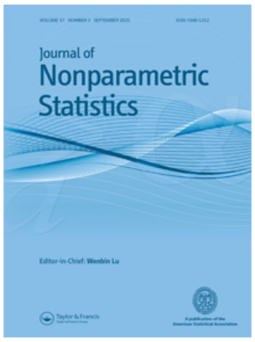

Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: www.tandfonline.com/journals/gnst20

Specification tests for generalised propensity scores using double projections

Juan Carlos Escanciano, Pedro H. C. Sant'Anna & Xiaojun Song

JOURNAL OF NONPARAMETRIC STATISTICS
https://doi.org/10.1080/10485252.2025.2551752

Specification tests for generalised propensity scores using double projections

Juan Carlos Escanciano [a], Pedro H. C. Sant'Anna[b] and Xiaojun Song[c]

[a]Department of Economics, Universidad Carlos III de Madrid, Getafe (Madrid), Spain; [b]Department of Economics, Emory University, Atlanta, Georgia, USA; [c]Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China

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 generalised propensity score models. The test procedure is based on two different projection arguments, leading to test statistics that are suitable for setups with many covariates and are (asymptotically) invariant to the estimation method used to estimate the nuisance parameters. We show that our proposed test is able to detect a broad class of local alternatives converging to the null at the usual parametric rate, and illustrate its attractive power properties through simulations.

ARTICLE HISTORY
Received 27 September 2024
Accepted 17 August 2025

KEYWORDS
Double projections; generalised 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 exposure, policy, or treatment on outcomes of interest. When the assignment to treatment is not randomised, 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 generalisations 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 generalisations (henceforth GPS) are popular, a primary concern with these methods is that the GPS is usually unknown and therefore has to be estimated. Given the availability of many pretreatment covariates and limited sample size, researchers usually adopt parametric models for the GPS. Such a common practice raises the important issue of model misspecification. Thus, assessing if your parametric putative model for the GPS is correctly specified is recommended.

CONTACT Xiaojun Song sxj@gsm.pku.edu.cn Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing 100871, China.

Supplemental data for this article can be accessed online at http://doi.org/10.1080/10485252.2025.2551752.

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, see Remark 2.1 for other applications. 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 2006a). 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 2009a; Escanciano and Goh 2014; Sant'Anna and Song 2019). As a result of this second projection, we can consider a broader 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, which supersedes and expands on Escanciano (2009b), is the first to combine the two aforementioned projections. Escanciano (2006a)'s test, although robust to dimensionality, is not robust to the estimation of nuisance parameters. A potential drawback of the test of Escanciano (2006a) is that its implementation relies on the wild bootstrap, and it is not directly applicable when the response variables are discrete or mixed, see Escanciano and Goh (2014), 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 invariant 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, prior to this paper, no other (specification) test available in the literature enjoys all these attractive properties, see González-Manteiga and Crujeiras (2013) for a review of existing methodologies. The results of Monte Carlo simulations indicate that these attractive properties translate to tests with excellent finite-sample properties, even when the dimension of the 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 whether the rejection 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).

The rest of this article is organised 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, and Section 5.2 tests a parametric index model against a semiparametric alternative. We evaluate the finite sample performance of the proposed tests through Monte Carlo simulations in Section 6. Section 7 concludes. Section 8 gathers the proofs of the main results. An online supplementary appendix contains further Monte Carlo simulation results and an empirical application.

2. Specification tests based on double projections

2.1. Setup and motivation

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

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

against

$$H_1 : \mathbb{P}\left(\mathbb{E}\left[e(t;\theta)\,|\,X\right] = 0\right) < 1 \quad \text{for any } \theta \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and some } t \in \mathcal{T}, \tag{2}$$

where $e(t;\theta)$ is a generalised 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 and the estimation of nuisance parameter θ_0. We note that in this paper, "many covariates" refers to the case where the dimension d_x of the covariates X is allowed to be large but finite. For the derivation of our asymptotic results, we restrict d_x to be a fixed value. Results with $d_x \to \infty$, as the sample size increases, are beyond the scope of this paper.

Although our framework is relatively general, it is worthwhile to motivate it within a popular and empirically relevant causal inference setup. For a generic matrix A, A' denotes the transpose of A. Let $J \in \mathbb{N}$ be a given positive integer. Let $(X', T, Y)'$ 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 $d_x \in \mathbb{N}$, $T \in \mathcal{T} = \{0, 1, \ldots, J\}$ is the treatment random variable, $Y = \sum_{t=0}^{J} 1(T = t)Y(t) \in \mathcal{Y} \subseteq \mathbb{R}$ is the observed outcome, and $Y(t)$ denotes the potential outcome when T is externally set to t. To motivate an important application of the GPS, let us consider the average causal effects of the form

$$\beta(t, s) \equiv \mathbb{E}\left[Y(t) - Y(s)\right]. \tag{3}$$

This is the average causal effect of exposing units to treatment t rather than treatment s. If assignment to treatment T is weakly unconfounded given the pre-treatment variables X, in the sense that $Y(t)$ is conditionally independent of the indicator $1(T=t)$ given X, for all $t \in \mathcal{T}$, then $\beta(t,s)$ in (3) is identified by the following weighting estimands:

$$\beta(t,s) = \mathbb{E}\left[\frac{Y1\left(T=t\right)}{p_t\left(X\right)}\right] - \mathbb{E}\left[\frac{Y1\left(T=s\right)}{p_s\left(X\right)}\right], \tag{4}$$

where $p_t(x) \equiv \mathbb{P}(T=t|X=x) = \mathbb{E}[1(T=t)\,|\,X=x]$ is the unknown GPS, assumed to be bounded away from zero for all t. To estimate $\beta(t,s)$ using (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, $\sum_{t=0}^{J} q_t(x,\theta_t) = 1$. 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 (e.g. Imbens 2000),

$$q_t\left(X,\theta_t\right) = \frac{\exp\left(X'\theta_t\right)}{1+\sum_{j=1}^{J}\exp\left(X'\theta_j\right)},\ t \in \{1,\ldots,J\}. \tag{5}$$

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 misspecification. 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 generalised 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, in the rest of the paper, we take $e(t;\theta)$ as the generalised error of the propensity score model, as mentioned above, although we can handle other models; see the following remark.

Remark 2.1: Our methodology applies to any other conditional moment restriction problem where $e(t;\theta)$ is a measurable function of the data Z and $g_t(X,\theta) = \partial\mathbb{E}[e(t;\theta)\,|\,X]/\partial\theta$ is known up to the finite-dimensional parameter θ. Example applications are given in, e.g. Whang (2001), Delgado et al. (2006), and Escanciano and Jacho-Chávez (2010). This setting includes, among many others, tests for parametric conditional mean models, where $e(1;\theta) = Y - q(X,\theta)$ for some parametric function $q(X,\theta)$, a scalar dependent variable Y, and $\mathcal{T} = \{1\}$, or joint tests for conditional mean and variance models in the spirit of Escanciano (2008), where $e(1;\theta) = Y - q_1(X,\theta_1)$, $e(2;\theta) = (Y - q_1(X,\theta_1))^2 - q_2(X,\theta_2)$, 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 $Y - \mathbb{E}[Y\,|\,X]$. Extensions of our double projection tests to nonparametric or semiparametric scores $g(X,\theta_0)$ are also possible, but they are beyond the scope of this paper.

2.2. Our proposed test

The characterisation of H_0 and H_1 in (1) and (2), respectively, makes explicit that we are interested in testing conditional moment restrictions (González-Manteiga and Crujeiras 2013). As argued by Escanciano (2006a), (1) can be equivalently characterised as

$$H_0 : R_t^{pro}(\beta, u; \theta_0) = 0 \text{ almost everywhere (a.e.) } \left(\beta', u\right)' \in \Pi_{pro},$$
$$\text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \tag{6}$$

where

$$R_t^{pro}(\beta, u; \theta) \equiv \mathbb{E}\left[e(t;\theta)1\left(X'\beta \leq u\right)\right],$$

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

Although one can find alternative characterisations of H_0, see, e.g. Bierens (1982), Stute (1997), and Dominguez and Lobato (2004), our main motivations for expressing H_0 as in (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 $X'\beta$, greatly reducing the dimensionality of the problem. Indeed, this dimension-reduction device has been proven valuable in various contexts that require dealing with many covariates; see, e.g. Escanciano (2006a), García-Portugués et al. (2014), Sun et al. (2017), Zhu et al. (2017), and Kim et al. (2020). However, it is worth mentioning that (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 (6), one natural way to proceed is to compute the generalised residual marked empirical process based on the projections $1(X'\beta \leq u)$,

$$R_{n,t}^{pro}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t; \widehat{\theta}_n) 1(X_i'\beta \leq u), \quad \left(\beta', u\right)' \in \Pi_{pro},$$

where $\widehat{\theta}_n$ is any $\sqrt{n}$-consistent estimator for θ_0, say the maximum likelihood estimator, and $e_i(t; \widehat{\theta}_n) \equiv 1(T_i = t) - q_t(X_i, \widehat{\theta}_n), i = 1, \ldots, n$, are the parametrically specified generalised 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 estimator $\widehat{\theta}_n$. 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, tests based on $R^{pro}_{n,t}(\beta,u;\widehat{\theta}_n)$ also require that $\sqrt{n}(\widehat{\theta}_n-\theta_0)$ admits an asymptotically linear representation. 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 the function $l(\cdot,\cdot,\cdot)$ is such that $\mathbb{E}[l(T,X,\theta_0)]=0$ and $\mathbb{E}[l(T,X,\theta_0)l(T,X,\theta_0)']$ exists and is positive definite. Note that different $\widehat{\theta}_n$ may lead to different $l(\cdot,\cdot,\cdot)$. Such a condition can be demanding, especially when one wishes to use estimation methods that involve penalisation (see, e.g. Knight and Fu 2000; Bühlmann and van de Geer 2011). The fact that $R^{pro}_{n,t}(\beta,u;\widehat{\theta}_n)$ 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, see, e.g. Escanciano and Goh (2014).

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

$$R^{dpro}_{n,t}(\beta,u;\widehat{\theta}_n)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(t;\widehat{\theta}_n)\mathcal{P}_{n,t}1(X_i'\beta\le u),\quad (\beta',u)'\in\Pi_{pro},$$

where the double-projected weight is

$$\mathcal{P}_{n,t}1\left(X_i'\beta\le u\right)\equiv 1\left(X_i'\beta\le u\right)-g_t(X_i,\widehat{\theta}_n)'\Delta^{-1}_{n,t}(\widehat{\theta}_n)G_{n,t}(\beta,u;\widehat{\theta}_n),\tag{7}$$

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

$$\Delta_{n,t}(\widehat{\theta}_n)=\frac{1}{n}\sum_{i=1}^{n}g_t(X_i,\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n)'\quad\text{and}$$

$$G_{n,t}(\beta,u;\widehat{\theta}_n)=\frac{1}{n}\sum_{i=1}^{n}g_t(X_i,\widehat{\theta}_n)1\left(X_i'\beta\le u\right).$$

We label $\mathcal{P}_{n,t}1(X'\beta\le u)$ as a double-projection because, as is evident from (7), it involves first using the projection proposed by Escanciano (2006a), $1(X'\beta\le u)$, and then projecting $1(X'\beta\le u)$ onto the tangent space of the nuisance parameters (see, e.g. Neyman 1959; Escanciano 2009b; Escanciano and Goh 2014; Sant'Anna and Song 2019). To the best of our knowledge, this paper is the first to incorporate this double-projection 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 $t\in\mathcal{T}$,

$$\mathbb{E}\left[g_t(X,\theta_0)\mathcal{P}_t1\left(X'\beta\le u\right)\right]\equiv 0,$$

almost everywhere in $(\beta', u)' \in \Pi_{pro}$, where

$$\mathcal{P}_t 1\left(X'\beta \le u\right) \equiv 1\left(X'\beta \le u\right) - g_t(X,\theta_0)'\Delta_t^{-1}(\theta_0)G_t(\beta,u;\theta_0), \tag{8}$$

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

The intuition behind (8) is very simple. First of all, note that, for each $t \in \mathcal{T}$, $(\beta', u)' \in \Pi_{pro}$, $\Delta_t^{-1}(\theta_0)G_t(\beta,u;\theta_0)$ is the vector of linear projection coefficients of regressing $1(X'\beta \le u)$ on the score function $g_t(X,\theta_0)$. Thus, it follows that

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

is the best linear predictor of $1(X'\beta \le u)$ given $g_t(X,\theta_0)$, and that (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 $(\beta', u)' \in \Pi_{pro}$,

$$R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\mathcal{P}_t 1(X_i'\beta \le u) + o_p(1), \tag{9}$$

for each $t \in \mathcal{T}$; see the proof of Theorem 3.1 in the next section. Moreover, $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$ 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 $R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)$, we can treat the residuals $e(t;\widehat{\theta}_n)$ as if they were the true errors $e(t;\theta_0)$; 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 typically have different asymptotically linear representations, resulting in different multiplier bootstraps and potentially different testing results. Of course, (9) allows us to avoid these problems.

In order to operationalise 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

$$CvM_n^{dpro} = \sum_{t\in\mathcal{T}} a_n(t)\int_{\Pi_{pro}}\left(R_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\,\mathrm{d}\beta, \tag{10}$$

where, for each t, $a_n(t)$ is a pre-specified (potentially random) non-negative weighting function, $F_{n,\beta}(u) = n^{-1}\sum_{i=1}^{n} 1(X_i'\beta \le u)$ is the empirical distribution function of the one-dimensional projected regressors $\{X_i'\beta\}_{i=1}^{n}$ 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 favour of the alternative H_1 whenever CvM_n^{dpro} in (10) is 'overly' large. For the sake of practical convenience, in the Monte Carlo simulations of Section 6 and the additional simulations and the empirical application 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. $a_n(t) = n^{-1} \sum_{i=1}^{n} 1(T_i = t)$. The investigation of the optimal choice of a for a given alternative is left for future research.

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, the CvM test statistic in (10) can be written in a closed-form expression and does not rely on any numerical integration method (cf. Escanciano 2006b). A direct consequence of these attractive computational features is that (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* (10) *with $\mathbb{S}^{d_x}$ the d_x-dimensional unit sphere. Then, we have*

$$CvM_n^{dpro} = \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=1}^{n} e_i^{pro}(t;\widehat{\theta}_n) e_j^{pro}(t;\widehat{\theta}_n) A_{ijr}, \tag{11}$$

with

$$e_i^{pro}(t;\widehat{\theta}_n) = e_i(t;\widehat{\theta}_n) - g_t(X_i,\widehat{\theta}_n)' \Delta_{n,t}^{-1}(\widehat{\theta}_n) \frac{1}{n} \sum_{j=1}^{n} g_t(X_j,\widehat{\theta}_n) e_j(t;\widehat{\theta}_n), \tag{12}$$

and

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

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

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

It is interesting to note that $e_i^{pro}(t;\widehat{\theta}_n)$ for $i = 1, \ldots, n$ in (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}. In fact, there is a computationally convenient matrix form expression of the test statistic, which is derived following the steps:

(1) Define the $n \times 1$ vector of residuals $\hat{\mathbf{e}}_t = (e_1(t;\widehat{\theta}_n), \ldots, e_n(t;\widehat{\theta}_n))'$.
(2) Compute the $n \times d_\theta$ matrix of scores $\mathbb{G}_t$ with ith row $g_t(X_i,\widehat{\theta}_n)'$, $i = 1, \ldots, n$.
(3) Let $\mathbb{A}$ denote the $n \times n$ matrix with ijth element $a_{ij} = \sum_{r=1}^{n} A_{ijr}/n$, $i,j = 1, \ldots, n$.
(4) Compute $\hat{\mathbb{W}}_t = \hat{\Pi}_t' \mathbb{A} \hat{\Pi}_t$, where $\hat{\Pi}_t = I - \mathbb{G}_t(\mathbb{G}_t'\mathbb{G}_t)^{-1}\mathbb{G}_t'$ and I is the identity matrix.

(5) Compute the test statistic $CvM_n^{dpro} = \sum_{t\in\mathcal{T}} a_n(t)\frac{1}{n}\hat{\mathbf{e}}_t'\hat{\mathbb{W}}_t\hat{\mathbf{e}}_t$.

The goffda R package (see García-Portugués et al. 2021), and more specifically its Adot function, computes the expression of $\mathbb{A}$ efficiently (in C++). Note that $\mathbb{A}$ needs to be computed only once to run the test.

3. Asymptotic results

3.1. Asymptotic null distribution

In this section, we formally investigate the limiting behaviour of double-projected generalised residual marked empirical process $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$ under the null hypothesis H_0 in (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 $\Psi_{pro}(du, d\beta) = F_\beta(du)d\beta$, where $F_\beta(\cdot)$ is the CDF of $X_i'\beta$. Recall that $p_t(x) = \mathbb{P}(T = t|X = x)$ for every $t \in \mathcal{T}$ are the true but unknown GPS. We list all the relevant regularity conditions as follows.

Assumption 3.1: The random sample $\{(X_i', T_i)', i = 1\ldots n\}$ consists of a sequence of independent and identically distributed random vectors from $(X', T)'$.

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 neighbourhood 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 $\sup_{\theta\in\Theta_0} ||g_t(\cdot,\theta)|| \le M(\cdot)$.

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 $||\widehat{\theta}_n - \theta_0|| = O_p(n^{-1/2})$ under H_0, and $||\widehat{\theta}_n - \theta^*|| = o_p(1)$ 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 (2006a). 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 most papers in the literature of specification testing based on empirical processes with estimated parameters.

To present our asymptotic results, 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). Then, it is shown that under the null

$R_{n,t}^{dpro}(\cdot,\cdot;\widehat{\theta}_n) \Longrightarrow R_{\infty,t}^{dpro}$, a centred Gaussian process with its covariance structure defined in Theorem 3.1.

The true generalised error is defined as $\varepsilon(t) = 1(T = t) - p_t(X)$, which satisfies $\mathbb{E}[\varepsilon(t) \mid X] = 0$ almost surely (*a.s.*) for each $t \in \mathcal{T}$ regardless of whether the null hypothesis is true. We state formally the asymptotic null behaviour 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* (1), *for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t\in\mathcal{T}} a(t) < \infty$, 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 $R_{\infty,t}^{dpro}$ and $R_{\infty,s}^{dpro}$ are Gaussian processes with mean zero and covariance structure

$$\mathbb{K}_{ts}^{dpro}((\beta_1, u_1), (\beta_2, u_2)) = \mathbb{E}\left[\sigma_{ts}^2(X)\mathcal{P}_t 1\left(X'\beta_1 \le u_1\right)\mathcal{P}_s 1\left(X'\beta_2 \le u_2\right)\right],$$

where $\sigma_{ts}^2(X) = \mathbb{E}[\varepsilon(t)\varepsilon(s)|X] = -p_t(X)p_s(X)1(t \ne s) + p_t(X)(1 - p_t(X))1(t = s)$ is the conditional covariance function of generalised errors $\varepsilon(t)$ and $\varepsilon(s)$ given X.

To prove Theorem 3.1 in Section 8, we first show that the asymptotic null behaviour 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).

Theorem 3.2: *Suppose Assumptions* 3.1– 3.4 *hold. Then, under the fixed alternative hypothesis H_1 in* (2), *for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t\in\mathcal{T}} a(t) < \infty$, we have that*

$$\frac{CvM_n^{dpro}}{n} \xrightarrow{p} \sum_{t\in\mathcal{T}} a(t) \int_{\Pi_{pro}} \left(\mathbb{E}\left[\left(p_t(X) - q_t\left(X, \theta^*\right)\right)\mathcal{P}_t 1\left(X'\beta \le u\right)\right]\right)^2 \Psi_{pro}(du, d\beta).$$

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

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

for some $(\beta', u)'$ 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) if $\mathbb{E}[(p_t(X)-q_t(X,\theta^*))\mathcal{P}_t 1(X'\beta\leq u)]=0$ for every $t \in \mathcal{T}$ and all $(\beta', u)' \in \Pi_{pro}$. 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 $q_t(X,\theta^*)$. Lack of power against these alternatives is not an important limitation. After all, all empirical-process-based tests have trivial local power against those directions, see Escanciano (2009a). As a result, the global power of all tests in the direction of the score will be also low (cf. Strasser 1990).

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

$$H_{1,n}:\mathbb{P}\left[p_t(X)=q_t(X,\theta_0)+\frac{r_t(X)}{\sqrt{n}}\right]=1\quad\text{for some }\theta_0\in\Theta\in\mathbb{R}^{d_\theta}\ \text{ and all } t\in\mathcal{T},\tag{13}$$

where, for each $t \in \mathcal{T}$, the direction of departure from H_0 is given by the function $r_t(X)$ (potentially different for each t), which is assumed to be $F_X(\cdot)$-integrable with zero mean and to 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* (13), *for any sequence $a_n(t)=a(t)+o_p(1)$, with $a(t)>0$ and $0<\sum_{t\in\mathcal{T}}a(t)<\infty$, we have that*

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

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

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

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 $(\beta', u)' \in \Pi_{pro}$ with a positive Lebesgue measure, our proposed Cramér–von Mises test statistic will have non-trivial power against local alternatives of the form (13). 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. $r_t(X) = \nu' g_t(X, \theta_0)$ *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. Escanciano (2009a) has shown that this is not a limitation of our test, but rather an unavoidable limitation of all empirical-processes-based tests. Also, this author has shown that the local power function of omnibus empirical-processes-based tests is essentially flat, except for a handful of alternatives. This may explain the close to zero power of these tests in some simulations. Unfortunately, it is hard to know in advance the deviations (scenarios) for which this happens, as for a given method, this will depend on the unknown data-generating process in a complicated way, see the principal components decomposition in Escanciano (2009a).

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 bootstrapped p-values or critical values, and demonstrate its asymptotic validity. Its implementation is as follows:

(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). Define $\hat{\mathbf{e}}_t^* = (e_1^*(t; \widehat{\theta}_n), \ldots, e_n^*(t; \widehat{\theta}_n))'$.

(2) Compute $CvM_{n,b}^{dpro,*} = \sum_{t \in \mathcal{T}} a_n(t) \frac{1}{n} (\hat{\mathbf{e}}_t^*)' \hat{\mathbb{W}}_t \hat{\mathbf{e}}_t^*$.

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

(4) Obtain the $(1 - \alpha)$ quantile of $\{CvM_{n,b}^{dpro,*} : b = 1 \ldots, B\}$, $c_{n,\alpha}^*$, 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 (1) if CvM_n^{dpro} is greater than the critical value $c_{n,\alpha}^*$, and fail to reject (1) otherwise.

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 cost of the proposed procedure. Second, thanks to the use of the second projection in $\mathcal{P}_{n,t} 1(X_i'\beta \leq u)$, 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 (11), A_{ijr} does not need to be computed for each bootstrap sample. Finally, it does not involve any tuning parameters such as bandwidths. These features significantly reduce 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

$$R_{n,t}^{dpro,*}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i e_i(t; \widehat{\theta}_n) \mathcal{P}_{n,t} 1\left(X_i'\beta \leq u\right)$$

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

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

whose closed-form expression is given in Step 2 above. It is shown in Section 8 that

$$R^{dpro,*}_{n,t}(\beta, u; \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i e_i(t; \theta_0) \mathcal{P}_t 1\left(X_i'\beta \le u\right) + o_p(1),$$

uniformly in $(\beta', u)'$ for each $t \in \mathcal{T}$. Thus, $R^{dpro,*}_{n,t}(\beta, u; \widehat{\theta}_n)$ 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* $CvM^{dpro,*}_n \overset{d}{\underset{*}{\to}} CvM^{dpro}_\infty$ *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* $\{(T_i, X_i')'\}_{i=1}^n$.

Theorem 4.1 states that the bootstrapped test statistic $CvM^{dpro,*}_n$ converges to the null limiting distribution of CvM^{dpro}_n conditional on the original sample. The fact that $CvM^{dpro,*}_n$ has the same limiting distribution under H_0 as CvM^{dpro}_n ensures size control, while the divergence of CvM^{dpro}_n under H_1, together with the convergence of $CvM^{dpro,*}_n$ to the same limiting distribution under any scenario, provides power to the proposed multiplier bootstrap test.

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, gaining additional insights into potential reasons for rejecting the putative model is essential.

There has been a large amount of interest in statistical inference in (single) index models, see, e.g. Stute and Zhu (2005), Maistre and Patilea (2019), and references therein. In the following, we investigate two relevant testing problems within 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.

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, $q_t(X,\theta) = q_t(X'\theta)$, where $X'\theta$ 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) The multinomial ordered discrete choice model has an intercept that varies with t. With unordered multinomial choices, one usually adopts a multiple index model, $q_t(X,\theta) = q_t(X'\theta_1, \ldots, X'\theta_J)$ as in the multinomial logit model (5). Since the binary and ordered multinomial cases are special cases of the unordered multinomial one, we focus on the latter.

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

$$H_0^{im} : \mathbb{P}\left(\mathbb{E}\left[e^{im}(t;\theta_0)|X\right] = 0\right) = 1 \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \tag{14}$$

where $\theta_0 = (\theta_{0,1}', \ldots, \theta_{0,J}')'$, 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 (1) against (2) but with $q_t(X,\theta) = q_t(\widetilde{X}_\theta)$ and the generalised 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 generalised residual marked double-projected empirical process as a special case of $R_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n)\mathcal{P}_{n,t}^{im} 1(X_i'\beta \leq u), \quad \left(\beta', u\right)' \in \Pi_{pro},$$

where $e_i^{im}(t;\widehat{\theta}_n) \equiv 1(T_i = t) - q_t(\widetilde{X}_{i,\widehat{\theta}_n})$ is the generalised residual under the multiple index model and the projection operator $\mathcal{P}_{n,t}^{im} 1(X_i'\beta \leq u)$ is as defined in (7) but with 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

$$H_0^{im1} : \mathbb{P}\left(\mathbb{E}\left[e^{im}(t;\theta_0)|\widetilde{X}_{\theta_0}\right] = 0\right) = 1 \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T}, \tag{15}$$

against the directional alternative

$$H_1^{im1} : \mathbb{P}\left(\mathbb{E}\left[1(T=t) - \mu_t(\widetilde{X}_\theta)|\widetilde{X}_\theta\right] = 0\right) = 1 \quad \text{for any } \theta \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and some } t \in \mathcal{T}, \tag{16}$$

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

Stute and Zhu (2005) considered similar testing problems specialised to the single-index setup and constructed their test statistics based on an integrated sample version of the conditional mean $\mathbb{P}(T = t \mid X'\theta_0 = v)$. We also employ an integrated moment approach, but consider double-projections and allow for multiple-index models.

The key insight we provide here is to re-express the null hypothesis (15) as

$$H_0^{im1} : \mathbb{E}\left[\mathrm{e}^{im}(t;\theta_0)1\left(\widetilde{X}'_{\theta_0}\beta \leq u\right)\right] = 0 \text{ a.e. } \left(\beta', u\right)' \in \Pi_{pro}^{im},$$
$$\text{for some } \theta_0 \in \Theta \subset \mathbb{R}^{d_\theta} \text{ and all } t \in \mathcal{T},$$

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 characterisation of the null immediately suggests using the following generalised residual marked double-projected empirical process with *estimated* multiple indexes,

$$M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t; \widehat{\theta}_n) \mathcal{P}_{n,t}^{im} 1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \leq u), \quad \left(\beta', u\right)' \in \Pi_{pro}^{im},$$

where

$$\mathcal{P}_{n,t}^{im} 1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \leq u) \equiv 1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \leq u) - g_t(X_i, \widehat{\theta}_n)' \Delta_{n,t}^{-1}(\widehat{\theta}_n) G_{n,t}^{im}(\beta, u; \widehat{\theta}_n), \tag{17}$$

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

$$G_{n,t}^{im}(\beta, u; \widehat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} g_t(X_i, \widehat{\theta}_n) 1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \leq u).$$

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}:

$$CvM_n^{im,dpro} = \sum_{t \in \mathcal{T}} a_n(t) \int_{\Pi_{pro}^{im}} \left(M_{n,t}^{dpro}(\beta, u; \widehat{\theta}_n)\right)^2 F_{n,\beta,\widehat{\theta}_n}^{im}(du)\, \mathrm{d}\beta, \tag{18}$$

where $F_{n,\beta,\widehat{\theta}_n}^{im}(u) = n^{-1} \sum_{i=1}^{n} 1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \leq u)$ is the empirical distribution function of the one-dimensional projected estimated indexes $\{\widetilde{X}'_{i,\widehat{\theta}_n}\beta\}_{i=1}^{n}$ 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 (18) resembles CvM_n^{dpro} in (10), with the caveat that now we use the vector of estimated linear indexes $\{\widetilde{X}_{i,\widehat{\theta}_n}\}_{i=1}^{n}$, instead of the observed covariates $\{X_i\}_{i=1}^{n}$ in the projection steps. With that replacement, the computation of the test remains the same.

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 Section 8, we have that, uniformly in $(\beta', u)' \in \Pi_{pro}^{im}$ and for each $t \in \mathcal{T}$,

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n) \left(1(\widetilde{X}'_{i,\theta_0}\beta \leq u) - g_t(X_i,\widehat{\theta}_n)' \Delta_{n,t}^{-1}(\widehat{\theta}_n) G_{n,t}^{im}(\beta,u;\widehat{\theta}_n)\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) \left(1(\widetilde{X}'_{i,\theta_0}\beta \leq u) - g_t(X_i,\theta_0)' \Delta_t^{-1}(\theta_0) G_t^{im}(\beta,u;\theta_0)\right) + o_p(1)$$

$$\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) \mathcal{P}_t^{im} 1(\widetilde{X}'_{i,\theta_0}\beta \leq u) + o_p(1),$$

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* (15), *for any sequence $a_n(t) = a(t) + o_p(1)$, with $a(t) > 0$ and $0 < \sum_{t\in\mathcal{T}} a(t) < \infty$, 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

$$\mathbb{K}_t^{im,dpro}\left((\beta_1,u_1),(\beta_2,u_2)\right) = \mathbb{E}\left[\widetilde{\sigma}_t^2(\widetilde{X}_{\theta_0}) \mathcal{P}_t^{im} 1\left(\widetilde{X}'_{\theta_0}\beta_1 \leq u_1\right) \mathcal{P}_t^{im} 1\left(\widetilde{X}'_{\theta_0}\beta_2 \leq u_2\right)\right],$$

where $\widetilde{\sigma}_t^2(\widetilde{X}_{\theta_0}) = \mathbb{E}[\varepsilon^2(t) \mid \widetilde{X}_{\theta_0}] = q_t(\widetilde{X}_{\theta_0})(1 - q_t(\widetilde{X}_{\theta_0}))$ is the conditional variance function of the generalised 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 $p_t(x) = \mathbb{P}(T = t \mid X = x)$ 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 $(\beta', u)' \in \Pi_{pro}^{im}$,

$$\frac{M_{n,t}^{dpro}(\beta,u;\widehat{\theta}_n)}{\sqrt{n}} \xrightarrow{p} \mathbb{E}\left[\left(p_t(X) - q_t(\widetilde{X}_{\theta^*})\right) \mathcal{P}_t^{im} 1\left(\tilde{X}'_{\theta^*}\beta \leq u\right)\right],$$

for each $t \in \mathcal{T}$ under H_1^{im1}, where, using obvious notation, $\mathcal{P}_t^{im} 1(\widetilde{X}'_{\theta^*}\beta \leq u)$ is the probability limit of (17).

To investigate the power properties of our test, we write

$$\begin{aligned}
&\mathbb{E}\left[\left(p_t(X) - q_t(\widetilde{X}_{\theta^*})\right) \mathcal{P}_t^{im} 1(\widetilde{X}'_{\theta^*}\beta \leq u)\right] \\
&\quad = \mathbb{E}\left[\left(p_t(X) - \mu_t(\widetilde{X}_{\theta^*})\right) \mathcal{P}_t^{im} 1(\widetilde{X}'_{\theta^*}\beta \leq u)\right] \\
&\qquad + \mathbb{E}\left[\left(\mu_t(\widetilde{X}_{\theta^*}) - q_t(\widetilde{X}_{\theta^*})\right) \mathcal{P}_t^{im} 1(\widetilde{X}'_{\theta^*}\beta \leq u)\right].
\end{aligned}$$

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 $q_t(v) - \mu_t(v)$ is a non-zero function not collinear to the score $g_t(X,\theta^*)$ would guarantee $\mathbb{E}[(\mu_t(\widetilde{X}_{\theta^*}) - q_t(\widetilde{X}_{\theta^*}))\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}'\beta \le u)] \ne 0$, thus implying that our test is consistent against such deviations in (16).

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 $\mathbb{E}[(p_t(X) - q_t(\widetilde{X}_{\theta^*}))\mathcal{P}_t^{im}1(\widetilde{X}_{\theta^*}'\beta \le u)] = 0$, even if the null hypothesis H_0^{im1} is false. Thus, 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 (15). 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 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 analogue of $CvM_n^{im,dpro}$, all one needs to do is replace $e_i(t;\widehat{\theta}_n)$ by $e_i^{im}(t;\widehat{\theta}_n)$. Given our results in Theorem 4.1, it is easy to show that the multiplier bootstrap analogue 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 the arguments.

6. Monte Carlo simulations

In this section, we conduct a series of Monte Carlo experiments to study the finite sample properties of the double projection-based tests in the context of observational treatment effect studies. We consider three different setups with (i) a binary treatment, (ii) multinomial, unordered treatments (henceforth multinomial treatments), and (iii) multinomial, ordered treatments (henceforth ordered treatments). For the sake of space, the Monte Carlo simulations for ordered treatments are gathered in the online supplementary appendix.

We compare the finite sample performance of our proposed omnibus test statistic CvM_n^{dpro} given in (11) with the directional test statistic $CvM_n^{im,dpro}$ given in (18), as well as with the Sant'Anna and Song (2019)'s CvM test statistic given for the binary treatment case by

$$CvM_n^{ss} = \frac{1}{n}\sum_{i=1}^{n}\left(R_n^{ss}\left(q\left(X_i,\widehat{\theta}_n\right)\right)\right)^2, \tag{19}$$

where $q(X_i,\widehat{\theta}_n) = \Phi(X_i'\widehat{\theta}_n)$ and $R_n^{ss}(u) \equiv n^{-1/2}\sum_{j=1}^{n} e_j(\widehat{\theta}_n)\mathcal{P}_n 1(q(X_j,\widehat{\theta}_n) \le u)$, with $e(\theta) = T - q(X,\theta)$,

$$\mathcal{P}_n 1\left(q(X,\theta) \le u\right) = 1\left(q(X,\theta) \le u\right) - g(X,\theta)'\Delta_n^{-1}(\theta)\, G_n(u,\theta),$$

$G_n(u,\theta) = n^{-1}\sum_{i=1}^{n} g(X_i,\theta)1(q(X_i,\theta) \le u)$, and $\Delta_n(\theta) = n^{-1}\sum_{i=1}^{n} g(X_i,\theta)g'(X_i,\theta)$.

To better understand these tests, consider for simplicity the binary case, and note that CvM_n^{dpro}, $CvM_n^{im,dpro}$ and CvM_n^{ss} focus on the respective hypotheses $\mathbb{E}[e(\theta_0)\,|\,X] = 0$, $\mathbb{E}[e(\theta_0)\,|\,\widetilde{X}_{\theta_0}] = 0$, and $\mathbb{E}[e(\theta_0)\,|\,q(X,\theta_0)] = 0$, almost surely, where $e(\theta) = T - q(X,\theta)$ is the parametric error term, $\widetilde{X}_{\theta_0}$ is the index and $q(X,\theta_0)$ the propensity score. Since the last two hypotheses are implications of the first, CvM_n^{dpro} leads to an omnibus test, while

$CvM_n^{im,dpro}$ and CvM_n^{ss} lead to directional ones. Depending on the alternative, one test may be better than the other, so we do not expect a uniformly larger power for the omnibus test over the directional tests. However, we expect the proposed omnibus test to be consistent against more alternatives than the directional tests.

Given that the simulation results in Sant'Anna and Song (2019) indicate that their test dominates several others in terms of size and power in the binary treatment setup, we only compare our proposed CvM_n^{dpro} statistic given in (11) to their CvM statistic. For the ordered and multinomial treatment setups, we consider extensions of the Sant'Anna and Song (2019)'s projection-based tests that are able to accommodate multi-valued treatment variables.

Critical values for the test statistic CvM_n^{dpro} are obtained using the multiplier bootstrap procedure described in Section 4, whereas for the single projection-based tests, we use the multiplier bootstrap procedure described in Sant'Anna and Song (2019). We consider sample sizes n equal to 200, 400, and 800. For each design, we consider 1000 Monte Carlo experiments. The multipliers $\{V_i,\ i = 1, \ldots, n\}$ used in the bootstrap implementations are independently generated as V with $\mathbb{P}(V = 1 - \kappa) = \kappa/\sqrt{5}$ and $\mathbb{P}(V = \kappa) = 1 - \kappa/\sqrt{5}$, where $\kappa = (\sqrt{5} + 1)/2$, as proposed by Mammen (1993). The bootstrapped critical values are approximated using $B = 999$ bootstrap replications. Moreover, we report simulation results from the test statistic $CvM_n^{im,dpro}$, which is used to test a parametric (multiple) index propensity score model against a semiparametric alternative. In all simulations, we use the constant weight $a_n(t) \equiv 1$ for all $t \in \mathcal{T}$ in CvM_n^{dpro} and $CvM_n^{im,dpro}$ as well as their bootstrap counterparts $CvM_n^{dpro,*}$ and $CvM_n^{im,dpro,*}$.

6.1. Binary treatment

We first consider the binary treatment case with $J = 1$. Consider the following data generating processes (DGPs), which are similar to Sant'Anna and Song (2019):

$$DGP1.\ T^* = -\frac{\sum_{j=1}^{10} X_j}{6} - \varepsilon;$$

$$DGP2.\ T^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 X_2}{2} - \varepsilon;$$

$$DGP3.\ T^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 \sum_{k=2}^{5} X_k}{4} - \varepsilon;$$

$$DGP4.\ T^* = -1.5 - \frac{\sum_{j=1}^{10} X_j}{6} + \frac{\sum_{k=1}^{10} X_k^2}{10} - \varepsilon;$$

$$DGP5.\ T^* = \frac{-0.1 + 0.1 \sum_{j=1}^{5} X_j}{\exp\left(-0.2 \sum_{k=1}^{10} X_j\right)} - \varepsilon.$$

For each of these five DGPs, $T = 1\{T^* > 0\}$, $\varepsilon \perp\!\!\!\perp X$, with $X = (1, X_1, X_2, \ldots, X_{10})'$ where $X_1 = Z_1$, $X_2 = (Z_1 + Z_2)/\sqrt{2}$, $X_k = Z_k$, $k = 3, \ldots, 10$, and $\{Z_k\}_{k=1}^{10}$ and ε are independent standard normal random variables. For each of these DGPs we consider the following potential outcomes:

$$Y(1) = 2m_1(X) + u(1) \quad \text{and} \quad Y(0) = m_1(X) + u(0),$$

where $m_1(X) = 1 + \sum_{j=1}^{10} X_j$, $u(1)$ and $u(0)$ are independent normal random variables with mean zero and variance one. The observed outcome is $Y = TY(1) + (1-T)Y(0)$, and the true average treatment effect (ATE) is 1. Although these outcome equations are not necessary to assess the size and power properties of the tests, they can be used to assess the utility of our proposed tests to distinguish between 'good' and 'bad' estimates of the ATE.

For $DGP1-DGP5$, the null hypothesis H_0 considered is

$$H_0 : \exists\theta_0 = (\delta_0, \delta_1, \ldots, \delta_{10})' \in \Theta : \mathbb{E}[T \mid X] = \Phi\left(X'\theta_0\right) \; a.s., \tag{20}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Note that under the null hypothesis in (20), the propensity score model satisfies a parametric single-index restriction with $q(x,\theta) = \Phi(x'\theta)$ and the corresponding score function $g(x,\theta) = \partial q(x,\theta)/\partial\theta = \phi(x'\theta)x$, with $\phi(\cdot)$ the PDF of the standard normal distribution. We estimate θ_0 using the probit maximum likelihood, i.e.

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \left[T_i \ln\left(\Phi\left(X_i'\theta\right)\right) + (1-T_i)\ln\left(1-\Phi\left(X_i'\theta\right)\right)\right].$$

Clearly, $DGP1$ falls under H_0, whereas $DGP2-DGP5$ fall under H_1, i.e. the negation of (20). Note that the treatment status T follows a heteroskedastic probit model in $DGP5$.

The simulation results are presented in Table 1. We report empirical rejection frequencies at the 5% significance level. Results for 10% and 1% significance levels are similar and are gathered in the online supplementary appendix. We also report the bias, root mean squared error (RMSE), and coverage of the 95% confidence interval of the following *stabilized* inverse probability weighted estimator:

$$ATE_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{w_{i,1}}{\bar{w}_{n,1}} - \frac{w_{i,0}}{\bar{w}_{n,0}}\right) Y_i, \tag{21}$$

where $w_{i,1} = T_i/q(X_i,\widehat{\theta}_n)$ and $w_{i,0} = (1-T_i)/(1-q(X_i,\widehat{\theta}_n))$ are the inverse probability weights (Hájek 1971), and $\bar{w}_{n,t}$ is the sample mean of $w_{i,t}$ for $t = 0, 1$. The 95% confidence interval is estimated via the percentile bootstrap with 499 draws.

We first analyze the size of our proposed test. From the results of $DGP1$, we find that the actual finite sample sizes of all considered tests are close to their nominal size, even when the sample size is as small as 200. In addition, when the propensity score is correctly specified, the bias of the ATE_n estimator in (21) is small, and the coverage probability is close to its nominal value even when $n = 200$.

Note that when the propensity score is misspecified in $DGP2$–$DGP5$, the ATE estimator (21) can be severely biased and its 95% confidence interval is too liberal, i.e. it can severely undercover the true ATE. Thus, tests with a higher power to detect such model misspecifications can prevent one from making misleading conclusions about the effectiveness of a given policy. Our proposed CvM_n^{dpro} test performs admirably in this task. Perhaps

Table 1. Monte Carlo results under designs *DGP*1–*DGP*5: Binary Treatment.

DGP	n	CvM_n^{dpro}	$CvM_n^{im,dpro}$	CvM_n^{ss}	Bias	RMSE	COV
1	200	0.060	0.059	0.057	−0.085	0.623	0.962
1	400	0.053	0.057	0.056	−0.007	0.504	0.948
1	800	0.053	0.055	0.054	−0.006	0.381	0.936
2	200	0.648	0.147	0.154	0.582	1.301	0.986
2	400	0.990	0.237	0.264	0.613	0.958	0.894
2	800	1.000	0.458	0.495	0.589	0.709	0.671
3	200	0.356	0.175	0.183	1.106	1.768	0.976
3	400	0.885	0.444	0.465	1.359	1.764	0.705
3	800	0.999	0.770	0.799	1.330	1.530	0.256
4	200	0.368	0.148	0.151	0.603	1.619	0.975
4	400	0.856	0.267	0.304	0.963	1.772	0.947
4	800	1.000	0.509	0.576	1.031	1.424	0.708
5	200	0.123	0.098	0.100	−0.106	0.358	0.974
5	400	0.265	0.191	0.192	−0.105	0.244	0.947
5	800	0.590	0.497	0.502	−0.118	0.192	0.892

Note: Simulations based on 1000 Monte Carlo experiments. 'CvM_n^{dpro}' stands for our proposed double-projected Cramér-von Mises test. '$CvM_n^{im,dpro}$' stands for the directional Cramér-von Mises test. 'CvM_n^{ss}' stands for Sant'Anna and Song (2019)'s test defined in (19). Finally, 'Bias', 'RMSE', and 'COV' stand for the average simulated bias, average simulated root mean squared error, and 95% coverage probability for the *ATE* estimator ATE_n as defined in (21). The 95% coverage probability is based on the percentile bootstrap with 499 draws. See the main text for further details.

what is more important to emphasise in terms of power is that in all alternative hypotheses and sample sizes analysed, our omnibus CvM_n^{dpro} test has substantially higher power than our directional test, $CvM_n^{im,dpro}$, and Sant'Anna and Song (2019)'s CvM_n^{ss} test. For instance, for $DGP2$ with $n = 200$, our proposed omnibus test is four times more powerful than Sant'Anna and Song (2019)'s test, which is already more powerful than other tests available in the literature, including covariate balancing tests and traditional specification tests based on the kernel methods; see Section 4.2 of Sant'Anna and Song (2019) for additional details. In addition, the power properties of $CvM_n^{im,dpro}$ and CvM_n^{ss} are similar. This, however, is not surprising given that the one-dimensional estimated conditioning variable $\Phi(X_i'\widehat{\theta}_n)$ in CvM_n^{ss} is only a strictly increasing transformation of the one-dimensional estimated linear-index $X_i'\widehat{\theta}_n$ (i.e. $\widetilde{X}_{i,\widehat{\theta}_n}$) in $CvM_n^{im,dpro}$.

6.2. Multinomial treatments

In this section, we consider unordered, multinomial treatments. Our DGPs are similar to Yang et al. (2016). The covariates X_1, X_2, X_3 are generated from a multivariate normal distribution with mean zero, variances of $(2, 1, 1)$ and covariances of $(1, -1, -0.5)$; X_4 follows a uniform distribution from -3 to 3; X_5 follows a chi-squared distribution with one degree of freedom; and X_6 follows a Bernoulli distribution with $p = 0.5$. Let $X = (X_1, \ldots, X_6)'$. We consider three treatment groups, $T = \{0, 1, 2\}$, whose assignment mechanism follows the multinomial logistic model

$$(T_0, T_1, T_2) \mid X \sim Multinomial\ \left(p_0\left(X\right), p_1\left(X\right), p_2\left(X\right)\right),$$

where T_t is the treatment indicator, i.e. $T_t = 1(T = t)$, and for $t = 0, 1, 2$,

$$p_t\left(X\right) = \frac{\exp\left(\phi_t\left(X\right)\right)}{\sum_{s=0}^{2} \exp\left(\phi_s\left(X\right)\right)}.$$

In what follows, we take $\phi_0(X) = 0$ and vary $\phi_1(X)$ and $\phi_2(X)$ as follows:

$$DGP6.\ \phi_1(X) = -1 + 0.4\sum_{j=1}^{6} X_j, \quad \phi_2(X) = -1 + 0.2\sum_{j=1}^{6} X_j;$$

$$DGP7.\ \phi_1(X) = -0.2\sum_{j=1}^{6} X_j + X_1X_6, \quad \phi_2(X) = -0.1\sum_{j=1}^{6} X_j + X_1X_4;$$

$$DGP8.\ \phi_1(X) = 0.3\sum_{j=1}^{6} X_j, \quad \phi_2(X) = -0.5 + 0.1\sum_{j=1}^{6} X_j^2;$$

$$DGP9.\ \phi_1(X) = -0.1 + \frac{\sum_{j=1}^{6} X_j}{5} + \frac{X_6\sum_{l=1}^{3} X_l}{2},$$

$$\phi_2(X) = -0.3\sum_{j=1}^{6} X_j - \frac{X_6(X_4 + X_5)}{2};$$

$$DGP10.\ \phi_1(X) = \sin\left(\sum_{j=1}^{6} X_j\right) + \sum_{l=1}^{3} X_l, \quad \phi_2(X) = 2\sin\left(\sum_{j=1}^{6} X_j\right) + \frac{1}{2}\sum_{l=1}^{3} X_l.$$

For each of these DGPs, we consider the potential outcomes

$$Y(0) = 1 + X'\beta_0 + u(0), \quad Y(1) = 20 + X'\beta_1 + u(1), \quad \text{and}$$
$$Y(2) = 6 + X'\beta_2 + u(2),$$

where $u(0)$, $u(1)$ and $u(2)$ are independent normal random variables with mean zero and variance 1, $\beta_0 = (-6, -6, -6, 6, 6, 6)'$, $\beta_1 = -\beta_0$, and $\beta_2 = (4, 4, 4, 4, 4, 4)'$. The observed outcome is $Y = 1(T = 0)Y(0) + 1(T = 1)Y(1) + 1(T = 2)Y(2)$, and the true $ATE_{1,0} = 1$, $ATE_{2,0} = 2$, and $ATE_{2,1} = 1$, where $ATE_{t,s} = \mathbb{E}[Y(t) - Y(s)]$.

Let $\alpha = (\alpha_1, \alpha_2)'$ and $\delta = (\delta_1', \delta_2')'$. For $DGP6-DGP10$, the H_0 considered is

$$H_0 : \exists\theta_0 = (\alpha', \delta')' \in \Theta : \mathbb{P}(T = t \,|\, X) = \frac{\exp\left(\alpha_t + X'\delta_t\right)}{\sum_{s=0}^{2}\exp\left(\alpha_s + X'\delta_s\right)} \quad a.s., \quad \text{for } t = 1, 2, \tag{22}$$

where, with some abuse of notation, we set $\alpha_0 = \delta_0 = 0$. Under (22), the GPS satisfies parametric single-index restrictions with the link function being the multinomial logistic CDF. We estimate θ_0 using the multinomial logit likelihood, i.e.

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta}\sum_{i=1}^{n}\sum_{t=0}^{2}\left[T_{i,t}\cdot\ln\left(\frac{\exp\left(\alpha_t + X_i'\delta_t\right)}{\sum_{s=0}^{2}\exp\left(\alpha_s + X_i'\delta_s\right)}\right)\right].$$

Clearly, $DGP6$ falls under H_0, whereas $DGP7-DGP10$ fall under H_1, i.e. the negation of (22). Note that either interactive, quadratic, or periodic functions of components in X enter $DGP7-DGP10$.

In the multinomial setup, our proposed test statistic CvM_n^{dpro} is given by (11), where, for $t = 1, 2$, $a_n(t) = 1$, the residual $e_i(t;\widehat{\theta}_n) = 1(T_i = t) - q_t(X_i, \widehat{\theta}_n)$,

$$q_t\left(X_i, \widehat{\theta}_n\right) = \Lambda_t^m\left(X_i, \theta\right) = \frac{\exp\left(\alpha_t + X_i'\delta_t\right)}{\sum_{s=0}^{2}\exp\left(\alpha_s + X_i'\delta_s\right)},$$

and the score function

$$g_t(X_i, \widehat{\theta}_n) = q_t(X_i, \widehat{\theta}_n)(1 - q_t(X_i, \widehat{\theta}_n))(1\ X_i)'.$$

Although Sant'Anna and Song (2019) only considered specification tests for binary treatments, we note that their tests can also be extended to test (22). We consider two different extensions of Sant'Anna and Song (2019)'s specification tests. The first variant of Sant'Anna and Song (2019)'s test statistic is given by

$$CvM_n^{ss1,m} = CvM_{n,1}^{ss1,m} + CvM_{n,2}^{ss1,m}, \tag{23}$$

where, for $t = 1, 2$, $CvM_{n,t}^{ss1,m} = n^{-1}\sum_{i=1}^{n}(R_{n,t}^{ss1,m}(q_1(X_i, \widehat{\theta}_n), q_2(X_i, \widehat{\theta}_n)))^2$, with

$$R_{n,t}^{ss1,m}(u_1, u_2) \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n} e_j(t;\widehat{\theta}_n)\mathcal{P}_{n,t}\left(1\left(\Lambda_1^m\left(X_j, \widehat{\theta}_n\right) \leq u_1\right)1\left(\Lambda_2^m\left(X_j, \widehat{\theta}_n\right) \leq u_2\right)\right),$$

and

$$\begin{aligned}&\mathcal{P}_{n,t}\left(1\left(\Lambda_1^m\left(X, \theta\right) \leq u_1\right)1\left(\Lambda_2^m\left(X, \theta\right) \leq u_2\right)\right)\\ &\quad = 1\left(\Lambda_1^m\left(X, \theta\right) \leq u_1\right)1\left(\Lambda_2^m\left(X, \theta\right) \leq u_2\right) - g_tX, \theta)'\Delta_{n,t}^{-1}\left(\theta\right)G_{n,t}^m\left(u_1, u_2; \theta\right),\end{aligned}$$

$g_t(X, \theta)$ and $\Delta_{n,t}^{-1}(\theta)$ being defined as before, and

$$G_{n,t}^m\left(u_1, u_2; \theta\right) = \frac{1}{n}\sum_{i=1}^{n} g_t(X_i, \theta)1\left(\Lambda_1^m\left(X, \theta\right) \leq u_1\right)1\left(\Lambda_2^m\left(X, \theta\right) \leq u_2\right).$$

The second variant of Sant'Anna and Song (2019) test statistic is given by

$$CvM_n^{ss2,m} = CvM_{n,1}^{ss2,m} + CvM_{n,2}^{ss2,m}, \tag{24}$$

where, for $t = 1, 2$, $CvM_{n,t}^{ss2,m} = n^{-1}\sum_{i=1}^{n}(R_{n,t}^{ss2,m}(\Lambda_t^m(X_i, \widehat{\theta}_n)))^2$, and

$$R_{n,t}^{ss2,m}(u) \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n} e_j(t;\widehat{\theta}_n)\mathcal{P}_{n,t}1\left(\Lambda_t^m\left(X_j, \widehat{\theta}_n\right) \leq u\right),$$

with

$$\mathcal{P}_{n,t}1\left(\Lambda_t^m\left(X, \theta\right) \leq u\right) = 1\left(\Lambda_t^m\left(X, \theta\right) \leq u\right) - g_t(X, \theta)'\Delta_{n,t}^{-1}\left(\theta\right)G_{n,t}\left(u, \theta\right),$$

and $g_t(X, \theta)$, $G_{n,t}(u, \theta)$ and $\Delta_{n,t}(\theta)$ defined analogously to the binary treatment setup.

It is important to emphasise that (23) and (24) are test statistics for *implications* of (22), and not for (22) itself. More precisely, (23) is a test statistic for the null hypothesis

$$H_0' : \exists\theta_0 = (\alpha', \delta')' \in \Theta : \mathbb{E}\left[e^m(t;\theta_0)|\Lambda_1^m(X,\theta_0), \Lambda_2^m(X,\theta_0)\right] = 0 \quad a.s. \text{ for } t = 0, 1,$$

whereas (24) is a test statistic based on the null hypothesis

$$H_0'' : \exists\theta_0 = (\alpha', \delta')' \in \Theta : \mathbb{E}\left[e^m(t;\theta_0)|\Lambda_t^m(X,\theta_0)\right] = 0 \quad a.s. \text{ for } t = 0, 1.$$

Importantly and in sharp contrast to CvM_n^{dpro}, both $CvM_n^{ss1,m}$ and $CvM_n^{ss2,m}$ are not consistent against general nonparametric alternatives H_1 in (2). Similarly, our directional test statistic $CvM_n^{im,dpro}$ for H_0^{im1} is consistent against the semiparametric alternatives H_1^{im1} in (16), but not necessarily consistent against general nonparametric alternatives H_1^{im}, the negation of H_0^{im} in (14).

The simulation results are presented in Table 2. We report empirical rejection frequencies at the 5% significance level. We also report the bias, RMSE, and coverage of the 95% confidence interval of the following *stabilized* inverse probability weighted estimators based on the identification result in (4):

$$ATE_{n,j,\ell} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{w_{i,j}}{\bar{w}_{n,j}} - \frac{w_{i,\ell}}{\bar{w}_{n,\ell}}\right)Y_i, \tag{25}$$

where

$$w_{i,0} = \frac{1\{T_i = 0\}}{q_0\left(X_i, \widehat{\theta}_n\right)}, \quad w_{i,1} = \frac{1\{T_i = 1\}}{q_1\left(X_i, \widehat{\theta}_n\right)} \quad \text{and} \quad w_{i,2} = \frac{1\{T_i = 2\}}{q_2\left(X_i, \widehat{\theta}_n\right)}$$

are the inverse probability weights, and $\bar{w}_{n,t}$ is the sample mean of $w_{i,t}$ for $t = 0, 1, 2$. The 95% confidence interval is estimated via the percentile bootstrap with 499 draws.

As before, we first discuss the size properties of the tests. From the results of $DGP6$, we find that all considered tests have good size properties, and the IPW estimators for the average treatment effects exhibit little to no bias. Their RMSEs decrease with sample size, and their coverage probabilities are very close to the nominal level. Among the considered test statistics, our directional test $CvM_n^{im,dpro}$ is the only one with some size distortions when sample size $n = 200$, but such distortions disappear as n increases.

In terms of power, note that, under $DGP7-DGP10$, our proposed double-projection omnibus test CvM_n^{dpro} outperforms $CvM_n^{im,dpro}$, $CvM_n^{ss1,m}$ and $CvM_n^{ss2,m}$ in all considered scenarios by a significant margin. We note that $CvM_n^{ss2,m}$ tends to outperform $CvM_n^{ss1,m}$ in all DGPs, except $DGP10$. We also note that the power performance of $CvM_n^{im,dpro}$ and $CvM_n^{ss2,m}$ is similar. Finally, it is evident from Table 2 that GPS misspecification can indeed lead to misleading conclusions about the treatment effect effectiveness.

7. Conclusions and directions for further research

In this article, we propose a new class of specification tests for GPS models based on novel double-projected weight functions. We have demonstrated that using double projections helps alleviate the 'curse of dimensionality' and avoids the complications associated with

Table 2. Monte Carlo results under designs $DGP6$–$DGP10$: Multinomial Treatment.

DGP	n	CvM_n^{dpro}	$CvM_n^{im,dpro}$	$CvM_n^{ss1,m}$	$CvM_n^{ss2,m}$	$Bias_{1,0}$	$Bias_{2,0}$	$RMSE_{1,0}$	$RMSE_{2,0}$	$COV_{1,0}$	$COV_{2,0}$
6	200	0.054	0.064	0.051	0.048	0.286	0.075	3.493	2.458	0.955	0.970
6	400	0.057	0.049	0.044	0.063	0.033	0.064	2.509	1.744	0.943	0.953
6	800	0.060	0.053	0.044	0.059	0.045	0.063	1.722	1.165	0.935	0.954
7	200	0.993	0.254	0.219	0.331	0.126	2.074	3.409	2.772	0.961	0.838
7	400	1.000	0.499	0.442	0.536	−0.166	1.974	2.363	2.278	0.962	0.637
7	800	1.000	0.770	0.718	0.728	−0.005	2.037	1.536	2.184	0.962	0.271
8	200	0.445	0.122	0.074	0.132	1.122	1.172	3.113	2.303	0.926	0.914
8	400	0.835	0.222	0.087	0.240	0.859	1.038	2.141	1.649	0.917	0.865
8	800	0.992	0.475	0.140	0.473	0.906	1.045	1.615	1.369	0.886	0.784
9	200	0.099	0.084	0.057	0.079	0.746	−0.829	3.232	3.815	0.945	0.888
9	400	0.149	0.123	0.076	0.116	0.558	−0.132	2.185	3.524	0.938	0.863
9	800	0.258	0.174	0.061	0.137	0.494	−0.167	1.608	3.042	0.927	0.834
10	200	0.133	0.076	0.090	0.075	−0.528	−1.369	4.043	2.803	0.954	0.872
10	400	0.302	0.102	0.157	0.137	−0.548	−1.108	3.406	2.347	0.939	0.823
10	800	0.747	0.176	0.332	0.281	−0.843	−1.285	2.424	1.818	0.921	0.689

Note: Simulations based on 1, 000 Monte Carlo experiments. 'CvM_n^{dpro}' stands for our proposed Cramér-von Mises tests. '$CvM_n^{im,dpro}$' stands for the directional Cramér-von Mises test. '$CvM_n^{ss1,m}$' and '$CvM_n^{ss2,m}$' stand for extensions of Sant'Anna and Song (2019)'s test defined in (23) and (24), respectively. Finally, '$Bias_{k,s}$', '$RMSE_{k,s}$', and '$COV_{k,s}$' stand for the average simulated bias, average simulated root mean squared error, and 95% coverage probability for the $ATE_{k,s}$ estimator $ATE_{n,k,s}$ as defined in (25). The 95% coverage probability is based on the percentile bootstrap with 499 draws. See the main text for further details.

'parameter estimation uncertainty' commonly encountered in specification testing. We have shown that our proposed CvM-type test statistics can be expressed in closed-form expressions 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 GPS. The simulation results and the empirical application (in the online supplementary appendix) 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 regressions are correctly specified; see, e.g. Bierens and Wang (2012), Rothe and Wied (2013), and Escanciano and Goh (2014). Another direction of future research could focus on developing tests for a semiparametric single-index assumption against a general nonparametric alternative. We leave a detailed analysis of these interesting extensions for future research.

8. Mathematical proofs

Proof: For notational simplicity, let us focus on $J = 1$ (the binary treatment case) and omit the dependence of $e_i(t;\widehat{\theta}_n)$ and $\mathcal{P}_{n,t}1(\beta' X_i \leq u)$ on t. The general case with $J \geq 2$ follows in a similar way.

First, we can rewrite CvM_n^{dpro} as

$$CvM_n^{dpro} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} \mathcal{P}_n 1\left(\beta' X_i \leq \beta' X_r\right)\mathcal{P}_n 1\left(\beta' X_j \leq \beta' X_r\right) \mathrm{d}\beta.$$

Recalling the expression of projection operator $\mathcal{P}_n 1(\beta' X_i \leq u)$, by simple algebra, CvM_n^{dpro} is further equal to

$$\begin{aligned}
CvM_n^{dpro} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} 1\left(\beta' X_i \leq \beta' X_r\right) 1\left(\beta' X_j \leq \beta' X_r\right) \mathrm{d}\beta \\
&\quad - 2\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n) \\
&\quad \times \int_{\mathbb{S}^{d_x}} 1\left(\beta' X_i \leq \beta' X_r\right) 1\left(\beta' X_s \leq \beta' X_r\right) \mathrm{d}\beta \\
&\quad + \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)) \\
&\quad \times g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} 1\left(\beta' X_s \leq \beta' X_r\right) 1\left(\beta' X_t \leq \beta' X_r\right) \mathrm{d}\beta \\
&\equiv B_{n1} - 2B_{n2} + B_{n3}.
\end{aligned}$$

As in Escanciano (2006a), if denoting $A_{ijr} = \int_{\mathbb{S}^{d_x}} 1(\beta' X_i \leq \beta' X_r)1(\beta' X_j \leq \beta' X_r)d\beta$, we have that

$$B_{n1} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)A_{ijr}.$$

Similarly,

$$\begin{aligned}
B_{n2} &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)A_{isr} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} e_i(\widehat{\theta}_n)\left[g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^{n} g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right]A_{ijr},
\end{aligned}$$

where the second equality follows by letting $j = s$, and

$$\begin{aligned}
B_{n3} &= \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)) \\
&\quad \times g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)A_{str} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[g'(X_i,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^{n} g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right] \\
&\quad \times \left[g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{t=1}^{n} g(X_t,\widehat{\theta}_n)e_t(\widehat{\theta}_n)\right]A_{ijr},
\end{aligned}$$

where the second equality follows by letting $i = s$ and $j = t$.

As a result,

$$
\begin{aligned}
CvM_n^{dpro} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[e_i(\widehat{\theta}_n) - g'(X_i,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^{n} g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right] \\
&\quad \times \left[e_j(\widehat{\theta}_n) - g'(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{t=1}^{n} g(X_t,\widehat{\theta}_n)e_t(\widehat{\theta}_n)\right] A_{ijr} \\
&\equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} e_i^{pro}(\widehat{\theta}_n)e_j^{pro}(\widehat{\theta}_n)A_{ijr}.
\end{aligned}
$$

This completes the proof of Lemma 2.1. ■

The proof of Theorem 3.1 is similar to the proofs of Theorem 1 and Corollary 1 in Sant'Anna and Song (2019). We first need to introduce several auxiliary lemmas. Henceforth, $x = (\beta', u)'$. The next lemma establishes the uniform asymptotic decomposition of the 'once' projected empirical process

$$
R_{n,t}^{pro}(x;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)1(\beta' X_i \le u), \quad x \in \Pi_{pro}.
$$

Lemma 8.1: *Suppose Assumptions* 3.1– 3.3 *hold. Then, for each $t \in \mathcal{T}$, we have that*

$$
\sup_{x\in\Pi_{pro}}\left| R_{n,t}^{pro}(x;\widehat{\theta}_n) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1(\beta' X_i \le u) + \sqrt{n}(\widehat{\theta}_n - \theta_0)' G_t(x;\theta_0)\right| = o_p(1),
$$

where $G_t(x;\theta) = \mathbb{E}[g_t(X,\theta)1(\beta' X \le u)]$ with $g_t(x,\theta) = \partial q_t(x,\theta)/\partial\theta$.

Proof: First note that, for each $t \in \mathcal{T}$, $R_{n,t}^{pro}(x;\widehat{\theta}_n)$ can be readily decomposed as

$$
\begin{aligned}
&R_{n,t}^{pro}(x;\widehat{\theta}_n) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1(\beta' X_i \le u) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(q_t(X_i,\widehat{\theta}_n) - q_t(X_i,\theta_0))1(\beta' X_i \le u).
\end{aligned}
$$

By the Mean Value Theorem (MVT) and Assumptions 3.2–3.3, the second term in the above expression is simply equal to

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta}_n - \theta_0)'\frac{1}{n}\sum_{i=1}^{n}\frac{\partial q_t(X_i,\widetilde{\theta}_n)}{\partial\theta}1(\beta' X_i \le u) \\
&\quad = \sqrt{n}(\widehat{\theta}_n - \theta_0)'\mathbb{E}\left[g_t(X,\theta_0)1(\beta' X_i \le u)\right] + o_p(1),
\end{aligned}
$$

with $\widetilde{\theta}_n$ lying between $\widehat{\theta}_n$ and θ_0, where the latter equality follows by the uniform law of large numbers (ULLN) of Newey and McFadden (1994, Lemma 2.4). This finishes the proof of Lemma 8.1. ■

To proceed, for each $t \in \mathcal{T}$, we introduce the following auxiliary quantity,

$$A_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t;\widehat{\theta}_n) g_t(X_i,\widehat{\theta}_n),$$

which admits the following decomposition.

Lemma 8.2: *Suppose Assumptions* 3.1– 3.3 *hold. Then, under the null hypothesis H_0 in* (1), *for each $t \in \mathcal{T}$, we have that*

$$A_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t;\theta_0) g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1),$$

where $\Delta_t(\theta) = \mathbb{E}[g_t(X,\theta)g_t'(X,\theta)]$.

Proof: We can rewrite $A_{n,t}$ as

$$\begin{aligned}
A_{n,t} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t;\theta_0) g_t(X_i,\theta_0) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e_i(t;\widehat{\theta}_n) - e_i(t;\theta_0)) g_t(X_i,\theta_0) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t;\theta_0)(g_t(X_i,\widehat{\theta}_n) - g_t(X_i,\theta_0)) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e_i(t;\widehat{\theta}_n) - e_i(t;\theta_0))(g_t(X_i,\widehat{\theta}_n) - g_t(X_i,\theta_0)) \\
&\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i(t;\theta_0) g_t(X_i,\theta_0) + A_{n,t1} + A_{n,t2} + A_{n,t3}.
\end{aligned}$$

We first show that $A_{n,t1} = -\sqrt{n}(\widehat{\theta}_n - \theta_0)'\Delta_t(\theta_0) + o_p(1)$. To this end, note that

$$\begin{aligned}
A_{n,t1} &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (q_t(X_i,\widehat{\theta}_n) - q_t(X_i,\theta_0)) g_t(X_i,\theta_0) \\
&= -\frac{1}{n} \sum_{i=1}^{n} g_t(X_i,\theta_0) \frac{\partial q_t(X_i,\widetilde{\theta}_n)}{\partial \theta'} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\
&= -\mathbb{E}[g_t(X,\theta_0)g_t'(X,\theta_0)]\sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1),
\end{aligned}$$

with $\widetilde{\theta}_n$ lying between $\widehat{\theta}_n$ and θ_0, where the second equality follows from the MVT and the third equality follows from the ULLN of Newey and McFadden (1994, Lemma 2.4), and Assumptions 3.2 and 3.3.

It thus remains to show that both $A_{n,t2}$ and $A_{n,t3}$ are asymptotically negligible. To this end, note that

$$\begin{aligned} A_{n,t2} &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} e_i(t;\theta_0) \frac{\partial g_t(X_i, \widetilde{\theta}_n)}{\partial \theta} \\ &= \sqrt{n}(\widehat{\theta}_n - \theta_0)' \mathbb{E}\left[e(t;\theta_0) \frac{\partial g_t(X, \theta_0)}{\partial \theta} \right] + o_p(1) \\ &= o_p(1), \end{aligned}$$

with $\widetilde{\theta}_n$ again lying between $\widehat{\theta}_n$ and θ_0, where the first equality follows from MVT, the second equality by ULLN of Newey and McFadden (1994, Lemma 2.4), and the last equality by Assumptions 3.2–3.3 as well as the law of iterated expectations (LIE) under the null hypothesis H_0.

For the last term $A_{n,t3}$, we have

$$\begin{aligned} \sqrt{n} A_{n,t3} &= -\sum_{i=1}^{n} (q_t(X_i, \widehat{\theta}_n) - q_t(X_i, \theta_0))(g_t(X_i, \widehat{\theta}_n) - g_t(X_i, \theta_0)) \\ &= -\sqrt{n}(\widehat{\theta}_n - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} \frac{\partial q_t(X_i, \widetilde{\theta}_n)}{\partial \theta} \frac{\partial g_t(X_i, \breve{\theta}_n)}{\partial \theta'} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ &= -\sqrt{n}(\widehat{\theta}_n - \theta_0)' \mathbb{E}\left[g_t(X, \theta_0) \frac{\partial g_t(X, \theta_0)}{\partial \theta'} \right] \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &= O_p(1), \end{aligned}$$

with $\widetilde{\theta}_n$ and $\breve{\theta}_n$ (potentially different) both lying between $\widehat{\theta}_n$ and θ_0, where the second equality follows by MVT, the third equality by ULLN of Newey and McFadden (1994, Lemma 2.4), and the last equality by Assumptions 3.2 and 3.3. Thus, $A_{n,t3} = O_p(n^{-1/2}) = o_p(1)$. This completes the proof of Lemma 8.2. ■

The next two lemmas establish the uniform convergence of $G_{n,t}(x;\widehat{\theta}_n)$ and $\Delta_{n,t}^{-1}(\widehat{\theta}_n)$ to $G_t(x;\theta_0)$ and $\Delta_t^{-1}(\theta_0)$ for each $t \in \mathcal{T}$, respectively.

Lemma 8.3: *Suppose Assumptions* 3.1– 3.3 *hold. Then, for each* $t \in \mathcal{T}$, *we have that*

$$\sup_{x \in \Pi_{pro}} \left| G_{n,t}(x;\widehat{\theta}_n) - G_t(x;\theta_0) \right| = o_p(1).$$

Proof: The proof follows directly from the ULLN of Newey and McFadden (1994, Lemma 2.4). ■

Lemma 8.4: *Suppose Assumptions* 3.1– 3.3 *hold. Then, for each* $t \in \mathcal{T}$, *we have that*

$$\Delta_{n,t}^{-1}(\widehat{\theta}_n) = \Delta_t^{-1}(\theta_0) + o_p(1).$$

Proof: The proof follows from the ULLN of Newey and McFadden (1994, Lemma 2.4) and the continuous mapping theorem. ■

With the assistance of Lemmas 8.1–8.4, we are ready to proceed with the proofs of our main theorems.

Proof: We first establish the uniform asymptotic representation of $R^{dpro}_{n,t}(x;\widehat{\theta}_n)$, which states that $R^{dpro}_{n,t}(x;\widehat{\theta}_n)$ is asymptotically invariant to $\widehat{\theta}_n$. Based on this representation, we can then readily prove the weak convergence of $R^{dpro}_{n,t}(x;\widehat{\theta}_n)$ to a centred Gaussian process with covariance structure given by $\mathbb{K}^{dpro}_t(x,x')$. Lastly, the limiting null distribution of CvM^{dpro}_n can be obtained by standard techniques.

By a straightforward decomposition, we have

$$\begin{aligned} R^{dpro}_{n,t}(x;\widehat{\theta}_n) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\widehat{\theta}_n)\left\{1\left(\beta' X_i \le u\right) - g_t(X_i,\widehat{\theta}_n)'\Delta^{-1}_{n,t}(\widehat{\theta}_n)G_{n,t}(x;\widehat{\theta}_n)\right\} \\ &= R^{pro}_{n,t}(x;\widehat{\theta}_n) - G_{n,t}(x;\widehat{\theta}_n)'\Delta^{-1}_{n,t}(\widehat{\theta}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n) \\ &\equiv R^{pro}_{n,t}(x;\widehat{\theta}_n) - G_{n,t}(x;\widehat{\theta}_n)'\Delta^{-1}_{n,t}(\widehat{\theta}_n)S_{n,t}. \end{aligned}$$

Then, by Lemmas 8.1–8.4,

$$\begin{aligned} &R^{dpro}_{n,t}(x;\widehat{\theta}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)1\left(\beta' X_i \le u\right) - G_t(x;\theta_0)'\sqrt{n}(\widehat{\theta}_n - \theta_0) \\ &\quad - G_t(x;\theta_0)'\Delta^{-1}_t(\theta_0)\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0)\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)\left\{1\left(\beta' X_i \le u\right) - G_t(x;\theta_0)'\Delta^{-1}_t(\theta_0)g_t(X_i,\theta_0)\right\} + o_p(1) \\ &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta' X_i \le u\right) + o_p(1), \end{aligned}$$

uniformly in $x \in \Pi_{pro}$.

The weak convergence of the infeasible double-projected empirical process

$$R^{dpro}_{n0,t}(x;\theta_0) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta' X_i \le u\right), \tag{26}$$

and consequently the weak convergence of $R^{dpro}_{n,t}(x;\widehat{\theta}_n)$ to the centred Gaussian process $R^{dpro}_{\infty,t}$ with covariance structure $\mathbb{K}^{dpro}_t(x,x')$ specified in Theorem 3.1 can be readily obtained by showing that the finite-dimensional distributions of $R^{dpro}_{n0,t}(x;\theta_0)$ converge to those of $R^{dpro}_{\infty,t}$ and the stochastic equicontinuity of $R^{dpro}_{n0,t}(x;\theta_0)$ by a straightforward application of the Donsker property of the class of linear indicator functions

$$\mathcal{F}=\left\{\nu \mapsto 1(\beta'\nu \le u):(\beta,u)\in \Pi_{pro}\right\}.$$

For the convergence in distribution of test statistic CvM_n^{dpro}, we will prove that

$$\sum_{t\in\mathcal{T}} a_n(t)\int_{\Pi_{pro}}\left(R_{n,t}^{dpro}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\,\mathrm{d}\beta \xrightarrow{d} \sum_{t\in\mathcal{T}} a(t)\int_{\Pi_{pro}}\left(R_{\infty,t}^{dpro}(x)\right)^2 F_{\beta}(du)\,\mathrm{d}\beta.$$

Given the assumption that $a_n(t)\xrightarrow{p} a(t)$ for each $t\in\mathcal{T}$, according to Slutsky's theorem, it suffices to show that, for each $t\in\mathcal{T}$,

$$\int_{\Pi_{pro}}\left(R_{n,t}^{dpro}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\,\mathrm{d}\beta \xrightarrow{d} \int_{\Pi_{pro}}\left(R_{\infty,t}^{dpro}(x)\right)^2 F_{\beta}(du)\,\mathrm{d}\beta.$$

First of all, note that the weak convergence of the double-projected process $R_{n,t}^{dpro}(x;\widehat{\theta}_n)$ and the Skorohod construction [see, e.g. Serfling 1980] yield

$$\sup_{(\beta,u)\in\Pi_{pro}}\left|R_{n,t}^{dpro}(x;\widehat{\theta}_n)-R_{\infty,t}^{dpro}(x)\right| \xrightarrow{a.s.} 0. \tag{27}$$

Note that the empirical distribution function $F_{n,\beta}(u)\equiv n^{-1}\sum_{i=1}^{n} 1(\beta'X_i\le u)$ estimates CDF $F_\beta(u):=\mathbb{P}(\beta'X\le u)$ *a.s.* uniformly for $(\beta,u)\in\Pi_{pro}$ by invoking the ULLN of Jennrich (1969) or the generalisation by Wolfowitz (1954) of the Glivenko–Cantelli theorem. That is,

$$\sup_{(\beta,u)\in\Pi_{pro}}\left|F_{n,\beta}(u)-F_\beta(u)\right| \xrightarrow{a.s.} 0. \tag{28}$$

Now write

$$\begin{aligned}
&\left|\int_{\Pi_{pro}}\left(R_{n,t}^{dpro}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\,\mathrm{d}\beta-\int_{\Pi_{pro}}\left(R_{\infty,t}^{dpro}(x)\right)^2 F_{\beta}(du)\,\mathrm{d}\beta\right|\\
&\quad\le\left|\int_{\Pi_{pro}}\left[\left(R_{n,t}^{dpro}(x;\widehat{\theta}_n)\right)^2-\left(R_{\infty,t}^{dpro}(x)\right)^2\right]F_{n,\beta}(du)\,\mathrm{d}\beta\right|\\
&\quad\quad+\left|\int_{\Pi_{pro}}\left(R_{\infty,t}^{dpro}(x)\right)^2\left[F_{n,\beta}(du)-F_\beta(du)\right]\mathrm{d}\beta\right|.
\end{aligned}$$

The first term of the right-hand side of the above inequality is $o(1)$ *a.s.* due to (27). The trajectories of the limiting process $R_{\infty,t}^{dpro}(x)$ are bounded and continuous *a.s.*. Then, by applying the Helly–Bray Theorem (see p. 97 in Rao 1965) to each of these trajectories and taking into account (28), we have that the second term of the right-hand side of the above inequality is also $o(1)$ *a.s.*. This completes the proof of Theorem 3.1. ■

Proof: Under Assumptions 3.1–3.3, uniformly in $x \in \Pi_{pro}$,

$$\begin{aligned}
&\sup_{x\in\Pi_{pro}} \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ e_i(t;\widehat{\theta}_n)\mathcal{P}_{n,t} 1\left(\beta' X_i \le u\right) - \mathbb{E}\left[e(t;\theta^*)\mathcal{P}1\left(\beta' X \le u\right)\right]\right\} \right| \\
&= \sup_{x\in\Pi_{pro}} \left| \frac{1}{\sqrt{n}} R^{dpro}_{n,t}(x;\widehat{\theta}_n) - \mathbb{E}\left[\left(p_t\left(X\right) - q_t\left(X,\theta^*\right)\right)\mathcal{P}1\left(\beta' X \le u\right)\right] \right| \\
&= o_p(1)
\end{aligned}$$

by ULLN of Newey and McFadden (1994, Lemma 2.4) and similar arguments as proving Lemmas 8.1, 8.3, and 8.4. ■

Proof: Note that under the sequence of local alternatives $H_{1,n}$ in (13), we have that, uniformly in $x \in \Pi_{pro}$:

$$\begin{aligned}
&R^{dpro}_{n,t}(x;\widehat{\theta}_n) \\
&\quad= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(e_i(t;\widehat{\theta}_n) - \frac{r_t(X_i)}{\sqrt{n}} \right) \mathcal{P}_{n,t} 1\left(\beta' X_i \le u\right) \\
&\quad\quad + \frac{1}{n} \sum_{i=1}^{n} r_t(X_i)\mathcal{P}_{n,t} 1\left(\beta' X_i \le u\right) \\
&\quad= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(e_i(t;\theta_0) - \frac{r_t(X_i)}{\sqrt{n}} \right) \mathcal{P}_{t} 1\left(\beta' X_i \le u\right) \\
&\quad\quad + \mathbb{E}\left[r_t(X)\mathcal{P}_t 1\left(\beta' X \le u\right)\right] + o_p(1) \\
&\quad\equiv R^{dpro}_{n1,t}(x;\theta_0) + \delta_t(x) + o_p(1) \\
&\quad\Rightarrow R^{dpro}_{\infty,t} + \delta_t,
\end{aligned}$$

where the second equality follows by similar arguments as proving Theorem 3.1 and by the ULLN.

Note that since $e_i(t;\theta_0) - n^{-1/2} r_t(X_i)$ forms a zero mean and an independent and identically distributed summand in the framework of local alternatives, we can establish the weak convergence of $R^{dpro}_{n1,t}(x;\theta_0)$ to a centred Gaussian process by checking the finite-dimensional distributions of $R^{dpro}_{n1,t}(x;\theta_0)$ and its stochastic equicontinuity, just as we have established for $R^{dpro}_{n0,t}(x;\theta_0)$ defined in (26). This yields that

$$R^{dpro}_{n1,t}(x;\theta_0) \Rightarrow R^{dpro}_{\infty,t}.$$

The last step thus follows and we finish the proof of Theorem 3.3. ■

Proof: As in Theorem 3.1, we have the following straightforward decomposition:

$$
\begin{aligned}
&R_{n,t}^{dpro,*}(x;\widehat{\theta}_n)\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)\left(1\left(\beta' X_i \le u\right) - g_t'(X_i,\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)G_{n,t}(x;\widehat{\theta}_n)\right) V_i\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)1\left(\beta' X_i \le u\right) V_i - G_{n,t}'(x;\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n)V_i\\
&\quad \equiv R_{n,t}^{pro,*}(x;\widehat{\theta}_n) - G_{n,t}'(x;\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)A_{n,t}^{*}.
\end{aligned}
$$

Conditional on the original sample $\{(X_i', T_i)'\}_{i=1}^n$, it follows from the stochastic equicontinuity argument and the consistency of $\widehat{\theta}_n$ to θ_0 that, uniformly in $x \in \Pi_{pro}$,

$$
\begin{aligned}
&R_{n,t}^{pro,*}(x;\widehat{\theta}_n)\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1\left(\beta' X_i \le u\right) V_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n) - e_i(t;\theta_0)\right)1\left(\beta' X_i \le u\right) V_i\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1\left(\beta' X_i \le u\right) V_i + o_p(1).
\end{aligned}
$$

Similarly, by the MVT, the consistency of $\widehat{\theta}_n$ to θ_0 and the properties of the sequence of multipliers $\{V_i\}_{i=1}^n$, we can show that

$$
\begin{aligned}
&A_{n,t}^{*}\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)g_t(X_i,\theta_0)V_i\\
&\qquad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\left(g_t(X_i,\widehat{\theta}_n) - g_t(X_i,\theta_0)\right)V_i\\
&\qquad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n) - e_i(t;\theta_0)\right)g_t(X_i,\theta_0)V_i\\
&\qquad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n) - e_i(t;\theta_0)\right)\left(g_t(X_i,\widehat{\theta}_n) - g_t(X_i,\theta_0)\right)V_i\\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)g_t(X_i,\theta_0)V_i + o_p(1).
\end{aligned}
$$

Thus, by Lemmas 8.3 and 8.4, uniformly in $x \in \Pi_{pro}$,

$$\begin{aligned}
&R_{n,t}^{dpro,*}(x;\widehat{\theta}_n) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\left\{1\left(\beta' X_i \le u\right) - G_t(x;\theta_0)'\Delta_t^{-1}(\theta_0)g_t(X_i,\theta_0)\right\}V_i + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta' X_i \le u\right)V_i + o_p(1) \\
&\equiv R_{n0,t}^{dpro,*}(x;\theta_0) + o_p(1),
\end{aligned}$$

leading to the multiplier bootstrapped version of $R_{n0,t}^{dpro}(x;\theta_0)$ defined in (26).

Using the properties of $\{V_i\}_{i=1}^{n}$, the rest of the proof then follows readily from the conditional multiplier central limit theorem applied to the (infeasible) multiplier bootstrapped double-projected process $R_{n0,t}^{dpro,*}(x;\theta_0)$ regardless of whether the null hypothesis H_0 holds or not (see van der Vaart and Wellner 1996, Theorem 2.9.6, p. 182), and the continuous mapping theorem. ■

Proof: We first define an intermediate doubly-projected empirical process

$$\widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n)\left(1(\widetilde{X}_{i,\theta_0}'\beta \le u) - g_t(X_i,\widehat{\theta}_n)'\Delta_{n,t}^{-1}(\widehat{\theta}_n)G_{n,t}^{im}(x;\widehat{\theta}_n)\right).$$

Following similar arguments in the proofs of Lemmas 8.1–8.4, under H_0^{im1} in (15), we can readily show that,

$$\begin{aligned}
(i)\quad &\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n)1(\widetilde{X}_{i,\theta_0}'\beta \le u) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\theta_0)1(\widetilde{X}_{i,\theta_0}'\beta \le u) - \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)' \\
&\quad \times \mathbb{E}\left[g_t(X,\theta_0)1(\widetilde{X}_{\theta_0}'\beta \le u)\right] + o_p(1) \\
&\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\theta_0)1(\widetilde{X}_{i,\theta_0}'\beta \le u) - \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)' G_t^{im}(x;\theta_0) + o_p(1)
\end{aligned}$$

uniformly in x and for each $t \in \mathcal{T}$,

$$\begin{aligned}
(ii)\quad &\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{im}(t;\theta_0)g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) + o_p(1)
\end{aligned}$$

for each $t \in \mathcal{T}$,

$$(iii)\quad G_{n,t}^{im}(x;\widehat{\theta}_n) = G_t^{im}(x;\theta_0) + o_p(1)$$

uniformly in x and for each $t \in \mathcal{T}$, and

$$(iv) \quad \Delta_{n,t}^{-1}(\widehat{\theta}_n) = \Delta_t^{-1}(\theta_0) + o_p(1)$$

for each $t \in \mathcal{T}$. As an immediate consequence of (i)-(iv), uniformly in x,

$$\begin{aligned}
&\widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) 1(\widetilde{X}'_{i,\theta_0}\beta \le u) - G_t^{im}(x;\theta_0)' \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \\
&\qquad - G_t^{im}(x;\theta_0)' \Delta_t^{-1}(\theta_0) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \right) + o_p(1) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) \left(1(\widetilde{X}'_{i,\theta_0}\beta \le u) - g_t(X_i,\theta_0)' \Delta_t^{-1}(\theta_0) G_t^{im}(x;\theta_0) \right) + o_p(1) \\
&\quad \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\theta_0) \mathcal{P}_t^{im} 1(\widetilde{X}'_{i,\theta_0}\beta \le u) + o_p(1).
\end{aligned}$$

In light of the above result, it then suffices to show that $M_{n,t}^{dpro}(x;\widehat{\theta}_n)$ is asymptotically uniformly equivalent to $\widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n)$. To this end, note that uniformly in x,

$$\begin{aligned}
&M_{n,t}^{dpro}(x;\widehat{\theta}_n) - \widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i^{im}(t;\widehat{\theta}_n) \left(1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \le u) - 1(\widetilde{X}'_{i,\theta_0}\beta \le u) \right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(e_i^{im}(t;\theta_0) - \left(q_t(\widetilde{X}_{i,\widehat{\theta}_n}) - q_t(\widetilde{X}_{i,\theta_0}) \right) \right) \left(1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \le u) - 1(\widetilde{X}'_{i,\theta_0}\beta \le u) \right) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i(t) \left(1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \le u) - 1(\widetilde{X}'_{i,\theta_0}\beta \le u) \right) \\
&\qquad - \sqrt{n}(\widehat{\theta}_n - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} g_t(X_i,\theta_0) \left(1(\widetilde{X}'_{i,\widehat{\theta}_n}\beta \le u) - 1(\widetilde{X}'_{i,\theta_0}\beta \le u) \right) + o_p(1) \\
&\quad \equiv M_{n,t1}^{dpro}(x) - \sqrt{n}(\widehat{\theta}_n - \theta_0)' M_{n,t2}^{dpro}(x) + o_p(1),
\end{aligned}$$

where $e_i^{im}(t;\theta_0) = \varepsilon_i(t)$ *a.s.* under H_0^{im1} in (15), and the second to last equality follows by the Taylor expansion of $q_t(\widetilde{X}_{i,\widehat{\theta}_n})$ around $q_t(\widetilde{X}_{i,\theta_0})$ and Assumptions 3.2 and 3.3. Since $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ by Assumption 3.3(ii), it then remains to show that both $M_{n,t1}^{dpro}(x)$ and $M_{n,t2}^{dpro}(x)$ are asymptotically uniformly negligible in x.

For the first term $M^{dpro}_{n,t1}(x)$, introduce an auxiliary process

$$\alpha_{n,t}(x;\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i(t)1(\widetilde{X}'_{i,\theta}\beta \leq u).$$

Since under H_0^{im1} for each $t \in \mathcal{T}$ the sequence of error terms $\{\varepsilon_1(t),\ldots,\varepsilon_n(t)\}$ is centred conditionally on $\{X_1,\ldots,X_n\}$, $\alpha_{n,t}(x;\theta)$ has i.i.d. centred summands. Then $M^{dpro}_{n,t1}(x)$ can be expressed as $\alpha_{n,t}(x;\widehat{\theta}_n) - \alpha_{n,t}(x;\theta_0)$. Clearly, the class of linear indicator functions is Donsker and $\alpha_{n,t}(\cdot,\cdot)$ is asymptotically equicontinuous. Since $\widehat{\theta}_n \to_p \theta_0$ under H_0^{im1} by Assumption 3.3(ii), $M^{dpro}_{n,t1}(x) = o_p(1)$ uniformly in x. The claim that $M^{dpro}_{n,t2}(x) = o_p(1)$ uniformly in x follows straightforwardly from the uniform convergence of

$$\frac{1}{n}\sum_{i=1}^{n} g_t(X_i,\theta_0)1(\widetilde{X}'_{i,\theta}\beta \leq u)$$

in x and θ together with the continuity of its limit.

The rest of the proof is similar to that of Theorem 3.1 and is thus omitted to avoid repetition. ■

Acknowledgments

We thank the Guest Editor and the referees for useful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Escanciano gratefully acknowledges financial support by MICIN/AEI/10.13039/501100011033, grant CEX2021-001181-M, Comunidad de Madrid, grants EPUC3M11 (V PRICIT) and H2019/HUM-5891, and grant PID2021-127794NB-I00 (MCI/AEI/FEDER, UE). Song gratefully acknowledges financial support from the National Natural Science Foundation of China (Grant Numbers 72373007, 72333001).

References

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(1), 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.
Delgado, M.A., Domínguez, M.A., and Lavergne, P. (2006), 'Consistent Tests of Conditional Moment Restrictions', *Annales d'Économie et de Statistique*, 81, 33–67.
Dominguez, M.A., and Lobato, I.N. (2004), 'Consistent Estimation of Models Defined by Conditional Moment Restrictions', *Econometrica*, 72(5), 1601–1615.

Escanciano, J.C. (2006a), 'A Consistent Diagnostic Test for Regression Models Using Projections', *Econometric Theory*, 22, 1030–1051.
Escanciano, J.C. (2006b), 'Goodness-of-Fit Tests for Linear and Nonlinear Time Series Models', *Journal of the American Statistical Association*, 101(474), 531–541.
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. (2009a), 'On the Lack of Power of Omnibus Specification Tests', *Econometric Theory*, 25(1), 162–194.
Escanciano, J.C. (2009b), 'Simple Bootstrap Tests for Conditional Moment Restrictions', Technical Report, Indiana University.
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 and Data Analysis*, 54(3), 625–636.
García-Portugués, E., Álvarez Liébana, J., Álvarez Pérez, G., and González-Manteiga, W. (2021), 'A Goodness-of-fit Test for the Functional Linear Model with Functional Response', *Scandinavian Journal of Statistics*, 48(2), 502–528. https://doi.org/10.1111/sjos.12486.
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.
Hájek, J. (1971), 'Discussion of 'An Essay on the Logical Foundations of Survey Sampling, Part I', by D. Basu,' in *Foundations of Statistical Inference*, eds. V. P. Godambe, and D. A. Sprott, Toronto: Holt, Rinehart, and Winston.
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.
Jennrich, R.I. (1969), 'Asymptotic Properties of Non-linear Least Squares Estimators', *The Annals of Mathematical Statistics*, 40(2), 633–643.
Kim, I., Balakrishnan, S., and Wasserman, L. (2020), 'Robust Multivariate Nonparametric Tests Via Projection-Averaging', *Annals of Statistics*, 48(6), 3417–3441.
Knight, K., and Fu, W. (2000), 'Asymptotics for Lasso-type Estimators', *Annals of Statistics*, 28(5), 1356–1378.
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', *Annals of Statistics*, 21(1), 255–285.
Newey, W.K., and McFadden, D. (1994), 'Large Sample Estimation and Hypothesis Testing,' in Handbook of Econometrics, Vol. 4, Amsterdam: North-Holland: Elsevier, chapter 36, pp. 2111–2245.
Neyman, J. (1959), 'Optimal Asymptotic Tests of Composite Statistical Hypotheses,' in Probability and Statistics: The Harald Cramer Volume, ed. U. Grenander, Stockholm: Almqvist & Wiksell, pp. 213–234.
Rao, C.R. (1965), *Linear Statistical Inference and Its Applications*, New York: Wiley.

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.
Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.
Strasser, H. (1990), 'Global Extrapolation of Local Efficiency', *Statistics & Risk Modeling*, 8(1), 11–26.
Stute, W. (1997), 'Nonparametric Model Checks for Regression', *Annals of Statistics*, 25(2), 613–641.
Stute, W., and Zhu, L.-X. (2005), 'Nonparametric Checks for Single-index Models', *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.
Wolfowitz, J. (1954), 'Generalization of the Theorem of Glivenko-Cantelli', *Annals of Statistics*, 25(1), 131–138.
Yang, S., Imbens, G.W., Cui, Z., Faries, D.E., and Kadziola, Z. (2016), 'Propensity Score Matching and Subclassification in Observational Studies with Multi-level Treatments', *Biometrics*, 72(4), 1055–1065.
Zhu, L., Xu, K., Li, R., and Zhong, W. (2017), 'Projection Correlation between Two Random Vectors', *Biometrika*, 104(4), 829–843.