to both the dimensionality of covariates and the estimation of nuisance parameters, and can be directly used for generic outcome models.

Overall, our proposed tests enjoy several attractive features. They (a) do not severely suffer from the "curse of dimensionality" when we have many covariates; (b) are data-driven and do not require tuning parameters such as bandwidths; (c) do not require estimators to be $n^{1/2}$-asymptotically linear, with n the sample size; and (d) are able to detect a broad class of local alternatives converging to the null at the parametric rate. In order to facilitate its practical implementation, we obtain closed-form expressions for our test statistics and show that critical values can be computed with the assistance of an easy-to-use multiplier-type bootstrap. To the best of our knowledge, no other (specification) test available in the literature enjoys all these attractive properties (e.g., Escanciano, 2006, Mora and Moro-Egido, 2008, Shaikh et al., 2009, Escanciano and Goh, 2014, García-Portugués et al., 2014, Iwasawa, 2015, Zhu et al., 2017, Sant'Anna and Song, 2019, Domínguez and Lobato (2020), and Kim et al., 2020). The results of Monte Carlo simulations indicate that these attractive properties translate to tests with excellent finite sample properties, even when the dimension of covariates is relatively large.

We also consider extensions of our basic setup to other more complex frameworks. For instance, practitioners routinely use parametric GPS models that incorporate index restrictions. Single-index models are popular with binary or ordered multinomial treatments. With more general unordered, multinomial choice models, practitioners often use multiple-index models such as multinomial logit/probit. If such specifications were to be rejected by our omnibus tests, it would not be clear if it is due to violations of the index restriction, the parametric distributional assumption, or both. We extend our proposal to test index models to shed light on these concerns. Using our double-projection arguments, we propose a test for a parametric index model against semiparametric alternatives that maintain the index restrictions (directional test). We also extend our framework to test a single-index GPS model with an unknown link function against general nonparametric alternatives. We highlight that our double-projection procedure can eliminate the *parametric*-type estimation effect in this semiparametric setup, though it does not eliminate the *nonparametric*-type effect coming from estimating the link function. This additional complication suggests that one should use a different, much more demanding multiplier-bootstrap procedure to compute critical values. Although we discuss such challenges, we believe that a detailed solution to these issues is beyond the scope of this paper, and we leave it for future research.

The rest of this article is organized as follows. In Section 2, we present the testing framework and introduce our proposed double-projection specification tests. The

asymptotic properties of our tests are established in Section 3, with the asymptotic null distribution derived in Section 3.1 and the asymptotic power properties studied in Section 3.2. In Section 4, we present a detailed description of an easy-to-implement multiplier-bootstrap procedure to compute critical values. In Section 5, we extend our testing procedure to test index-type models; Section 5.1 tests a parametric index model against a nonparametric alternative, Section 5.2 tests a parametric index model against a semiparametric alternative, and Section 5.3 tests a semiparametric single-index model against a nonparametric alternative. We then illustrate the usefulness of our tests through an empirical application in Section 6. Section 7 concludes. Mathematical proofs and Monte Carlo simulation results are gathered in an online supplementary appendix.[1]

2 Specification tests based on double projections

2.1 Setup and motivation

In this paper, we seek to test hypotheses of the type of

$$H_0 : \mathbb{P}\left(\mathbb{E}\left[e(t;\theta_0)|X\right] = 0\right) = 1 \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \qquad (2.1)$$

against

$$H_1 : \mathbb{P}\left(\mathbb{E}\left[e(t;\theta)|X\right] = 0\right) < 1 \text{ for any } \theta \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and/or some } t \in \mathcal{T}, \qquad (2.2)$$

where $e(t;\theta)$ is a generalized error term indexed by $t \in \mathcal{T} \subset \mathbb{N}$, and $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$, with Θ a compact space with $d_\theta \geq 1$ a given positive integer. For notational simplicity, we suppress the dependence on t in θ and related quantities. Clearly, the null hypothesis H_0 is composite. Our main goal is to propose tests of H_0 against H_1 that are robust to the presence of many covariates[2] and the estimation of nuisance parameter θ_0.

Although our framework is general, it is worth motivating within a popular and empirically relevant causal inference setup. For a generic $d_1 \times d_2$ matrix A, $A^\top$ denotes the transpose of A. Let $J \in \mathbb{N}$ be a given finite positive integer. Let $(X^\top, T, Y)^\top$ be a random vector in a $(d_x + 2)$-dimensional Euclidean space, where $X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ is an observable $d_x \times 1$ vector of covariates with $d_x \in \mathbb{N}$, $T \in \mathcal{T} \subseteq \{0, 1, \dots, J\}$ is the treatment random variable, $Y = \sum_{t=0}^{\mathcal{T}} 1\,(T = t)\,Y\,(t) \in \mathcal{Y} \subseteq \mathbb{R}$ is the observed outcome, and $Y\,(t)$ denotes the potential outcome when T is externally set to t. For the sake of simplicity, let's focus

---

on average causal effects of the form

$$\beta(t,s) \equiv \mathbb{E}\left[Y(t) - Y(s)\right], \tag{2.3}$$

the average causal effect of exposing units to treatment t rather than treatment s. We can also cover quantile treatment effects and/or treatment effects for the treated subpopulation (Lee, 2018, Ao et al., 2019).

Imbens (2000) shows that, if assignment to treatment T is weakly unconfounded given the pre-treatment variables X, $\beta(t,s)$ in (2.3) is identified by the following weighting estimands:

$$\beta(t,s) = \mathbb{E}\left[\frac{Y1(T=t)}{p_t(X)}\right] - \mathbb{E}\left[\frac{Y1(T=s)}{p_s(X)}\right], \tag{2.4}$$

where $p_t(x) \equiv \mathbb{P}(T=t|X=x) = \mathbb{E}[1(T=t)|X=x]$ is the unknown GPS. To estimate $\beta(t,s)$ using (2.4), researchers usually assume a parametric model $q_t(x,\theta_t)$ for $p_t(x)$, where $q_t(x,\theta_t) : \mathcal{X} \times \Theta_t \mapsto [0,1]$ denotes a family of parametric functions known up to the finite-dimensional parameter vector θ_t. By construction, $\sum_{j=0}^{J} q_j(x,\theta_j) = 1$. For example, when treatments are multi-valued, qualitatively distinct, and without a logical ordering, a popular specification for $q_t(x,\theta_t)$ is the multinomial logit model[3] (e.g., Imbens, 2000),

$$q_t(X,\theta_t) = \frac{\exp(X'\theta_t)}{1 + \sum_{j=1}^{J}\exp(X'\theta_j)}, \quad t \in \{1,\ldots,J\}. \tag{2.5}$$

The parameter vector $\{\theta_t, t \in \mathcal{T}\}$ can be estimated using the maximum likelihood approach. Researchers can then estimate $\beta(t,s)$ through the inverse probability weighting (IPW) approach.

Given the widespread empirical practice of adopting parametric models for the GPS, a natural concern is potential model misspecifications. In particular, if the working model $q_t(x,\theta_t)$ for the GPS $p_t(x)$ is misspecified, causal effects estimators such as the inverse probability weighted estimator are in general biased and policy recommendations based on them can be highly misleading (e.g., Linden et al., 2016). To assess whether this is the case, one can frame this as a specification test for $q_t(x,\theta_t)$ that fits into our testing framework by simply setting $e(t;\theta)$ to be the parametrically specified generalized error under the null for every $t \in \mathcal{T}$, i.e., $e(t;\theta) \equiv 1(T=t) - q_t(X,\theta)$.

For motivational and concreteness sake, we take $e(t;\theta)$ as the generalized error of the propensity score model as mentioned above in the rest of the paper, though we can handle much more general models. See, e.g., Remark 1 for additional details.

---

3 One limitation with the multinomial logit model is the independence of irrelevant alternatives (IIA) assumption imposed by it. One alternative to avoid the IIA assumption is the multinomial probit model. But if $J \geq 2$, multinomial probit model meets practical obstacles as it involves computing $(J+1)$-dimensional integrals.

2.2 Our proposed test

The characterization of H_0 and H_1 in (2.1) and (2.2), respectively, makes explicit that we are interested in testing conditional moment restrictions (González-Manteiga and Crujeiras, 2013). As argued by Escanciano (2006), (2.1) can be equivalently characterized as

$$H_0 : R_t^{pro}(\beta, u; \theta_0) = 0 \text{ almost everywhere (a.e.) } \left(\beta^\top, u\right)^\top \in \Pi_{pro},$$
$$\text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \qquad (2.6)$$

where

$$R_t^{pro}(\beta, u; \theta) \equiv \mathbb{E}\left[e(t; \theta)1\left(X^\top \beta \le u\right)\right],$$

and $\Pi_{pro} \equiv \mathbb{S}^{d_x} \times \overline{\overline{\mathbb{R}}}$ denotes the projected space, with $\overline{\overline{\mathbb{R}}} = [-\infty, \infty]$ the extended real line and $\mathbb{S}^{d_x}$ the unit ball in $\mathbb{R}^{d_x}$, i.e., $\mathbb{S}^{d_x} = \left\{\beta \in \mathbb{R}^{d_x} : ||\beta|| = 1\right\}$ with $||A|| = \left[\text{tr}\left(AA^\top\right)\right]^{1/2}$ denoting the Euclidean norm for a generic matrix A.

Although one can find alternative characterizations of H_0, see, e.g., Bierens (1982), Stute (1997), and Dominguez and Lobato (2004), our main motivations for expressing H_0 as in (2.6) are that (i) $R_t^{pro}(\beta, u; \theta_0)$ is based on unconditional moment restrictions, implying that we can avoid the use of tuning parameters such as bandwidths when estimating $R_t^{pro}(\beta, u; \theta_0)$; and (ii) $R_t^{pro}(\beta, u; \theta_0)$ depends on covariates only through the one-dimensional projection $X^\top \beta$, greatly reducing the dimensionality of the problem. Indeed, this dimension-reduction device has been proven valuable in various contexts that need to deal with many covariates; see, e.g., Escanciano (2006), García-Portugués et al. (2014), Sun et al. (2017), Zhu et al. (2017), and Kim et al. (2020); for an overview, see Guo and Zhu (2017). However, it is worth mentioning that (2.6) involves not only a single process $R_t^{pro}(\beta, u; \theta_0)$ as is commonly the case in the specification testing literature (see Escanciano, 2008 for an exception), but J different processes $R_t^{pro}(\beta, u; \theta_0)$ associated with the J different treatment levels t.

From (2.6), one natural way to proceed is to compute the generalized residual marked empirical process based on the projections $1\left(X^\top \beta \le u\right)$,

$$R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(t; \widehat{\theta}_n)1(X_i^\top \beta \le u), \ \ \left(\beta^\top, u\right)^\top \in \Pi_{pro},$$

where $\widehat{\theta}_n$ is any $\sqrt{n}$-consistent estimator for θ_0, say the maximum likelihood estimator, and $e_i(t; \widehat{\theta}_n) \equiv 1(T_i = t) - q_t(X_i, \widehat{\theta}_n)$, $i = 1, \dots, n$, are the parametrically specified generalized residuals under the null H_0. Then, one can use continuous functionals of $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ to measure its distance from zero and assess if H_0 is rejected or not.

Although natural, there are some potential drawbacks of using $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$. For instance, the underlying null limiting distribution of $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ depends on the esti-

mator $\widehat{\theta}_n$. Indeed, the asymptotic null distribution of tests based on $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ will depend on whether one estimates θ_0 using maximum likelihood, nonlinear least squares, or by the method of estimating equations, even though the underlying specification for the null model is the same across these estimation methods. Perhaps more importantly, as noted by Escanciano (2006), tests based on $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ also require that $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ admits an asymptotically linear representation.[4] Such a condition can be demanding, especially when one wishes to use estimation methods that involve penalization (see, e.g., Knight and Fu, 2000 and Bühlmann and van de Geer, 2011). The fact that $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ is not invariant to $\widehat{\theta}_n$ also precludes an easy-to-implement multiplier bootstrap procedure to obtain critical values. This is inconvenient, especially when the response variables are discrete or mixed, since the wild bootstrap method requires regenerating dependent variables and fails to mimic the original data structure.

Given these potential drawbacks, we follow an alternative route that leads to estimator-invariant tests. More specifically, our proposed test statistics are continuous functionals of the following generalized residual marked empirical process based on the double projections:

$$R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t; \widehat{\theta}_n) \mathcal{P}_{n,t} 1(X_i^\top \beta \leq u), \quad (\beta^\top, u)^\top \in \Pi_{pro},$$

where the double-projected weight is

$$\mathcal{P}_{n,t} 1\left(X_i^\top \beta \leq u\right) \equiv 1\left(X_i^\top \beta \leq u\right) - g_t(X_i, \widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) G_{n,t}(\beta, u; \widehat{\theta}_n), \qquad (2.7)$$

where, for each $t \in \mathcal{T}$, $g_t(x, \theta) \equiv \partial q_t(x, \theta)/\partial\theta$, denotes the score function associated with the parametric model $q_t(x, \theta)$, and

$$\Delta_{n,t}(\widehat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} g_t(X_i, \widehat{\theta}_n) g_t^\top(X_i, \widehat{\theta}_n) \text{ and } G_{n,t}(\beta, u; \widehat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} g_t(X_i, \widehat{\theta}_n) 1\left(X_i^\top \beta \leq u\right).$$

We label $\mathcal{P}_{n,t} 1\left(X^\top \beta \leq u\right)$ as a double-projection because, as it is evident from (2.7), it involves first using the projection proposed by Escanciano (2006), $1\left(X^\top \beta \leq u\right)$, and then projecting $1\left(X^\top \beta \leq u\right)$ onto the tangent space of the nuisance parameters (see, e.g., Neyman, 1959, Escanciano and Goh, 2014, and Sant'Anna and Song, 2019). To the best of our knowledge, this paper is the first to incorporate this double-projection

---

[4] Throughout the paper, an estimator $\widehat{\theta}_n$ is said to be root-n asymptotically linear if it satisfies the following asymptotic expansion under H_0:

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(T_i, X_i, \theta_0) + o_p(1),$$

where $l(\cdot, \cdot, \cdot)$ is such that $\mathbb{E}[l(T, X, \theta_0)] = 0$ and $L(\theta_0) = \mathbb{E}[l(T, X, \theta_0)l(T, X, \theta_0)^\top]$ exists and is positive definite. Note that different $\widehat{\theta}_n$ may have different $l(\cdot, \cdot, \cdot)$.

argument, which, in practice, translates to test statistics that are robust against the "curse of dimensionality" and whose limiting null distributions are asymptotically invariant to $\widehat{\theta}_n$. This latter property follows from the fact that, for each $t \in \mathcal{T}$,

$$\mathbb{E}\left[g_t(X,\theta_0)\mathcal{P}_t 1\left(X^\top \beta \le u\right)\right] \equiv 0,$$

almost everywhere in $\left(\beta^\top, u\right)^\top \in \Pi_{pro}$, where

$$\mathcal{P}_t 1\left(X^\top \beta \le u\right) \equiv 1\left(X^\top \beta \le u\right) - g_t(X,\theta_0)^\top \Delta_t^{-1}(\theta_0)G_t(\beta,u;\theta_0), \tag{2.8}$$

with $\Delta_t(\theta) = \mathbb{E}\left[g_t(X,\theta)g_t(X,\theta)^\top\right]$ and $G_t(\beta,u;\theta) = \mathbb{E}\left[g_t(X,\theta)1\left(X^\top \beta \le u\right)\right]$. Note also that $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$ does not depend on tuning parameters such as bandwidths.

The intuition behind (2.8) is very simple. First of all, note that, for each $t \in \mathcal{T}$, $\left(\beta^\top, u\right)^\top \in \Pi_{pro}$, $\Delta_t^{-1}(\theta_0)G_t(\beta,u;\theta_0)$ is the vector of linear projection coefficients of regressing $1\left(X^\top \beta \le u\right)$ on the score function $g_t(X,\theta_0)$. Thus, it follows that

$$g_t(X,\theta_0)^\top \Delta_t^{-1}(\theta_0)G_t(\beta,u;\theta_0)$$

is the best linear predictor of $1\left(X^\top \beta \le u\right)$ given $g_t(X,\theta_0)$, and that (2.8) is nothing more than the associated projection error, which, by definition, is orthogonal to $g_t(X,\theta_0)$.

This orthogonality condition exploited by the double-projection procedure has important consequences. For example, under some weak regularity conditions, uniformly in $\left(\beta^\top, u\right)^\top \in \Pi_{pro}$,

$$R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)\mathcal{P}_t 1(X_i^\top \beta \le u) + o_p(1), \tag{2.9}$$

for each $t \in \mathcal{T}$; see the proof of Theorem 3.1 in the next section. As so, $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$ is asymptotically invariant to the choice of the estimator $\widehat{\theta}_n$, which, in turn, facilitates a simple multiplier bootstrap method to simulate critical values of test statistics based on $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$. Indeed, given that we can "ignore" estimation effects when computing $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$, all we need to do is "perturbate" the residuals $e(t;\widehat{\theta}_n)$; see Section 4 for details. Here, it is worth stressing that this is only feasible due to the usage of the second projection. Without it, we would need to either rule out discrete/mixed outcomes, and/or further impose that $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ admits an asymptotically linear representation. Even in these cases, different estimators usually have different asymptotically linear representations, leading to different multiplier bootstraps with potentially different testing results. Of course, (2.9) allows us to avoid these problems.

In order to operationalize our testing procedure, we need to choose a norm to measure the distance of $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$ from zero. We propose using the popular Cramér–von Mises

(CvM thereafter)-type test statistic

$$CvM_n^{dpro} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \right)^2 F_{n,\beta}(du)\, d\beta, \qquad (2.10)$$

where, for each t, $a_n(t)$ is a pre-specified (potentially random) non-negative weighting function, $F_{n,\beta}(u) = n^{-1} \sum_{i=1}^{n} 1\left(X_i^\top \beta \leq u \right)$ is the empirical distribution function of the one-dimensional projected regressors $\left\{ X_i^\top \beta \right\}_{i=1}^{n}$ for any fixed projected direction $\beta \in \mathbb{S}^{d_x}$, and $d\beta$ is the rescaled uniform density on the unit sphere $\mathbb{S}^{d_x}$.

We reject the null H_0 in favor of the alternative H_1 whenever CvM_n^{dpro} in (2.10) is "overly" large. For the sake of practical convenience, in the empirical application in Section 6 and Monte Carlo simulations in the online supplementary appendix, we use the constant weight $a_n(t) \equiv 1$ for all $t \in \mathcal{T}$, though other sensible data-driven choices are feasible, e.g., $a_n(t) = n^{-1} \sum_{i=1}^{n} 1(T_i = t)$. Due to space constraint, we do not intend to search for the optimal weight that yields the highest power.

At this stage, one may wonder why we have chosen to use a CvM-type instead of a Kolmogorov–Sminov-type test statistic. The reason is computational: as we show below in Lemma 2.1, (2.10) can be written in a closed-form expression and does not rely on any numerical integration method. Furthermore, one does not need to compute a different projection for each $\left(\beta^\top, u \right)^\top \in \Pi_{pro}$, which is arguably a "natural" step if one were to take Neyman (1959)'s "debiasing" proposal literally. A direct consequence of these attractive computational features is that (2.10) can be easily implemented even with many covariates and many treatment levels. In addition, the closed-form expression given in Lemma 2.1 can readily be used to calculate the multiplier bootstrapped version of our CvM test statistic in Section 4.

Lemma 2.1 *Let CvM_n^{dpro} be defined in (2.10) with $\mathbb{S}^{d_x}$ the d_x-dimensional unit sphere. Then, we have*

$$CvM_n^{dpro} = \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} e_i^{pro}(t; \widehat{\theta}_n) e_j^{pro}(t; \widehat{\theta}_n) A_{ijr}, \qquad (2.11)$$

with

$$e_i^{pro}(t; \widehat{\theta}_n) = e_i(t; \widehat{\theta}_n) - g_t\big(X_i, \widehat{\theta}_n\big)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) \frac{1}{n} \sum_{s=1}^{n} g_t(X_s, \widehat{\theta}_n) e_s(t; \widehat{\theta}_n), \quad i = 1, \ldots, n, \quad (2.12)$$

and

$$A_{ijr} = \int_{\mathbb{S}^{d_x}} 1\left(X_i^\top \beta \leq X_r^\top \beta \right) 1\left(X_j^\top \beta \leq X_r^\top \beta \right) d\beta = A_{ijr}^{(0)} \frac{\pi^{d_x/2 - 1}}{\Gamma\left(d_x/2 \right)},$$

where $\Gamma(\cdot)$ is the gamma function, arccos *is the inverse cosine function, and*

$$
A_{ijr}^{(0)} = \begin{cases} 2\pi, & if \ X_i = X_r = X_j, \\ \pi, & if \ X_i = X_j, \ or \ X_i = X_r, \ or \ X_j = X_r, \\ \left| \pi - \arccos\left(\dfrac{(X_i - X_r)^\top (X_j - X_r)}{\|X_i - X_r\| \, \|X_j - X_r\|} \right) \right|, & otherwise. \end{cases}
$$

It is interesting to note that $e_i^{pro}(t; \widehat{\theta}_n)$ for $i = 1, \ldots, n$ in (2.12) are simply the ordinary least squares residuals from regressing $e_i(t; \widehat{\theta}_n)$ on $g_t(X_i, \widehat{\theta}_n)$. Lemma 2.1 yields an explicit formula for our double-projected test statistic CvM_n^{dpro}. It builds on Escanciano (2006) and García-Portugués et al. (2014), who derived expressions for the CvM-type functionals of "single-projected" empirical processes akin to $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$. In fact, it can be shown that the CvM test statistic based on $R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n)$ has the following closed-form expression:

$$
\begin{aligned}
CvM_n^{pro} &= \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n) \right)^2 F_{n,\beta}(du) \, d\beta \\
&= \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^n e_i(t; \widehat{\theta}_n) e_j(t; \widehat{\theta}_n) A_{ijr},
\end{aligned}
$$

which is robust to the dimensionality d_x of X but not robust to the choice of estimator $\widehat{\theta}_n$. With the introduction of the second projection, we are able to obtain a CvM test statistic that is robust to both dimensionality d_x and estimator $\widehat{\theta}_n$.

Just like in Escanciano (2006) and García-Portugués et al. (2014), A_{ijr} represents the surface area of particular spherical regions depending on whether X's are the same across observations: it is the whole sphere $\mathbb{S}^{d_x}$ when $X_i = X_r = X_j$, a hemisphere of $\mathbb{S}^{d_x}$ when $X_i = X_j$ or $X_i = X_r$ or $X_i = X_r$ (but not all X's are the same), or a spherical wedge of width angle given by the third entry of $A_{ijr}^{(0)}$. Thus, as discussed in Escanciano (2006) and García-Portugués et al. (2014), we can express $A_{ijr} = A_{ijr}^{(0)} \pi^{d_x/2 - 1} / \Gamma(d_x/2)$, as in Lemma 2.1 above.

In connection with the expression of CvM_n^{pro}, what is new in Lemma 2.1 is that, instead of computing infinitely many projected residuals, one for each $(\beta^\top, u)^\top \in \Pi_{pro}$ as it is suggested by (2.7), it suffices to use the sequence of projected parametric residuals $\left\{ e_i^{pro}(t; \widehat{\theta}_n), \ i = 1, \ldots n, \ t \in \mathcal{T} \right\}$ as defined in (2.12), which do not depend on the projection direction β. Note that the construction of $e_i^{pro}(t; \widehat{\theta}_n)$ and its intuitive interpretation are only possible with the help of the second projection.

3 Asymptotic results

3.1 Asymptotic null distribution

In this section, we formally investigate the limiting behavior of double-projected generalized residual marked empirical process $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ under the null hypothesis H_0 in (2.1) and consequently that of the test statistic CvM_n^{dpro} based on it.

First, let us denote by $F_X(\cdot)$ the cumulative distribution function (CDF) of covariates X. Also, let $\Psi_{pro}(du, d\beta) = F_\beta(du)d\beta$. Recall that $p_t(x) = \mathbb{P}(T = t | X = x)$ for every $t \in \mathcal{T}$ are the true but unknown GPS.[5] We list all the relevant regularity conditions as follows.

Assumption 3.1 *The random sample $\left\{ \left(X_i^\top, T_i \right)^\top, i = 1 \ldots n \right\}$ consists of a sequence of independent and identically distributed random vectors from $(X^\top, T)^\top$.*

Assumption 3.2 *For each $t \in \mathcal{T} \subseteq \mathbb{N}$, the propensity score model $q_t(X, \theta)$ is known up to the finite-dimensional parameter θ, and is twice continuously differentiable in a neighborhood Θ_0 of θ_0 with $\Theta_0 \subset \Theta$. The score function $g_t(X, \theta) = \partial q_t(X, \theta)/\partial \theta$ satisfies that there exists a $F_X(\cdot)$-integrable function $M(\cdot)$ such that $\sup_{\theta \in \Theta_0} \|g_t(\cdot, \theta)\| \leq M(\cdot)$.*

Assumption 3.3 *(i) The parameter space Θ is a compact subset of $\mathbb{R}^{d_\theta}$; (ii) the true parameter θ_0 belongs to the interior of Θ; and (iii) the estimator $\widehat{\theta}_n$ satisfies that $\left\|\widehat{\theta}_n - \theta_0\right\| = O_p(n^{-1/2})$ under H_0, and $\left\|\widehat{\theta}_n - \theta^*\right\| = o_p(1)$ under H_1 for some θ^* in the parameter space Θ.*

Assumption 3.4 *The integrating function $\Psi_{pro}(\cdot)$ is absolutely continuous with respect to the Lebesgue measure on Π_{pro}.*

Assumptions 3.1-3.4 are standard in the specification testing literature; see, e.g., Escanciano (2006). Note that we only require the root-n consistency of $\widehat{\theta}_n$ in Assumption 3.3(iii) rather than $\widehat{\theta}_n$ being root-n asymptotically linear as required by Escanciano (2006, Assumption A3(b)).

To present our asymptotic results, henceforth, we adopt the following notation. For a generic set $\mathcal{G}$, let $l^\infty(\mathcal{G})$ be the Banach space of all uniformly bounded real functions on $\mathcal{G}$, equipped with the uniform metric $\|f\|_{\mathcal{G}} \equiv \sup_{z \in \mathcal{G}} |f(z)|$. We study the weak convergence of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ and its related processes as elements of $l^\infty(\Pi_{pro})$, where $\Pi_{pro} \equiv \mathbb{S}^{d_x} \times [-\infty, \infty]$ with $\mathbb{S}^{d_x}$ the unit ball in $\mathbb{R}^{d_x}$. Let "$\Longrightarrow$" denote weak convergence on $(l^\infty(\Pi_{pro}), \mathcal{B}_\infty)$ in the sense of J. Hoffmann–Jϕrgensen, where $\mathcal{B}_\infty$ denotes the corresponding Borel σ-algebra - see e.g. Definition 1.3.3 in van der Vaart and Wellner (1996).

---

5 One can alternatively denote $p_t(x)$ as the true but unknown model underlying a conditional moment restriction; see Remark 1.

Then it is shown that under the null $R_{n,t}^{dpro}(\cdot, \cdot; \widehat{\theta}_n) \Longrightarrow R_{\infty,t}^{dpro}$, a centered Gaussian process with its covariance structure defined in Theorem 3.1.

The true generalized error is defined as $\varepsilon(t) = 1(T = t) - p_t(X)$, which satisfies $\mathbb{E}[\varepsilon(t)|X] = 0$ almost surely (a.s.) for each $t \in \mathcal{T}$ regardless of whether the null hypothesis is true. We state formally the asymptotic null behavior of our test statistic CvM_n^{dpro} in the following theorem.

Theorem 3.1 *Suppose Assumptions 3.1-3.4 hold. Then, under the null hypothesis H_0 in (2.1), for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t \in \mathcal{T}} a(t) \leq C < \infty$, we have that*

$$CvM_n^{dpro} \xrightarrow{d} CvM_\infty^{dpro} \equiv \sum_{t \in \mathcal{T}} a(t) \int_{\Pi_{pro}} \left(R_{\infty,t}^{dpro}(\beta, u) \right)^2 \Psi_{pro}(du, d\beta),$$

where $R_{\infty,t}^{dpro}$ is a Gaussian process with mean zero and covariance structure

$$\mathbb{K}_t^{dpro}((\beta, u), (\beta', u')) = \mathbb{E}\left[\sigma_t^2(X) \mathcal{P}_t 1 \left(X^\top \beta \leq u \right) \mathcal{P}_t 1 \left(X^\top \beta' \leq u' \right) \right], \qquad (3.1)$$

where $\sigma_t^2(X) = \mathbb{E}\left[\varepsilon^2(t)|X \right] = p_t(X)(1 - p_t(X)) = q_t(X, \theta_0)(1 - q_t(X, \theta_0))$ is the conditional variance function of generalized error $\varepsilon(t)$ given X under the null.

To prove Theorem 3.1, we first show that the asymptotic null behavior of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ does not depend on $\widehat{\theta}_n$. Based on this result, we combine the weak convergence of the doubly-projected empirical process $R_{n,t}^{dpro}$ with the continuous mapping theorem (see, e.g., van der Vaart and Wellner, 1996, Theorem 1.3.6) to derive the asymptotic distribution of our proposed Cramér–von Mises test statistic CvM_n^{dpro} under the null H_0.

3.2 Asymptotic power

In this section, we study the asymptotic power properties of the CvM_n^{dpro} test statistic based on $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ under the fixed (i.e., global) alternative and a certain sequence of local alternatives converging to the null H_0 at the usual parametric rate. We first consider the fixed alternative hypothesis H_1 in (2.2).

Theorem 3.2 *Suppose Assumptions 3.1-3.4 hold. Then, under the fixed alternative hypothesis H_1 in (2.2), for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t \in \mathcal{T}} a(t) \leq C < \infty$, we have that*

$$\frac{CvM_n^{dpro}}{n} \xrightarrow{p} \sum_{t \in \mathcal{T}} a(t) \int_{\Pi_{pro}} \left(\mathbb{E}\left[(p_t(X) - q_t(X, \theta^*)) \mathcal{P}_t 1 \left(X^\top \beta \leq u \right) \right] \right)^2 \Psi_{pro}(du, d\beta).$$

It follows from Theorem 3.2 that, under the fixed alternative alternative H_1, as long as the unconditional expectation

$$\mathbb{E}\left[(p_t(X) - q_t(X, \theta^*)) \mathcal{P}_t 1 \left(X^\top \beta \leq u \right) \right] \neq 0$$

for some $(\beta^\top, u)^\top$ and for some treatment level $t \in \mathcal{T}$, CvM_n^{dpro} will diverge to positive infinity at the n rate, indicating that CvM_n^{dpro} is able to detect such a fixed alternative with probability tending to one. On the other hand, CvM_n^{dpro} might not be consistent against all fixed alternative hypotheses in (2.2) if $\mathbb{E}\left[(p_t(X) - q_t(X, \theta^*)) \mathcal{P}_t 1\left(X^\top \beta \leq u\right)\right] = 0$ for every $t \in \mathcal{T}$. Specifically, the test statistic cannot distinguish those alternatives such that, for every $t \in \mathcal{T}$, the difference between $p_t(X)$ and $q_t(X, \theta^*)$ is collinear to the score function $g_t(X, \theta^*)$ associated with $q_t(X, \theta^*)$. However, we do not think this type of alternative is of main empirical concern. If one were concerned with this, one could use alternative testing procedures such as Stute (1997), Escanciano (2006), and Domínguez and Lobato (2020).

We now proceed to consider the asymptotic local power properties of our proposed test statistic. Toward this end, we study the asymptotic distribution of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ under a certain sequence of Pitman-type local alternatives converging to the null at a parametric rate. In particular, we consider the data-generating process for the sequence of local alternatives given by

$$H_{1,n} : \mathbb{P}\left[p_t(X) = q_t(X, \theta_0) + \frac{r_t(X)}{\sqrt{n}}\right] = 1 \text{ for some } \theta_0 \in \Theta \in \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \quad (3.2)$$

where, for each $t \in \mathcal{T}$, the direction of departure from H_0 is given by function $r_t(X)$ (potentially different for each t), which is assumed to be $F_X(\cdot)$-integrable with zero mean and satisfy $\mathbb{P}(r_t(X) = 0) < 1$.

Theorem 3.3 *Suppose Assumptions 3.1-3.4 hold. Then, under the sequence of local alternatives $H_{1,n}$ in (3.2), for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t \in \mathcal{T}} a(t) \leq C < \infty$, we have that*

$$CvM_n^{dpro} \overset{d}{\to} CvM_{1,\infty}^{dpro} \equiv \sum_{t \in \mathcal{T}} a(t) \int_{\Pi_{pro}} \left(R_{\infty,t}^{dpro}(\beta, u) + \delta_t(\beta, u)\right)^2 \Psi_{pro}(du, d\beta),$$

where $R_{\infty,t}^{dpro}$ is the same Gaussian process as defined in Theorem 3.1, and δ_t is a deterministic shift function given by

$$\delta_t(\beta, u) = \mathbb{E}\left[r_t(X) \mathcal{P}_t 1\left(X^\top \beta \leq u\right)\right].$$

An immediate consequence of Theorem 3.3 is that whenever there exists some $t \in \mathcal{T}$ such that the deterministic shift function $\delta_t(\beta, u) \neq 0$ for at least some $(\beta^\top, u)^\top \in \Pi_{pro}$ with a positive Lebesgue measure, our proposed Cramér–von Mises test statistic will have non-trivial power against local alternatives of the form (3.2). A pathological situation in which our test will only have trivial local power against such local alternatives is when $r_t(X)$ is a linear combination of score function $g_t(X, \theta_0)$ for every treatment level $t \in \mathcal{T}$, i.e., $r_t(X) = \nu^\top g_t(X, \theta_0)$ *a.s.* for some nonzero vector ν. In such a case, the limiting

distribution of CvM_n^{dpro} under H_0 and $H_{1,n}$ is the same so that $H_{1,n}$ cannot be detected. However, such a specific class of local alternatives is arguably of limited practical interest.

4 A multiplier bootstrap procedure

Since the limiting null distribution of our test statistic CvM_n^{dpro} is non-pivotal, we propose a simple-to-implement multiplier bootstrap procedure to obtain the bootstrapped p-values or critical values and show its asymptotic validity. Below is its implementation:

1. For each $t \in \mathcal{T}$, $i = 1, \ldots n$, generate $e_i^*(t; \widehat{\theta}_n) = V_i \, e_i(t; \widehat{\theta}_n)$, where $\{V_i, i = 1, \ldots, n\}$ is a sequence of independent and identically distributed random variables with mean zero, variance one, and finite third moment; e.g., Rademacher random variables with $\mathbb{P}(V = -1) = \mathbb{P}(V = 1) = 1/2$ (Liu, 1988) or Bernoulli random variable with $\mathbb{P}(V = 1 - \kappa) = \kappa/\sqrt{5}$ and $\mathbb{P}(V = \kappa) = 1 - \kappa/\sqrt{5}$, where $\kappa = (\sqrt{5} + 1)/2$ (Mammen, 1993).

2. Compute

$$\left(CvM_n^{dpro,*}\right)^b = \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^n e_i^{pro,*}(t; \widehat{\theta}_n) e_j^{pro,*}(t; \widehat{\theta}_n) A_{ijr}, \qquad (4.1)$$

where

$$e_i^{pro,*}(t; \widehat{\theta}_n) = e_i^*(t; \widehat{\theta}_n) - g_t^\top(X_i, \widehat{\theta}_n) \Delta_{n,t}^{-1}(\widehat{\theta}_n) \frac{1}{n} \sum_{s=1}^n g_t(X_s, \widehat{\theta}_n) e_s^*(t; \widehat{\theta}_n), \quad i = 1, \ldots, n. \tag{4.2}$$

3. Repeat Steps 1 and 2 B times, and collect $\left\{ \left(CvM_n^{dpro,*}\right)^b, b = 1 \ldots, B \right\}$.

4. Obtain the $(1 - \alpha)$ quantile of $\left\{ \left(CvM_n^{dpro,*}\right)^b, b = 1 \ldots, B \right\}$, $c_{n,\alpha}^*$, and set it as the critical value for the test with significance level α, for $0 < \alpha < 1$.

5. Reject the null hypothesis H_0 in (2.1) if CvM_n^{dpro} is greater than the critical value $c_{n,\alpha}^*$, and fail to reject (2.1) otherwise.

Note that $e_i^{pro,*}(t; \widehat{\theta}_n)$ for $i = 1, \ldots, n$ in (4.2) are simply the ordinary least squares residuals from regressing $e_i^*(t; \widehat{\theta}_n)$ on $g_t(X_i, \widehat{\theta}_n)$ that do not depend on the projection direction β. As such, $e_i^{pro,*}(t; \widehat{\theta}_n)$ are nothing but the bootstrap counterparts of $e_i^{pro}(t; \widehat{\theta}_n)$ in (2.12).

The multiplier bootstrapped test statistic $CvM_n^{dpro,*}$ has attractive theoretical and empirical properties. First, it does not require computing new parameter estimates at each bootstrap draw, reducing the computational intensity of the proposed procedure.

Second, thanks to the use of the second projection in $\mathcal{P}_{n,t}1\left(X_i^\top \beta \le u\right)$, its implementation does not require using estimators that admit an asymptotically linear representation and thus allows for a wider range of estimators. Third, thanks to the closed-form representation in (2.11), A_{ijr} does not need to be computed for each bootstrap sample. Finally, it does not involve any tuning parameters such as bandwidths. All these features greatly alleviate the computational cost of our testing procedure.

To establish the asymptotic validity of the proposed multiplier bootstrap procedure described in Steps 1–5 above, let

$$R_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i \, e_i(t; \widehat{\theta}_n) \mathcal{P}_{n,t}1\left(X_i^\top \beta \le u\right)$$

denote the multiplier bootstrapped version of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$, with the sequence of multiplies $\{V_i\}_{i=1}^n$ as described in Step 1. In addition, the multiplier bootstrapped version of CvM_n^{drpo} is

$$CvM_n^{dpro,*} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(R_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n)\right)^2 F_{n,\beta}(du) \, d\beta,$$

whose closed-form expression is given by (4.1). It is shown in Appendix A that

$$R_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i \, e_i(t; \theta_0) \mathcal{P}_t1\left(X_i^\top \beta \le u\right) + o_p(1),$$

uniformly in $(\beta^\top, u)^\top$ for each $t \in \mathcal{T}$. Thus, $R_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n)$ is asymptotically invariant to $\widehat{\theta}_n$, and this holds under H_0, H_1, and $H_{1,n}$. The next theorem formally states the asymptotic validity of the multiplier bootstrap procedure.

Theorem 4.1 *Suppose Assumptions 3.1-3.4 hold. Then, we have $CvM_n^{dpro,*} \overset{d}{\underset{*}{\to}} CvM_\infty^{dpro}$ in probability under the bootstrap law, where CvM_∞^{dpro} is the same distribution as defined in Theorem 3.1, and $\overset{d}{\underset{*}{\to}}$ denotes weak convergence under the bootstrap law, i.e., conditional on the original sample $\left\{(T_i, X_i^\top)^\top\right\}_{i=1}^n$.*

Theorem 4.1 states that the bootstrapped test statistic $CvM_n^{dpro,*}$ converges to the null limiting distribution of CvM_n^{dpro} conditional on the original sample. The fact that $CvM_n^{dpro,*}$ has the same limiting distribution under H_0, H_1, and $H_{1,n}$ is what allows the proposed multiplier bootstrap procedure to work.

Remark 1 *Our proposed double-projection procedure can be readily used to test the validity of conditional moment restrictions $\mathbb{E}\left[e(t; \theta_0)|X\right] = 0$ a.s. for all $t \in \mathcal{T}$, for an appropriately defined generalized error $e(t; \theta_0)$. This includes classical regression models, where $\mathcal{T} = \{1\}$ the outcome of interest in Y, the parametric model for $\mathbb{E}[Y|X]$ is $q(X, \theta)$, and $e(1; \theta) = Y - q(X, \theta)$. One can also test conditional mean and variance*

models in the spirit of *Escanciano (2008)*, where $\mathcal{T} = \{1, 2\}$, $e(1; \theta) = Y - q_1(X, \theta_1)$, $e(2; \theta) = (Y - q_1(X, \theta_1))^2 - q_2(X, \theta_2)$, with $q_1(X, \theta_1)$ being a parametric model for $\mathbb{E}[Y|X]$, and $q_2(X, \theta_2)$ a parametric model for the conditional variance of $Y - \mathbb{E}[Y|X]$. In these cases, all one needs to do is adjust the definition of the errors and score functions. Importantly, we do not require that parametric errors (and the outcome variables) to be continuous.

We also stress that it is straightforward to extend our proposal to test even more general conditional moment restrictions of the following form:

$$H_0^{cmr} : \mathbb{P}\left(\mathbb{E}[\rho(Z, \theta_0)|X] = 0\right) = 1 \ \text{for some} \ \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta},$$
$$H_1^{cmr} : \mathbb{P}\left(\mathbb{E}[\rho(Z, \theta)|X] = 0\right) < 1 \ \text{for any} \ \theta \in \Theta \subset \mathbb{R}^{d_\theta},$$

where $\rho(z, \theta)$ is a known function, not necessarily differentiable with respect to θ; see, e.g., *Whang (2001)*, *Delgado et al. (2006)*, and *Escanciano and Jacho-Chávez (2010)*. A special case is when $Z = (Y, X)$, $\rho(Z, \theta) = Y - q(X, \theta)$ for some parametric function $q(X, \theta)$, and Y is the dependent variable, as we just discussed above. Another interesting case is when $\rho(Z, \theta) = 1(Y - q(X, \theta) \leq 0) - \tau$ with $\tau \in (0, 1)$; then $q(X, \theta_0)$ is the τ-quantile of the conditional distribution of Y given X. For the general case, however, the dependent variable may be related to the regressors through the implicit function $\rho(\cdot, \cdot)$. Then, by letting $e(\theta) = \rho(Z, \theta)$ and $g(X, \theta) = \partial\mathbb{E}[\rho(Z, \theta)|X]/\partial\theta$, we can straightforwardly generalize our double-projection methodology to test H_0^{cmr} against H_1^{cmr}.

5 Specification tests for multiple-index models

In the previous sections, we proposed omnibus-type specification tests that aim to detect the inadequacy of general parametric models. In particular, the working parametric models do not assume any dimension-reducing structure under the null and/or under the alternative. In many applications, however, using models with single/multiple index structures is common. In such cases, getting some additional insights about potential reasons for rejecting the putative model is interesting.

There has been a large amount of interest in statistical inference in (single) index models. Horowitz and Härdle (1994), Fan and Li (1996), Fan and Liu (1997), Aït-Sahalia et al. (2001), Härdle et al. (2001), Stute and Zhu (2002), Xia et al. (2004), Stute and Zhu (2005), Xia (2009), Escanciano and Song (2010), Guo et al. (2016) and Maistre and Patilea (2019), among others, proposed various approaches to testing generalized linear models or testing parametric or semiparametric (single) index assumptions. In the following, we investigate three relevant testing problems in the framework of index-type GPS. In Section 5.1, we briefly describe a method for testing an index model (IM hereafter) with a known link function against a nonparametric alternative (i.e., not restricting ourselves to an IM

with a known link function). To further uncover the potential sources of misspecification for the assumed IM, in Section 5.2, we discuss the problem of directionally testing an IM with a known link function against a semiparametric alternative of IM with an unknown link function. Section 5.3 outlines how to test a semiparametric single-index model against a nonparametric alternative. We emphasize that the test we discuss in Section 5.3 is very different from all other tests discussed in the paper as the null model is semiparametric, i.e., it involves an infinite-dimensional nuisance function.

5.1 Testing a parametric IM against a nonparametric alternative

For practical convenience, researchers often impose some dimension-reducing structure on the GPS. When the treatment is binary or ordered multinomial, a popular choice is the class of single-index models, namely, $q_t(X, \theta) = q_t(X^\top \theta)$, where $X^\top \theta$ is the single linear-index, and $q_t(\cdot) : \mathbb{R} \mapsto [0, 1]$ is a known link function for each $t \in \mathcal{T}$ (e.g., the multinomial logit link with $q_t(\cdot)$ specified as the cumulative logistic distribution function or the probit link with $q_t(\cdot)$ specified as the cumulative normal distribution function).[6] With unordered multinomial choices, one usually adopts a multiple index model, $q_t(X, \theta) = q_t(X^\top \theta_1, \ldots, X^\top \theta_J)$ as in the multinomial logit model (2.5). Since the binary and ordered multinomial cases are special cases of the unordered multinomial one, we focus on the latter.

For a generic $\theta = (\theta_1^\top, \ldots, \theta_J^\top)^\top$, let $\widetilde{X}_\theta = (X^\top \theta_1, \ldots, X^\top \theta_J)$ denote the vector of linear indexes. In this section, the null hypothesis of interest is

$$H_0^{im} : \mathbb{P}\left(\mathbb{E}\left[e^{im}(t; \theta_0)|X\right] = 0\right) = 1 \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \qquad (5.1)$$

where $\theta_0 = (\theta_{0,1}^\top, \ldots, \theta_{0,J}^\top)^\top$, and $e^{im}(t; \theta) \equiv 1(T = t) - q_t(\widetilde{X}_\theta)$. The alternative hypothesis H_1^{im} is the negation of H_0^{im}. This is an omnibus test of a multiple-index parametric model against a nonparametric alternative.

Testing H_0^{im} against H_1^{im} is equivalent to testing (2.1) against (2.2) but with $q_t(X, \theta) = q_t(\widetilde{X}_\theta)$ and the generalized error $e(t; \theta)$ replaced by $e^{im}(t; \theta) \equiv 1(T = t) - q_t(\widetilde{X}_\theta)$. As such, we can consider the following generalized residual marked double-projected empirical process as a special case of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{im}(t; \widehat{\theta}_n) \mathcal{P}_{n,t}^{im} 1(X_i^\top \beta \leq u), \quad \left(\beta^\top, u\right)^\top \in \Pi_{pro},$$

where $e_i^{im}(t; \widehat{\theta}_n) \equiv 1(T_i = t) - q_t(\widetilde{X}_{i,\widehat{\theta}_n})$ is the generalized residual under the multiple index model and the projection operator $\mathcal{P}_{n,t}^{im} 1(X_i^\top \beta \leq u)$ is as defined in (2.7) but with

---

6 The multinomial ordered discrete choice model has an intercept that varies with t.

the score function $g_t(x, \theta) = \partial q_t(\widetilde{x}_\theta)/\partial\theta$ for each $t \in \mathcal{T}$. The associated theoretical results are omitted as they are the same as those in previous sections.

5.2 Testing a parametric IM against a semiparametric alternative

In this section, we discuss how one can directionally test a parametric index model against a semiparametric index model, where the semiparametric component comes from link functions being unknown. Formally, we want to test the null hypothesis

$$H_0^{im1} : \mathbb{P}\left(\mathbb{E}\left[e^{im}(t; \theta_0)|\widetilde{X}_{\theta_0}\right] = 0\right) = 1 \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \quad (5.2)$$

against the directional alternative

$$H_1^{im1} : \mathbb{P}\left(\mathbb{E}\left[1(T = t) - \mu_t(\widetilde{X}_\theta)|\widetilde{X}_\theta\right] = 0\right) = 1 \text{ for any } \theta \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and/or some } t \in \mathcal{T}, \quad (5.3)$$

where, $\mu_t(\cdot) : \mathbb{R}^J \mapsto [0, 1]$ is an unknown link function such that $\mathbb{P}\left(\mu_t(\widetilde{X}_\theta) = q_t(\widetilde{X}_\theta)\right) < 1$ for any $\theta \in \Theta$ and/or some $t \in \mathcal{T}$.

Horowitz and Härdle (1994) and Fan and Liu (1997) considered similar testing problems specialized to the single-index setup and constructed their test statistics by comparing a nonparametric estimate of $\mathbb{P}(T = t|X^\top\widehat{\theta}_n = v)$ with $q_t(v)$. We pursue a different route by using the integrated moment approach with double-projections and allowing for multiple-index models.

The key insight we provide here is to re-express the null hypothesis (5.2) as

$$H_0^{im1} : \mathbb{E}\left[e^{im}(t; \theta_0)1\left(\widetilde{X}_{\theta_0}^\top\beta \leq u\right)\right] = 0 \text{ a.e. } \left(\beta^\top, u\right)^\top \in \Pi_{pro}^{im},$$
$$\text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \quad (5.4)$$

where $\Pi_{pro}^{im} \equiv \mathbb{S}^J \times \overline{\mathbb{R}}$ denotes the projected space under the index model. In light of our previous discussions, this characterization of the null immediately suggests using the following generalized residual marked double-projected empirical process with *estimated* multiple indexes,

$$M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t; \widehat{\theta}_n)\mathcal{P}_{n,t}^{im}1(\widetilde{X}_{i,\widehat{\theta}_n}^\top\beta \leq u), \quad \left(\beta^\top, u\right)^\top \in \Pi_{pro}^{im},$$

where

$$\mathcal{P}_{n,t}^{im}1(\widetilde{X}_{i,\widehat{\theta}_n}^\top\beta \leq u) \equiv 1(\widetilde{X}_{i,\widehat{\theta}_n}^\top\beta \leq u) - g_t(X_i, \widehat{\theta}_n)^\top\Delta_{n,t}^{-1}(\widehat{\theta}_n)G_{n,t}^{im}(\beta, u; \widehat{\theta}_n), \quad (5.5)$$

with $g_t(X_i, \widehat{\theta}_n)$ and $\Delta_{n,t}(\widehat{\theta}_n)$ defined as before, and

$$G_{n,t}^{im}(\beta, u; \widehat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} g_t(X_i, \widehat{\theta}_n) 1(\widetilde{X}_{i,\widehat{\theta}_n}^{\top} \beta \leq u).$$

Based on $M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$, we can use the following class of CvM-type test statistics to test H_0^{im1} against H_1^{im1}:

$$CvM_n^{im,dpro} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}^{im}} \left(M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \right)^2 F_{n,\beta,\widehat{\theta}_n}^{im}(du) \, d\beta, \tag{5.6}$$

where $F_{n,\beta,\widehat{\theta}_n}^{im}(u) = n^{-1} \sum_{i=1}^{n} 1\left(\widetilde{X}_{i,\widehat{\theta}_n}^{\top} \beta \leq u \right)$ is the empirical distribution function of the one-dimensional projected estimated indexes $\left\{ \widetilde{X}_{i,\widehat{\theta}_n}^{\top} \beta \right\}_{i=1}^{n}$ for any fixed projected direction $\beta \in \mathbb{S}^J$, and $d\beta$ is the rescaled uniform density on the unit sphere $\mathbb{S}^J$.

Note that $CvM_n^{im,dpro}$ in (5.6) resembles CvM_n^{dpro} in (2.10), with the caveat that now we use the vector of estimated linear indexes $\left\{ \widetilde{X}_{i,\widehat{\theta}_n} \right\}_{i=1}^{n}$, instead of the observed covariates $\{X_i\}_{i=1}^{n}$ in the projection steps. As this difference changes the computational aspect of the test statistic and also its asymptotic properties, we discuss these differences in the following.

In terms of computation, by building on Lemma 2.1 we can show that $CvM_n^{im,dpro}$ in (5.6) can be expressed in closed-form as

$$CvM_n^{im,dpro} = \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} e_i^{im,pro}(t; \widehat{\theta}_n) e_j^{im,pro}(t; \widehat{\theta}_n) A_{ijr}^{im}, \tag{5.7}$$

where

$$e_i^{im,pro}(t; \widehat{\theta}_n) = e_i^{im}(t; \widehat{\theta}_n) - g_t(X_i, \widehat{\theta}_n)^{\top} \Delta_{n,t}^{-1}(\widehat{\theta}_n) \frac{1}{n} \sum_{s=1}^{n} g_t(X_s, \widehat{\theta}_n) e_s^{im}(t; \widehat{\theta}_n), \quad i = 1, \ldots, n, \tag{5.8}$$

and $A_{ijr}^{im} = A_{ijr}^{im(0)} \times \left(\pi^{J/2-1} \right) / \Gamma(J/2)$ with

$$A_{ijr}^{im(0)} = \begin{cases} 2\pi, & if \quad \widetilde{X}_{i,\widehat{\theta}_n} = \widetilde{X}_{r,\widehat{\theta}_n} = \widetilde{X}_{j,\widehat{\theta}_n}, \\ \pi, & if \quad \widetilde{X}_{i,\widehat{\theta}_n} = \widetilde{X}_{j,\widehat{\theta}_n}, \ or \ \widetilde{X}_{i,\widehat{\theta}_n} = \widetilde{X}_{r,\widehat{\theta}_n}, \ or \ \widetilde{X}_{j,\widehat{\theta}_n} = \widetilde{X}_{r,\widehat{\theta}_n}, \\ \pi - \arccos\left(\frac{\left(\widetilde{X}_{i,\widehat{\theta}_n} - \widetilde{X}_{r,\widehat{\theta}_n} \right)^{\top} \left(\widetilde{X}_{j,\widehat{\theta}_n} - \widetilde{X}_{r,\widehat{\theta}_n} \right)}{\left\| \widetilde{X}_{i,\widehat{\theta}_n} - \widetilde{X}_{r,\widehat{\theta}_n} \right\| \left\| \widetilde{X}_{j,\widehat{\theta}_n} - \widetilde{X}_{r,\widehat{\theta}_n} \right\|} \right), & otherwise. \end{cases}$$

This characterization helps us implement the multiplier bootstrap procedure and on other practical computational aspects without needing to compute a continuum of projections.

In terms of the statistical properties of our test statistic $CvM_n^{im,dpro}$, if we show that there is no estimation effect from using the estimated indexes in the weighting functions, we can resort to Theorem 3.1 to establish size control. Indeed, as we show in Appendix

A, we have that, uniformly in $(\beta^\top, u)^\top \in \Pi_{pro}^{im}$ and for each $t \in \mathcal{T}$,

$$
\begin{aligned}
& M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \\
={} & \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{im}(t; \widehat{\theta}_n) \left(1(\widetilde{X}_{i,\theta_0}^\top \beta \le u) - g_t(X_i, \widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) G_{n,t}^{im}(\beta, u; \widehat{\theta}_n) \right) + o_p(1) \\
={} & \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{im}(t; \theta_0) \left(1(\widetilde{X}_{i,\theta_0}^\top \beta \le u) - g_t(X_i, \theta_0)^\top \Delta_t^{-1}(\theta_0) G_t^{im}(\beta, u; \theta_0) \right) + o_p(1) \\
\equiv{} & \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{im}(t; \theta_0) \mathcal{P}_t^{im} 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u) + o_p(1),
\end{aligned}
$$

directly leading to the following result.

Proposition 5.1 *Suppose Assumptions 3.1-3.3 hold. Consider a multiple-index specification, $q_t(X, \theta) = q_t(\widetilde{X}_\theta)$ for all $t \in \mathcal{T}$, such that $q_t(\cdot)$ is known up to the finite-dimensional parameters θ. Assume that the underlying distribution of the projected-multiple indexes, $F_{\beta,\theta_0}(\cdot)$, is absolutely continuous with respect to the Lebesgue measure. Then, under the null hypothesis H_0^{im1} in (5.2), for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t \in \mathcal{T}} a(t) \le C < \infty$, we have that*

$$
CvM_n^{im,dpro} \xrightarrow{d} CvM_\infty^{im,dpro} \equiv \sum_{t \in \mathcal{T}} a(t) \int_{\Pi_{pro}^{im}} \left(M_{\infty,t}^{dpro}(\beta, u) \right)^2 F_{\beta,\theta_0}(du) d\beta,
$$

where $M_{\infty,t}^{dpro}$ is a Gaussian process with mean zero and covariance structure

$$
\mathbb{K}_t^{im,dpro}\left((\beta, u), (\beta', u')\right) = \mathbb{E}\left[\widetilde{\sigma}_t^2(\widetilde{X}_{\theta_0}) \mathcal{P}_t^{im} 1\left(\widetilde{X}_{\theta_0}^\top \beta \le u \right) \mathcal{P}_t^{im} 1\left(\widetilde{X}_{\theta_0}^\top \beta' \le u' \right) \right], \quad (5.9)
$$

where $\widetilde{\sigma}_t^2(\widetilde{X}_{\theta_0}) = \mathbb{E}\left[\varepsilon^2(t) | \widetilde{X}_{\theta_0} \right] = q_t(\widetilde{X}_{\theta_0})\left(1 - q_t(\widetilde{X}_{\theta_0}) \right)$ is the conditional variance function of the generalized error $\varepsilon(t)$ given $\widetilde{X}_{\theta_0}$ under the null H_0^{im1}.

In order to study the power properties of our proposed test, recall that $p_t(x) = \mathbb{P}(T = t | X = x)$ is the true, unknown GPS, which does not have to satisfy the semiparametric index restriction stated in the alternative hypothesis H_1^{im1}. To proceed with the analysis of the asymptotic global power properties of $CvM_n^{im,dpro}$, we can show that, uniformly in $(\beta^\top, u)^\top \in \Pi_{pro}^{im}$,

$$
\frac{M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)}{\sqrt{n}} \xrightarrow{p} \mathbb{E}\left[\left(p_t(X) - q_t(\widetilde{X}_{\theta^*}) \right) \mathcal{P}_t^{im} 1\left(\widetilde{X}_{\theta^*}^\top \beta \le u \right) \right],
$$

for each $t \in \mathcal{T}$ under H_1^{im1}, where, using obvious notation, $\mathcal{P}_t^{im} 1\left(\widetilde{X}_{\theta^*}^\top \beta \le u \right)$ is the probability limit of (5.5).

If we ignore the less interesting class of alternative hypotheses under which the difference between $q_t(\widetilde{X}_{\theta^*})$ and $\mu_t(\widetilde{X}_{\theta^*})$ is collinear with the score function $g_t(X, \theta^*)$, we have

that

$$\mathbb{E}\left[\left(p_t(X) - q_t(\widetilde{X}_{\theta^*})\right)\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}^\top\beta \le u)\right]$$
$$=\mathbb{E}\left[\left(p_t(X) - \mu_t(\widetilde{X}_{\theta^*})\right)\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}^\top\beta \le u)\right]$$
$$+ \mathbb{E}\left[\left(\mu_t(\widetilde{X}_{\theta^*}) - q_t(\widetilde{X}_{\theta^*})\right)\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}^\top\beta \le u)\right].$$

On one hand, when $p_t(x) = \mu_t(\tilde{x}_{\theta^*})$ holds (i.e., the true GPS satisfies the semiparametric IM assumption under H_1^{im1}), the fact that $q_t(v) \ne \mu_t(v)$ would be able to guarantee $\mathbb{E}\left[\left(\mu_t(\widetilde{X}_{\theta^*}) - q_t(\widetilde{X}_{\theta^*})\right)\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}^\top\beta \le u)\right] \ne 0$, thus implying that our test is consistent against H_1^{im1} in (5.3).

On the other hand, if the true model does not satisfy an index restriction, i.e., $p_t(x) \ne \mu_t(\tilde{x}_{\theta^*})$, it is possible that $\mathbb{E}\left[\left(p_t(X) - q_t(\widetilde{X}_{\theta^*})\right)\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}^\top\beta \le u)\right] = 0$, even if the null hypothesis H_0^{im1} is false. As so, tests based on $M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ are inconsistent against the more general class of alternative hypotheses consisting of the negation of (5.2). These findings, however, are typical features of directional-type tests; see, e.g., Horowitz and Härdle (1994). It is also straightforward to show that our test $CvM_n^{im,dpro}$ is able to detect a broad range of local alternatives, similar to the discussion in Theorem 3.3. We omit the details to avoid repetition.

In practice, one can leverage the closed-form representation of our test statistic in (5.7) to compute critical values and/or p-values with the assistance of a convenient multiplier bootstrap procedure. Similar to Theorem 4.1, we can show that to compute the bootstrap analog of $CvM_n^{im,dpro}$, all one needs to do is replace $e_i^{im,pro}(t; \widehat{\theta}_n)$ in (5.7) with its "perturbed" version,

$$e_i^{im,pro,*}(t; \widehat{\theta}_n) = e_i^{im,*}(t; \widehat{\theta}_n) - g_t(X_i, \widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^n g_t(X_s, \widehat{\theta}_n)e_s^{im,*}(t; \widehat{\theta}_n), \quad (5.10)$$

where $e_i^{im,*}(t; \widehat{\theta}_n) = V_i\, e_i^{im}(t; \widehat{\theta}_n)$, $i = 1, \ldots, n$, and $\{V_i, i = 1, \ldots, n\}$ is a sequence of i.i.d. random variables with mean zero, variance one, and finite third moment. Given our results in Theorem 4.1, it is easy to show that the multiplier bootstrap analog of $CvM_n^{im,dpro}$ converges to the null distribution of $CvM_n^{im,dpro}$ conditional on the original sample, establishing its asymptotic validity. We omit the details to avoid repetition of arguments.

5.3 Testing a semiparametric single-index model against a nonparametric alternative

So far, we have focused our attention on testing parametric models against omnibus or directional alternatives. Suppose that a researcher rejects a putative parametric IM, and

then decides to use a more flexible semiparametric model while still maintaining the index restriction. By doing so, the researcher can still alleviate the "curse of dimensionality". In this context, however, a natural question arises: is the semiparametric index model correctly specified? In this section, we describe how one can use a variant of our double-projection procedure to answer this type of question. Given that the setup here is very different from the rest of the paper, we restrict our attention to single-index models (SIM). Extensions for multiple-index models are possible but involve much more tedious arguments and steps.

More precisely, we are now interested in testing the following semiparametric-type null hypothesis

$$H_0^{sim} : \ \mathbb{P}\left(\mathbb{E}\left[e^{sim}(t;\theta_0,\mu_t)|X\right] = 0\right) = 1$$
$$\text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and some } \mu_t(\cdot) \in \mathcal{C}_k, \text{ and all } t \in \mathcal{T}, \qquad (5.11)$$

where $e_i^{sim}(t;\theta_0,\mu_t) \equiv 1(T_i = t) - \mu_t(X_i^\top \theta_0)$, and $\mu_t(\cdot) : \mathbb{R} \mapsto [0,1]$ is an unknown link function that belongs to $\mathcal{C}_k$, the space of real-valued, continuous (measurable) functions with bounded k-th derivative, $k > 1$. The alternative hypothesis H_1^{sim} is the negation of H_0^{sim}, i.e,

$$H_1^{sim} : \ \mathbb{P}\left(\mathbb{E}\left[e^{sim}(t;\theta,\tilde{\mu}_t)|X\right] = 0\right) < 1,$$
$$\text{for any } \theta \in \Theta \subset \mathbb{R}^{d_\theta}, \text{ any } \tilde{\mu}_t(\cdot) \in \mathcal{C}_k, \text{ and/or some } t \in \mathcal{T}. \qquad (5.12)$$

Henceforth, we assume that the linear combination $X^\top \theta_0$ admits a probability density function (PDF) $f(\cdot)$. Let $K(v)$ be the univariate kernel function and h be a bandwidth parameter that shrinks to zero at an appropriate rate as $n \to \infty$. In order to test the above semiparametric conditional moment restriction, in light of our previous discussion and Escanciano (2006),[7] it is natural first to consider the following density-weighted generalized residual marked projected empirical process:

$$S_{n,t}^{pro}(\beta,u;\widehat{\theta}_n,\widehat{\mu}_t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{e}_i^{sim}(t;\widehat{\theta}_n,\widehat{\mu}_t)\widehat{f}(X_i^\top \widehat{\theta}_n)1(X_i^\top \beta \le u), \quad \left(\beta^\top, u\right)^\top \in \Pi_{pro},$$

where $\widehat{\theta}_n$ is any $\sqrt{n}$-consistent estimator for θ_0, say, the minimum average conditional variance estimator of Xia et al. (2002) and Xia (2006), $\widehat{e}_i^{sim}(t;\widehat{\theta}_n,\widehat{\mu}_t) \equiv 1(T_i = t) - \widehat{\mu}_t(X_i^\top \widehat{\theta}_n)$ is the semiparametric generalized residual,

$$\widehat{\mu}_t(X_i^\top \widehat{\theta}_n) = \frac{\frac{1}{n-1}\sum_{j \ne i}^n \frac{1}{h} K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h}\right) 1(T_j = t)}{\widehat{f}(X_i^\top \widehat{\theta}_n)}$$

---

7 We stress that Escanciano (2006) does not consider tests for semiparametric models like the one being discussed here.

and

$$\widehat{f}(X_i^\top \widehat{\theta}_n) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{h} K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h}\right)$$

are the standard leave-one-out kernel estimates for single-index model $\mu_t(X_i^\top \theta_0) = \mathbb{E}[1(T_i = t)|X_i^\top \theta_0]$ and the PDF $f(X_i^\top \theta_0)$, respectively.[8] Density-weighted statistics like $S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ avoid random denominators, which simplifies the large sample derivations and also usually translates to better finite sample properties.

Clearly, $S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ satisfies the following decomposition:

$$S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1(T_i = t) - \mu_t(X_i^\top \widehat{\theta}_n)\right) \widehat{f}(X_i^\top \widehat{\theta}_n) 1(X_i^\top \beta \leq u)$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\widehat{\mu}_t(X_i^\top \widehat{\theta}_n) - \mu_t(X_i^\top \widehat{\theta}_n)\right) \widehat{f}(X_i^\top \widehat{\theta}_n) 1(X_i^\top \beta \leq u)$$

$$\equiv S_{n,t1}^{pro}(\beta, u; \widehat{\theta}_n, \mu_t) - S_{n,t2}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t),$$

where $S_{n,t1}^{pro}(\beta, u; \widehat{\theta}_n, \mu_t)$ is affected by the finite-dimensional estimator of $\widehat{\theta}_n$ but not by the infinite-dimensional estimator $\widehat{\mu}_t$, whereas $S_{n,t2}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ is affected by both. By working with each of these terms separately, we show in Appendix A that, under H_0^{sim}, uniformly in $(\beta^\top, u)^\top \in \Pi_{pro}$,

$$S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{sim}(t; \theta_0, \mu_t) f(X_i^\top \theta_0) \left(1(X_i^\top \beta \leq u) - \mathbb{E}[1(X_i^\top \beta \leq u)|X_i^\top \theta_0]\right)$$
$$- \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \mathbb{E}\left[\mu_t'(X^\top \theta_0) f(X^\top \theta_0) \left(X - \mathbb{E}[X|X^\top \theta_0]\right) 1(X^\top \beta \leq u)\right] + o_p(1), \quad (5.13)$$

where $\mu_t'(v) = d\mu_t(v)/dv$ is the (unknown) first derivative of $\mu_t(v)$ for each $t \in \mathcal{T}$. It is interesting to remark that the quantity

$$\sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \mathbb{E}\left[\mu_t'(X^\top \theta_0) f(X^\top \theta_0) \left(X - \mathbb{E}[X|X^\top \theta_0]\right) 1(X^\top \beta \leq u)\right] \quad (5.14)$$

can be regarded as the *total* or *overall* parametric-type estimation effect due to using $\widehat{\theta}_n$ to estimate θ_0, while the quantity

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i^{sim}(t; \theta_0, \mu_t) f(X_i^\top \theta_0) \mathbb{E}\left[1(X_i^\top \beta \leq u)|X_i^\top \theta_0\right] \quad (5.15)$$

can be regarded as the *nonparametric*-type estimation effect due to using the nonparametric estimator $\widehat{\mu}_t(v)$ to replace $\mu_t(v)$ if the linear-index $v = x^\top \theta_0$ were known.

---

8 The kernel density estimate $\widehat{f}(X_i^\top \widehat{\theta}_n)$ is employed in $S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ to avoid random denominators and ease theoretical derivations; see, e.g., Powell et al. (1989) and Delgado and González Manteiga (2001).

To formally establish (5.13), we need the following assumptions, on top of Assumptions 3.1 and 3.3.

Assumption 5.1 *The semiparametric propensity score model $\mu_t(v)$ is k-times continuously differentiable in v with uniformly bounded derivatives, i.e., $\sup_v |d^j q_t(v)/dv^j| < C_j$ with positive constant C_j for $j = 1, \ldots, k$, $k \geq 1$.*

Assumption 5.2 *The kernel function $K(v)$ is bounded, symmetric around zero, and M-times continuously differentiable in v with uniformly bounded derivatives ($M \geq 5$ is an integer). In addition, $K(v)$ satisfies $\int K(v)dv = 1$, $\int v^j K(v) = 0$ for $j = 1, \ldots, k-1$, $\int v^k K(v) \neq 0$, $\int K^2(v)dv < \infty$, $\int |v|^k |K(v)|dv < \infty$, and $|v|^k |K(v)| \to 0$ as $|v| \to \infty$.*

Assumption 5.3 *The positive bandwidth sequence $h = h_n$ satisfies $h \to 0$, $nh^3 \to \infty$ and $nh^{2k} \to 0$ as $n \to \infty$.*

Assumption 5.1 assumes the link function is smooth, while Assumption 5.2 requires that $K(v)$ is a higher-order kernel. These are common assumptions in the literature on nonparametric kernel estimation. Assumption 5.3 is a standard assumption on the bandwidth that balances between biases and variances of U-processes, which appear in the proof. In particular, $nh^3 \to \infty$ is a condition used in proving that the second order terms of U-processes are asymptotically uniformly negligible, while the condition $nh^{2k} \to 0$ is used to control the bias terms. Note that if the standard normal kernel is used, it is differentiable to any order and $k = 2$. Although our assumptions may be stronger than necessary, they make the proofs much easier.

Given the linear representation in (5.13), it is straightforward to establish the asymptotic null distribution of $S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ when $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ admits an asymptotically linear representation. Even in that case, it is clear that its asymptotic null distribution would depend on how $\widehat{\theta}_n$ is obtained. Based on these more demanding and arguably less-desirable conditions, it is easy to build on (5.13) to propose a CvM-type test, establish its power properties and use a multiplier-bootstrap akin to the one discussed in Section 4 to compute asymptotically-valid critical values; see, e.g., Xia et al. (2004) for a related proposal that relies on the standard indicator weight $1(X \leq \tilde{x})$ instead of the one-dimensional projected indicator weight $1(X^\top \beta \leq u)$.[9]

Instead of following this "more-traditional" path, one may propose a test insensitive to the choice of estimators for the nuisance parameters. In what follows, we argue that using a double-projection argument like those introduced in Section 2.2, we can eliminate the parametric-type estimation effect (5.14), but not the nonparametric-type of estimation effect (5.15). Still, using the double-projection procedure allows researchers to consider

---

9 Their bootstrap procedure also requires outcomes to be continuous, ruling out discrete response variables such as in the (generalized) propensity score models considered here.

a wider variety of estimators for θ_0 than if they were to propose test-statistics based on $S_{n,t}^{pro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$.

To see this, let the (infeasible) double-projection weights be defined as

$$\mathcal{P}_t^{sim}1(X_i^\top \beta \leq u) \equiv 1(X_i^\top \beta \leq u) - \mu_t'(X_i^\top \theta_0)f(X_i^\top \theta_0)\eta(X_i, \theta_0)^\top \Delta_t^{-1}(\theta_0)G_t(\beta, u; \theta_0),$$

where $\eta(x, \theta) = x - \varphi(x^\top \theta)$ with $\varphi(v) = \mathbb{E}[X|X^\top \theta = v]$,

$$\Delta_t(\theta) = \mathbb{E}[\mu_t'(X^\top \theta)^2 f(X^\top \theta)^2 \eta(X, \theta)\eta(X, \theta)^\top],$$

and

$$G_t(\beta, u; \theta) = \mathbb{E}[\mu_t'(X^\top \theta)f(X^\top \theta)\eta(X, \theta)1(X^\top \beta \leq u)].$$

The feasible version of $\mathcal{P}_t^{sim}1(X_i^\top \beta \leq u)$ is given by

$$\mathcal{P}_{n,t}^{sim}1(X_i^\top \beta \leq u) \equiv 1(X_i^\top \beta \leq u) - \widehat{\mu}_t'(X_i^\top \widehat{\theta}_n)\widehat{f}(X_i^\top \widehat{\theta}_n)\widehat{\eta}(X_i, \widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n)G_{n,t}(\beta, u; \widehat{\theta}_n), \quad (5.16)$$

where

$$\widehat{\mu}_t'(X_i^\top \widehat{\theta}_n) = \frac{\frac{1}{(n-1)^2 h^3}\sum_{j\neq i}^n \sum_{j'\neq i}^n K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h}\right)K^{(1)}\left(\frac{(X_i - X_{j'})^\top \widehat{\theta}_n}{h}\right)(1(T_{j'} = t) - 1(T_j = t))}{\widehat{f}(X_i^\top \widehat{\theta}_n)^2}$$

is the leave-one-out kernel estimator for $\mu_t'(X_i^\top \theta_0) \equiv d\mu_t(v)/dv|_{v=X_i^\top \theta_0}$, $\widehat{\eta}(X_i, \widehat{\theta}_n) = X_i - \widehat{\varphi}(X_i^\top \widehat{\theta}_n)$ with $\widehat{\varphi}(X_i^\top \widehat{\theta}_n)$ the leave-one-out kernel estimator for $\varphi(X_i^\top \theta_0)$ is given by

$$\widehat{\varphi}(X_i^\top \widehat{\theta}_n) = \frac{\frac{1}{(n-1)h}\sum_{j\neq i}^n K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h}\right)X_j}{\widehat{f}(X_i^\top \widehat{\theta}_n)},$$

and

$$\Delta_{n,t}(\widehat{\theta}_n) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}_t'(X_i^\top \widehat{\theta}_n)^2\widehat{f}(X_i^\top \widehat{\theta}_n)^2\widehat{\eta}(X_i, \widehat{\theta}_n)\widehat{\eta}(X_i, \widehat{\theta}_n)^\top,$$

$$G_{n,t}(\beta, u; \widehat{\theta}_n) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}_t'(X_i^\top \widehat{\theta}_n)\widehat{f}(X_i^\top \widehat{\theta}_n)\widehat{\eta}(X_i, \widehat{\theta}_n)1(X_i^\top \beta \leq u).$$

Consider the following density-weighted generalized residual marked double-projected empirical process

$$S_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{e}_i^{sim}(t; \widehat{\theta}_n, \widehat{\mu}_t)\widehat{f}(X_i^\top \widehat{\theta}_n)\mathcal{P}_{n,t}^{sim}1(X_i^\top \beta \leq u), \quad \left(\beta^\top, u\right)^\top \in \Pi_{pro}.$$

Note that, to compute $\mathcal{P}_{n,t}^{sim}1(X_i^\top \beta \leq u)$, one needs to estimate several infinite-dimensional parameters, including the derivatives $\mu_t'(\cdot)$. This makes the computation of $S_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ substantially more challenging than all the other double-projection empirical process previously considered in this paper.

Following similar arguments as proving (5.13), it is possible to show that, under some

additional regularity conditions, uniformly in $(\beta^\top, u)^\top$, and under H_0^{sim},

$$S_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{sim}(t; \theta_0, \mu_t)) f(X_i^\top \theta_0) \left(\mathcal{P}_t^{sim} 1(X_i^\top \beta \le u) - \mathbb{E}[\mathcal{P}_t^{sim} 1(X^\top \beta \le u) | X_i^\top \theta_0] \right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{sim}(t; \theta_0, \mu_t)) f(X_i^\top \theta_0) \left(\mathcal{P}_t^{sim} 1(X_i^\top \beta \le u) - \mathbb{E}[1(X^\top \beta \le u) | X_i^\top \theta_0] \right) + o_p(1),$$

where the second step follows because $\mathbb{E}[\eta(X_i, \theta_0)|X_i^\top \theta_0] = 0$ $a.s.$. The good news now is that we do not have to deal with the parametric estimation effect due to using $\widehat{\theta}_n$, and therefore can avoid requiring the more stringent condition that $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ admits an asymptotically linear representation. On the other hand, the *nonparametric*-type estimation effect given in (5.15) is still present.

The fact that we do not eliminate the nonparametric-type estimation effect using our double projections leads to some complications that were not present before. For instance, note that we can compute the following CvM test statistic in closed-form:

$$CvM_n^{sim,dpro} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(S_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t) \right)^2 F_{n,\beta}(du) \, d\beta$$

$$= \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} \widehat{e}_i^{sim,pro}(t; \widehat{\theta}_n) \widehat{e}_j^{sim,pro}(t; \widehat{\theta}_n) A_{ijr},$$

where

$$\widehat{e}_i^{sim,pro}(t; \widehat{\theta}_n) = \widehat{e}_i^{sim}(t; \widehat{\theta}_n) \widehat{f}(X_i^\top \widehat{\theta}_n) - \widehat{\mu}_t'(X_i^\top \widehat{\theta}_n) \widehat{f}(X_i^\top \widehat{\theta}_n) \widehat{\eta}(X_i, \widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) \widehat{Q}_{n,t}$$

and

$$\widehat{Q}_{n,t} = \frac{1}{n} \sum_{s=1}^{n} \widehat{\mu}_t'(X_s^\top \widehat{\theta}_n) \widehat{f}(X_s^\top \widehat{\theta}_n)^2 \widehat{\eta}(X_s, \widehat{\theta}_n) \widehat{e}_s^{sim}(t; \widehat{\theta}_n).$$

However, because the underlying null distribution of $S_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ is not invariant to the nonparametric estimator for the link function $\widehat{\mu}_t$, we need to account for its estimation effect in the implementation of a multiplier-bootstrap procedure. That is, a CvM-type multiplier bootstrapped test statistic can be constructed as

$$CvM_n^{sim,dpro,*} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(S_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t) \right)^2 F_{n,\beta}(du) \, d\beta,$$

where

$$S_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \widehat{e}_i^{sim}(t; \widehat{\theta}_n) \widehat{f}(X_i^\top \widehat{\theta}_n) \left(\mathcal{P}_{n,t}^{sim} 1(X_i^\top \beta \le u) - \widehat{\phi}(\beta, u; X_i^\top \widehat{\theta}_n) \right),$$

with $\{V_i\}_{i=1}^{n}$ a sequence of multipliers defined in Section 4 and $\widehat{\phi}(\beta, u; X_i^\top \widehat{\theta}_n)$ the leave-

one-out kernel estimator for $\phi(\beta, u; X_i^\top \theta_0) \equiv \mathbb{E}[1(X^\top \beta \leq u)|X^\top \theta_0 = v]$ given by

$$\widehat{\phi}(\beta, u; X_i^\top \widehat{\theta}_n) = \frac{\frac{1}{(n-1)h} \sum_{j \neq i}^n K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h}\right) 1(X_j^\top \beta \leq u)}{\widehat{f}(X_i^\top \widehat{\theta}_n)}.$$

Unfortunately, given that $S_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n, \widehat{\mu}_t)$ depends on $\phi(\beta, u; X_i^\top \theta_0)$, which is indexed by β and u, it is not clear if one can get an easy-to-use closed-form expression for $CvM_n^{sim,dpro,*}$. This suggests that one would need to resort to numerical integration procedures to compute critical values for $CvM_n^{sim,dpro}$ based on $CvM_n^{sim,dpro,*}$. Although this is feasible, it also seems much more computationally challenging. Given that these topics are sufficiently different from the rest of the paper, we leave their theoretical justifications and detailed implementation guidelines for future research.

6   Data illustration

In this section, we apply our tests to analyze the goodness-of-fit of different GPS models used to study the effect of maternal smoking on birth weight. The dataset, available at http://www.stata-press.com/data/r13/cattaneo2.dta, is the excerpt from Almond et al. (2005) and Cattaneo (2010) previously used by Lee et al. (2017). It consists of observations from white mothers in Pennsylvania in the USA; like Lee et al. (2017), we further restrict our sample to white and non-Hispanic mothers, total $3,754$ observations. The treatment variable, T, is a multi-valued variable that is equal to 0 if the mother does not smoke during the pregnancy, equal to 1 if the mother smokes, on average, between one and five cigarettes a day during the pregnancy, equal to 2 if the mother smokes, on average, between six and ten cigarettes a day during the pregnancy, and equal to 3 if the mother smokes, on average, more than eleven cigarettes a day during the pregnancy. The set of pre-treatment covariates X we use are the mother's age, number of prenatal care visits, and indicator variables for alcohol consumption during pregnancy, first prenatal visit in the first trimester, whether there was a previous birth where the newborn died, twelve years of education (complete high-school), and more than twelve years of educations (some college). The outcome of interest is the infant's birth weight measured in grams.

We start our analysis by analyzing the effect of the mother being a smoker during the pregnancy $(T > 0)$ versus not smoking during the pregnancy $(T = 0)$. Given the binary nature of the "being a smoker" treatment, we estimate the propensity score using a logistic regression model with linear predictors including all aforementioned covariates. We then apply our proposed specification test to assess the goodness-of-fit of this simple propensity score model, using $9,999$ bootstrap replications. Our procedure yields a p-value of 0.18, suggesting that our proposed testing procedure does not find any evidence

of model misspecification at the usual significance levels.

Next, we move our attention to analyzing the effect of maternal smoking intensity during pregnancy, T. Given that the treatment T is clearly ordered, we estimate the GPS using an ordered logit regression model with all covariates entering the model in a linear fashion. Although natural, we note that the ordered logit model imposes important restrictions on the data such as a proportional odds restriction. In practice, however, such restrictions may be too rigid for a given application. Indeed, our proposed specification test with $9,999$ bootstrap replications yield a p-value of 0.08, suggesting that the ordered logit model is misspecified at the 10% significance level.

A relatively straightforward way to relax the proportional odd restrictions inherited in the ordered logit model is to ignore that the treatment T is ordered and estimate the GPS using a multinomial logit linear regression model. In contrast with the ordered logit model, the multinomial logit model does not impose that the regressors' coefficients are the same across different treatment levels. Our specification test with $9,999$ bootstrap replications yield a p-value of 0.72, suggesting that the multinomial logit model is a more suitable model for maternal smoking intensity during pregnancy T than the order logit model.

Table 1 shows the estimates of the causal effects of smoking on birth weight based on the inverse probability weighting estimators for binary and multi-valued treatments, ATE_n as in (B.3) and $ATE_{n,j,\ell}$ as in (B.7), respectively. Although qualitatively similar, one can notice some differences in the standard errors and confidence intervals between the treatment effects estimates based on misspecified ordered logit model and those based on the multinomial logit model for the GPS. This illustrates one of the potential pitfalls of model misspecification.

7   Conclusions and directions for further research

In this article, we proposed a new class of specification tests for GPS models based on novel double-projected weight functions. We have shown that using double projections helps ameliorate the "curse of dimensionality" and avoids the complications associated with "parameter estimation uncertainty" commonly encountered in specification testing. We have shown that our proposed test statistics can be written in closed form, and that one can use an easy-to-implement multiplier bootstrap procedure to compute critical values as accurately as desired. We have also extended our double-projection proposal to test parametric multiple-index or semiparametric single-index GPS. The simulation results (in Appendix B) and the empirical application highlight that our proposed tests can serve as a valuable diagnostic tool in the context of multi-valued treatment effects.

We anticipate that one can extend our proposal to test whether putative parametric conditional distributions, distributional regressions, or linear/nonlinear quantile re-

Table 1: Results from the empirical illustration: point estimates, standard error by the bootstrap, and 95% confidence interval

	Point Estimate	Standard Error	95% Confidence Interval
(a) The causal effect of mother being a smoker during pregnancy on infant's birth weight Estimators based on a logit propensity score model			
$\mathbb{E}\left[Y\left(1\right)-Y\left(0\right)\right]$	-270	28	(-326, -215)
(b) The causal effect of mother's smoking intensity during pregnancy on infant's birth weight Estimators based on an ordered logit GPS model			
$\mathbb{E}\left[Y\left(1\right)-Y\left(0\right)\right]$	-279	59	(-401, -164)
$\mathbb{E}\left[Y\left(2\right)-Y\left(0\right)\right]$	-244	46	(-338, -156)
$\mathbb{E}\left[Y\left(3\right)-Y\left(0\right)\right]$	-279	39	(-357, -202)
(c) The causal effect of mother's smoking intensity during pregnancy on infant's birth weight Estimators based on a multinomial logit GPS model			
$\mathbb{E}\left[Y\left(1\right)-Y\left(0\right)\right]$	-250	49	(-346, -154)
$\mathbb{E}\left[Y\left(2\right)-Y\left(0\right)\right]$	-241	49	(-338, -148)
$\mathbb{E}\left[Y\left(3\right)-Y\left(0\right)\right]$	-267	42	(-347, -184)

Note: Standard errors are computed using the empirical bootstrap with $9,999$ draws. 95% confidence intervals based on the percentile bootstrap with $9,999$ draws. See the main text for further details.

gressions are correctly specified; see, e.g., Bierens and Wang (2012), Rothe and Wied (2013) and Escanciano and Goh (2014). Due to the lack of a closed-form expression of $CvM_n^{sim,dpro,*}$ at the end of Section 5.3, another direction of future research could focus on developing alternative bootstrap methods to obtain critical values of tests that are designed to test the semiparametric single-index assumption against a general nonparametric alternative and are robust to both the dimensionality of covariates and the estimation of parametric/nonparametric-type nuisance parameters. We leave a detailed analysis of these interesting extensions for future research.

References

Aït-Sahalia, Y., Bickel, P. J., and Stoker, T. M. (2001), "Goodness-of-fit tests for kernel regression with an application to option implied volatilities," *Journal of Econometrics*, 105(2), 363–412.

Almond, D., Chay, K., and Lee, D. (2005), "The Costs of Low Birth Weight," *The Quarterly Journal of Economics*, 120(3), 1031–1083.

Ao, W., Calonico, S., and Lee, Y.-Y. (2019), "Multivalued Treatments and Decomposition Analysis: An Application to the WIA Program," *Journal of Business & Economic Statistics*, 0(0), 1–14.

Bickel, P. J., Ritov, Y., and Stoker, T. M. (2006), "Tailor-made tests for goodness of fit to semiparametric hypotheses," *Annals of Statistics*, 34(2), 721–741.

Bierens, H. J. (1982), "Consistent model specification tests," *Journal of Econometrics*, 20(1982), 105–134.

Bierens, H. J., and Wang, L. (2012), "Integrated conditional moment tests for parametric conditional distributions," *Econometric Theory*, 28(2), 328–362.

Bühlmann, P., and van de Geer, S. (2011), *Statistics for High-Dimensional Data*, Springer Series in Statistics, Berlin, Heidelberg: Springer Berlin Heidelberg.

Busso, M., Dinardo, J., and McCrary, J. (2014), "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators," *The Review of Economics and Statistics*, 96(5), 885–895.

Cattaneo, M. D. (2010), "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.

Delgado, M. A., Domínguez, M. A., and Lavergne, P. (2006), "Consistent tests of conditional moment restrictions," *Annales d'Économie et de Statistique*, pp. 33–67.

Delgado, M. A., and González Manteiga, W. (2001), "Significance testing in nonparametric regression based on the bootstrap," *The Annals of Statistics*, 29(5), 1469–1507.

Dominguez, M. A., and Lobato, I. N. (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72(5), 1601–1615.

Domínguez, M. A., and Lobato, I. N. (2020), "Specification testing with estimated variables," *Econometric Reviews*, 39(5), 476–494.

Escanciano, J. C. (2006), "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Escanciano, J. C. (2008), "Joint and marginal specification tests for conditional mean and variance models," *Journal of Econometrics*, 143(1), 74–87.

Escanciano, J. C., and Goh, S. C. (2014), "Specification analysis of linear quantile models," *Journal of Econometrics*, 178, 495–507.

Escanciano, J. C., and Jacho-Chávez, D. T. (2010), "Approximating the critical values of Cramér-von Mises tests in general parametric conditional specifications," *Computational statistics & data analysis*, 54(3), 625–636.

Escanciano, J. C., and Song, K. (2010), "Testing single-index restrictions with a focus on average derivatives," *Journal of Econometrics*, 156(2), 377–391.

Fan, Y., and Li, Q. (1996), "Consistent model specification tests: omitted variables and semiparametric functional forms," *Econometrica*, 64(4), 865–890.

Fan, Y., and Liu, Z. (1997), "A Simple Test for a Parametric Single Index Model," *Opsearch*, 34(3), 186–195.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014), "A Goodness-of-Fit Test for the Functional Linear Model with Scalar Response," *Journal of Computational and Graphical Statistics*, 23(3), 761–778.

González-Manteiga, W., and Crujeiras, R. M. (2013), "An updated review of Goodness-of-Fit tests for regression models," *Test*, 22(3), 361–411.

Guo, X., Wang, T., and Zhu, L. (2016), "Model checking for parametric single-index models: a dimension reduction model-adaptive approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5), 1013–1035.

Guo, X., and Zhu, L. (2017), "A Review on Dimension-Reduction Based Tests For Regressions," in *From Statistics to Mathematical Finance*, Cham: Springer International Publishing, pp. 105–125.

Härdle, W., Mammen, E., and Proença, I. (2001), "A bootstrap test for single index models," *Statistics: a journal of theoretical and applied statistics*, 35(4), 427–451.

Horowitz, J. L., and Härdle, W. (1994), "Testing a parametric model against a semiparametric alternative," *Econometric theory*, 10(5), 821–848.

Imbens, G. (2000), "The role of the propensity score in estimating dose-response functions," *Biometrika*, 87(3), 706–710.

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference in Statistics, Social and Biometical Sciences*, Cambridge, MA: Cambridge University Press.

Iwasawa, M. (2015), "A joint specification test for response probabilities in unordered multinomial choice models," *Econometrics*, 3(3), 667–697.

Kim, I., Balakrishnan, S., and Wasserman, L. (2020), "Robust Multivariate Nonparametric Tests via Projection-Averaging," *Annals of Statistics*, To Appear, 1–34.

Knight, K., and Fu, W. (2000), "Asymptotics for Lasso-type estimators," *Annals of Statistics*, 28(5), 1356–1378.

Lee, S., Okui, R., and Whang, Y.-J. J. (2017), "Doubly robust uniform confidence band for the conditional average treatment effect function," *Journal of Applied Econometrics*, 32(7), 1207–1225.

Lee, Y. Y. (2018), "Efficient propensity score regression estimators of multivalued treatment effects for the treated," *Journal of Econometrics*, 204(2), 207–222.

Linden, A., Uysal, S. D., Ryan, A., and Adams, J. L. (2016), "Estimating causal effects for multivalued treatments: A comparison of approaches," *Statistics in Medicine*, 35(4), 534–552.

Liu, R. Y. (1988), "Bootstrap Procedures under some Non-I.I.D. Models," *Annals of Statistics*, 16(4), 1696–1708.

Lopez, M. J., and Gutman, R. (2017), "Estimation of causal effects with multiple treatments: A review and new ideas," *Statistical Science*, 32(3), 432–454.

Maistre, S., and Patilea, V. (2019), "Nonparametric model checks of single-index assumptions," *Statistica Sinica*, 29(1), 113–138.

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models," *The Annals of Statistics*, 21(1), 255–285.

Mora, J., and Moro-Egido, A. I. (2008), "On specification testing of ordered discrete choice models," *Journal of Econometrics*, 143(1), 191–205.

Neyman, J. (1959), "Optimal Asymptotic Tests of Composite Statistical Hypotheses," in *Probability and Statistics: The Harald Cramér Volume*, ed. U. Grenander, Stockholm: Almqvist & Wiksell, pp. 213–234.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, pp. 1403–1430.

Rothe, C., and Wied, D. (2013), "Misspecification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, 108(501), 314–324.

Sant'Anna, P. H., and Song, X. (2019), "Specification tests for the propensity score," *Journal of Econometrics*, 210(2), 379–404.

Shaikh, A. M., Simonsen, M., Vytlacil, E. J., and Yildiz, N. (2009), "A specification test for the propensity score using its distribution conditional on participation," *Journal of Econometrics*, 151(1), 33–46.

Stute, W. (1997), "Nonparametric model checks for regression," *The Annals of Statistics*, 25(2), 613–641.

Stute, W., and Zhu, L.-X. (2002), "Model Checks for Generalized Linear Models," *Scandinavian Journal of Statistics*, 29(3), 535–545.

Stute, W., and Zhu, L.-X. (2005), "Nonparametric checks for single-index models," *The Annals of Statistics*, 33(3), 1048–1083.

Sun, Z., Chen, F., Zhou, X., and Zhang, Q. (2017), "Improved model checking methods for parametric models with responses missing at random," *Journal of Multivariate Analysis*, 154, 147–161.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Whang, Y.-J. (2001), "Consistent specification testing for conditional moment restrictions," *Economics Letters*, 71(3), 299–306.

Xia, Y. (2006), "Asymptotic Distributions for Two Estimators of the Single-Index Model," *Econometric Theory*, 22(06).

Xia, Y. (2009), "Model checking in regression via dimension reduction," *Biometrika*, 96(1), 133–148.

Xia, Y., Li, W. K., Tong, H., and Zhang, D. (2004), "A goodness-of-fit test for single-index models," *Statistica Sinica*, pp. 1–28.

Xia, Y., Tong, H., Li, W. K., and Zhu, L.-X. (2002), "An adaptive estimation of dimension reduction space," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3), 363–410.

Zhu, L., Xu, K., Li, R., and Zhong, W. (2017), "Projection correlation between two random vectors," *Biometrika*, 104(4), 829–843.