Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 07-005

"Asymptotically Optimal Tests for Single-Index Restrictions with a Focus on Average Partial Effects"

by

Juan Carlos Escanciano and Kyungchul Song

http://ssrn.com/abstract=960530

Asymptotically Optimal Tests for Single-Index Restrictions with a Focus on Average Partial Effects[1]

Juan Carlos Escanciano[2] and Kyungchul Song[3]

January 29, 2007.

Abstract: This paper proposes an asymptotically optimal specification test of single-index models against alternatives that lead to inconsistent estimates of a covariate's average partial effect. The proposed tests are relevant when a researcher is concerned about a potential violation of the single-index restriction only to the extent that the estimated average partial effects suffer from a nontrivial bias due to the misspecification. Using a pseudo-norm of average partial effects deviation and drawing on the minimax approach, we find a nice characterization of the least favorable local alternatives associated with misspecified average partial effects as a single direction of Pitman local alternatives. Based on this characterization, we define an asymptotic optimal test to be a semiparametrically efficient test that tests the significance of the least favorable direction in an augmented regression formulation, and propose such a one that is asymptotically distribution-free, with asymptotic critical values available from the χ_1^2 table. The testing procedure can be easily modified when one wants to consider average partial effects with respect to binary covariates or multivariate average partial effects.

Key words and phrases: Average Partial Effects, Omnibus tests, Optimal tests, Semiparametric Efficiency, Efficient Score.

JEL Classification Number: C14

[1]We are grateful to Javier Hualde, Ignacio Lobato, Manuel Domínguez, Wenceslao González-Manteiga, Quang Vuong, and seminar participants at Penn State University for useful comments. All errors belong to us. Corresponding Address: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6297. Email: kysong@sas.upenn.edu.

[2]Department of Economics, Indiana University. Parts of this paper were written while I was visiting the Cowles Foundation at Yale University and the Economics Department at Cornell University, whose hospitality is gratefully acknowledged. Research funded by the Spanish Ministerio de Educación y Ciencia, reference numbers SEJ2004-04583/ECON and SEJ2005-07657/ECON.

[3]Department of Economics, University of Pennsylvania

1 Introduction

Suppose that a researcher is interested in testing a conditional moment restriction

$$\mathbf{E}\left[\rho(S;\beta)|X\right] = 0 \text{ for some } \beta \in \mathcal{B} \tag{1}$$

where S and X represent random vectors and $\rho(s;\beta)$ is a function of s indexed by $\beta \in \mathcal{B}$ with $\mathcal{B}$ denoting a finite or infinite dimensional parameter space. A typical power analysis of a test involves studying the asymptotic power against alternatives of the form:

$$\mathbf{E}_n\left[\rho(S;\beta)|X\right] = a_n(X) \text{ for some } \beta \in \mathcal{B}$$

for a sequence a_n of functions, where $\mathbf{E}_n$ denotes the expectation under the local alternatives. An *omnibus* test is a test designed to have nontrivial power against essentially all the local alternatives that represent the negation of the null in (1) and converge to the null hypothesis at a rate not too fast. In particular, when $a_n(x) = b_n a(x)$ for a fixed function a and a decreasing sequence $b_n \to 0$, the alternatives are often called *Pitman local alternatives* (e.g. Nikitin (1995)) and the function a is referred to as the *direction* of the alternatives.

Although considering an omnibus test is naturally the first idea when there is no *a priori* preference of alternatives that receive more attention than others, it is worth noting that there are several known limitations of omnibus tests. Most notably, Janssen (2000) has shown that every omnibus test of goodness-of-fit has a power envelope function that is almost flat except on a finite dimensional space of alternatives. The few directions that span this finite dimensional space often lack motivation in practice, and change dramatically, corresponding to an apparently innocuous change of the test statistic. This finding leads him to remark as follows:

> A well-reflected choice of tests requires some knowledge of preferences concerning alternatives which may come from the practical experiment. (Janssen (2000), p.240)

It appears that the idea of incorporating an *a priori* interest in a subset of alternatives into a test of nonparametric or semiparametric models has not received much attention in the literature. The literature on testing nonparametric or semiparametric restrictions is dominantly concerned with the omnibus approach, and a few studies in the literature of nonparametric specification tests that deal with a single direction or several directions of Pitman local alternatives (e.g. Stute (1997)) often lack practical motivation for the specific choice of such directions.

This paper studies a concrete example of a semiparametric test with a focus on a subset of alternatives that is specifically motivated by the interest of the model's user. Suppose that a researcher is interested in testing the single-index restriction:

$$\mathbf{E}\left[Y - \mu(X'\theta)|X\right] = 0,$$

where θ is a finite dimensional parameter and μ is an unknown function, but does not worry about the violation of the restriction as long as the identification of the average partial effect of a covariate of interest remains intact. This particular interest in a subset of alternatives seems natural when one's use of the single-index restriction is motivated by its facility in identifying average partial effects. This constitutes an interesting situation that marks departure from both the omnibus approach and the directional approach that exist in the literature. In this situation, an omnibus test may not be an optimal solution because the test will waste its power on alternatives that are of no interest to the econometrician. The situation is also distinguished from that of the directional approach in the literature because the set of alternatives of focus here have a clear, practical motivation and are constituted by an infinite number of directions, not just several of them.

This paper introduces a new notion of optimality of a test in a situation where a particular interest in a subset of alternatives leads one to exclude those alternatives that satisfy a certain linear functional equation. Here in the context of testing single-index restriction, the equation corresponds to the equality between restricted and unrestricted average partial effects. This new notion of optimality is constituted by two steps. The first step involves excluding the set of uninteresting alternatives that satisfy the linear functional equation by using an appropriate pseudo-norm which is based on the linear functional. Then, drawing on the minimax approach (e.g. Ingster and Suslina (2003)), we select from the remaining alternatives those that are least favorable, in other words, we select those that are as closest as possible to the null hypothesis (here, of single-index restriction) in terms of the L_2-norm. The comparison of tests then can be made based on their asymptotic power properties against this set of least favorable alternatives. In this context of testing a conditional moment restriction with uninteresting alternatives identified by a linear functional equation, we find that the selected least favorable alternatives are characterized as Pitman local alternatives with a single direction given by the Riesz Representation of the linear functional. Then, in the second step, following Choi, Hall, and Schick (1996), we define an optimal test to be a test that achieves a semiparametric power envelope which is a hypothesis testing analogue of the semiparametric efficiency bound in estimation theory. More specifically, this optimal test is an asymptotically uniformly most powerful test that is derived from the local asymptotic normality (LAN) of the semiparametric model where experiments of local shifts encompass all the parametric submodels that pass through the probabilities under the null hypothesis.

As mentioned before, the investigation is expedited by our finding that the set of least favorable directions in L_2 distance after the exclusion of the uninteresting alternatives is characterized by a single direction of Pitman local alternatives. This finding reveals that in the conditional moment tests, the elimination of alternatives that satisfy a linear functional equation renders the problem of minimax rate optimality trivial with the parametric optimal rate $n^{-1/2}$. (For minimax rate optimality, see Horowitz and Spokoiny (2001) and Guerre and Lavergne (2002) and references therein.)

We construct an asymptotic optimal test that is based on the series estimation. In order to deal with the asymptotic properties of the test, we establish a general result of uniform asymptotic representation of empirical processes that involve a series-based conditional mean estimator (see Lemma 1U in the appendix.) Here are the findings from the asymptotic theory. First, the estimation of θ_0 is *ancillary* to the asymptotic optimality of the test. In other words, lack of knowledge of θ_0 does not affect the semiparametric power envelope. Second, the direction of Pitman local alternatives that give the maximal local asymptotic power lies in the set of interesting alternatives that give a misspecified average partial effect. Note that this is not necessarily ensured by usual omnibus tests that disregard the particular focus on the interesting alternatives. Third, the space of local alternatives against which the optimal test has nontrivial local asymptotic power does not in general coincide with the space of interesting alternatives. This is due to the fact that the direction against which the test has no local asymptotic power due to the elimination of uninteresting alternatives is "tilted" by the optimal incorporation of the information in the null hypothesis of single-index restriction. This demonstrates that the notion of optimality crucially depends on the formulation of the null hypothesis and the information it contains.

There have been a plethora of researches investigating inference in single-index models. Duan and Li (1989), Powell, Stock and Stoker (1989), Härdle, Hall and Ichimura (1993), Ichimura (1993), Klein and Spady (1993) or Hristache, Juditsky and Spokoiny (2001), among others, studied the estimation problem. Newey and Stoker (1993) proposed an efficient estimation of an average partial effect in a more general setting. See also Delecroix, Härdle and Hristache (2003). As compared to estimation, the problem of testing single-index restrictions has received less attention in the literature. Fan and Li (1996) and Aït-Sahalia, Bickel and Stoker (2001) proposed omnibus tests based on a (weighted) residual sum of squares in the spirit of Härdle and Mammen (1993). Recently, Stute and Zhu (2005) and Xia, Li, Tong and Zhang (2004) proposed bootstrap-based omnibus tests. Our paper deviates from this omnibus approach, as it acknowledges priority of correct identification of average partial effects in the specification test.

The rest of the paper is organized as follows. In Section 2 we define the basic environment of hypothesis testing which is of focus in this paper. Section 3 introduces the notion of asymptotic optimality of the tests and presents asymptotically optimal tests. Section 4 is devoted to the asymptotic theory of the proposed test. Section 5 discusses extensions including the cases of a binary covariate and of multivariate average partial effects. In Section 6, we conclude. Besides the mathematical proofs of the results, the appendix also contains a brief review of semiparametric efficient tests and a general uniform asymptotic representation of a semiparametric marked empirical process that is of independent interest.

2 Testing Framework

2.1 Single-Index Restrictions: The Null Hypothesis

Let a random vector $Z = (Y, X)$ in $\mathbf{R}^{1+d_X}$ follow a distribution P_0, where Y and X are related by

$$Y = m_0(X) + \varepsilon,$$

for some real-valued function $m_0(\cdot)$, and a random variable ε such that $\mathbf{E}[\varepsilon|X] = 0$ almost surely (a.s.). Here $\mathbf{E}[Y|X]$ indicates the conditional expectation of Y given X under the probability measure P_0. Throughout the paper, we assume that $\mathbf{E}[||X||] < \infty$ and $\mathbf{E}[\varepsilon^2] < \infty$. The function $m_0(\cdot)$ is identified as the conditional mean function $\mathbf{E}[Y|X = \cdot]$.

As opposed to the notation P_0 which denotes the true data generating process behind (Y, X), we use the notation P as a generic probability that serves as a potential distribution of (Y, X) and has a well-defined conditional expectation $m_P(X) \triangleq \mathbf{E}_P[Y|X]$. (We use the notation $\triangleq$ for a definitional relation.) Let $L_2(P)$ be the space of square integrable random variables with respect to P and let $||\cdot||_{2,P}$ and $||\cdot||_2$ indicate the $L_2(P)$ norm and the $L_2(P_0)$ norm respectively. Finally, the notation $||\cdot||$ denotes the Euclidean norm defined as $||a|| \triangleq \sqrt{tr(a'a)}$, for $a \in \mathbf{R}^d$, and the notation $||\cdot||_\infty$, the sup norm : $||f||_\infty \triangleq \sup_v |f(v)|$.

The null hypothesis of a single index restriction is written as

$$H_0 : m_0(X) = \mu(X'\theta) \text{ a.s. for some } \theta \in \Theta \subset \mathbf{R}^d \text{ and some } \mu(\cdot) \in \mathcal{M},$$

where the parameter θ is a vector in a compact subset of the Euclidean space, $\Theta \subset \mathbf{R}^d$, and $\mathcal{M}$ is a space of measurable functions on $\mathbf{R}$. Let us denote the vector of parameters $\beta \triangleq (\theta, \mu)$, and for the space of parameters, we introduce the notation $\mathcal{B} = \Theta \times \mathcal{M}$. Then, the class of probabilities under H_0 is

$$\mathcal{P}_0 = \{P \in \mathcal{P} : \exists(\theta, \mu) \in \mathcal{B} \text{ s.t. } P\left\{m_P(X) = \mu(X'\theta)\right\} = 1\}.$$

The alternatives are probabilities in $\mathcal{P}_1 = \mathcal{P}\backslash\mathcal{P}_0$.

The null hypothesis of a single index restriction may constitute identifying restrictions for parameters θ and μ that may cease to hold under the alternatives. In this paper, we confine our attention to the probability model $\mathcal{P}$ such that under each potential data generating process $P \in \mathcal{P}$, a single parameter $\theta_P \in \Theta$ is identified and has a $\sqrt{n}$-consistent estimator $\hat{\theta}$. More specifically, we assume that there is a unique solution:

$$\theta_P \triangleq \arg\min_{\theta \in \Theta} \mathbf{E}_P[\left(Y - \mathbf{E}_P[Y|X'\theta]\right)^2], \tag{2}$$

for each $P \in \mathcal{P}$. This identified parameter θ_P may change as we move from one data generating process to another within $\mathcal{P}$ and hence its dependence upon P is made explicit by its subscript. Once θ_P is identified, the function $\mu_P(\cdot)$ is identified as

$$\mu_P(v) = \mathbf{E}_P\left[Y|X'\theta_P = v\right]. \tag{3}$$

We simply write $\mu_0 \triangleq \mu_{P_0}$ and $\theta_0 \triangleq \theta_{P_0}$. For simplicity, we assume that $\mathcal{P}$ is chosen such that for all $P \in \mathcal{P}$, identified parameters $\beta_P \triangleq (\theta_P, \mu_P)$ belong to $\mathcal{B} = \Theta \times \mathcal{M}$.

2.2 Average Partial Effects: The Alternative Hypothesis

An omnibus test focuses on the whole space of alternatives $\mathcal{P}_1$. On the contrary, in this paper we consider a situation where a researcher's main interest lies in the estimation of average partial effects. For each $P \in \mathcal{P}$, the (nonparametric) average partial effect with respect to X_1 is defined by[4]

$$\mathbf{E}_P \left[\frac{\partial}{\partial x_1} \mathbf{E}_P[Y|X = x]|_{x=X} \right],$$

provided the regression function $m_P(x) = \mathbf{E}_P[Y|X = x]$ is differentiable[5]. For each $P \in \mathcal{P}_0$, the average partial effect is equal to

$$\mathbf{E}_P \left[\frac{\partial}{\partial x_1} \mu_P(x'\theta_P)|_{x=X} \right].$$

We aim to design a test that detects only those alternatives that are associated with divergence between the restricted (i.e. model-based) and unrestricted (i.e. nonparametric) average partial effects. Therefore, *uninteresting* alternatives in this situation are those that satisfy the following equality

$$\int \frac{\partial m_P(x)}{\partial x_1} P(dx) = \int (\frac{\partial \mu_P}{\partial x_1})(x'\theta_P) P(dx). \tag{4}$$

For each $P \in \mathcal{P}$, we define a linear functional M_P on $\mathcal{D}_P \triangleq \{g \in L_2(P) : \mathbf{E}_P[| (\partial g/\partial x_1)(X)|] < \infty\}$ by

$$M_P g \triangleq \int \left(\frac{\partial g}{\partial x_1} \right)(x) P(dx), \tag{5}$$

and define

$$e(Z; \beta) \triangleq Y - \mu(X'\theta) \text{ and } r_P(x) \triangleq \mathbf{E}_P[e(Z; \beta_P)|X = x].$$

Then, noting $r_P(x) = m_P(x) - \mu_P(x'\theta_P)$ and assuming that $r_P \in \mathcal{D}_P$, one can see from (4) that the alternatives that lead to a correct estimation of average partial effects using the single-index restrictions are those P's such that $M_P r_P = 0$, whereas alternatives that lead to bias in the estimation of average partial effects are the ones with $M_P r_P \neq 0$. We define a subset $\mathcal{P}_M$ of *uninteresting* alternatives in $\mathcal{P}_1$ by

$$\mathcal{P}_M \triangleq \{P \in \mathcal{P}_1 : M_P r_P = 0\}. \tag{6}$$

The subset $\mathcal{P}_M$ of alternatives is uninteresting in the sense that the violation of the null hypothesis due to $P_0 \in \mathcal{P}_M$ does not cause bias to the average partial effects identified under the null hypothesis. The space of all the alternatives $\mathcal{P}_1$ is decomposed into $\mathcal{P}_M \cup \mathcal{P}_M^c$ where $\mathcal{P}_M^c = \mathcal{P}_1 \backslash \mathcal{P}_M$. The

[4] In the article we mainly focus on the average partial effect with respect to a single continuous covariate X_1. For multivariate extensions of our approach or extensions to a binary covariate, see Section 5.

[5] All the results of the paper can be applied similarly to non-differentiable regressions $m_0(x)$ by changing the operator M in (5) accordingly.

alternative hypothesis is then written as

$$H_1 : P_0 \in \mathcal{P}_M^c. \tag{7}$$

This article develops a test that optimally concentrates its local asymptotic power on the subset $\mathcal{P}_M^c$ of alternatives.

We stress that the null hypothesis in this paper is not whether the equivalence of the restricted and unrestricted average partial effects holds (i.e. $P_0 \in \mathcal{P}_M$) but whether the single index restriction holds (i.e., $P_0 \in \mathcal{P}_0$). We may formulate a test that tests whether the restricted average partial effects and the unrestricted partial effects are the same. In the case when this test is omnibus, the alternative hypothesis is precisely the same as in (7). However, when one attempts to construct an asymptotically optimal test in the sense that is adopted in this paper, it makes a difference in general how the null hypothesis (and of course, the alternative hypothesis) is formulated. For details, see the discussion after Theorem 1 in Section 4.

3 Asymptotic Optimality of Semiparametric Tests

3.1 Characterizing the Alternatives of Focus

In this subsection, we provide a useful characterization of interesting alternatives. Recall that for each $P \in \mathcal{P}$, we let $m_P(x) = \mathbf{E}_P[Y|X = x]$. The hypothesis testing problem of single-index restriction is written as

$$\begin{aligned} H_0 &: m_0 \in \mathcal{G}_{P_0} \text{ and} \\ \widetilde{H}_1 &: m_0 \notin \mathcal{G}_{P_0} \end{aligned} \tag{8}$$

where $\mathcal{G}_P \triangleq \{m \in \mathcal{D}_P : m(x) = \mu(x'\theta),\ (\theta, \mu) \in \Theta \times \mathcal{M}\}$.

A minimax approach compares tests based on the local power at the least favorable alternatives that give the maximum of Type II error over the space of alternatives. Since the least favorable alternatives can be arbitrarily close to the null in the setup of (8), giving a trivial maximum Type II error equal to one minus Type I error, it is often suggested to consider alternatives $P \in \mathcal{P}_1$ such that $H_1(r_n) : \inf_{m \in \mathcal{G}_P} ||m_P - m|| > r_n$, where $r_n \to 0$ and $|| \cdot ||$ is a norm on $\mathcal{G}_P$. Clearly, the notion of asymptotic optimality of nonparametric/semiparametric tests critically depends on the norm $|| \cdot ||$ taken for the space $\mathcal{G}_P$. For example, one may consider an $L_2(P)$-norm and proceed with the notion of rate-optimality. (e.g. Ingster (1993), Horowitz and Spokoiny (2001), Guerre and Lavergne (2002), Ingster and Suslina (2003)). Or using the Kolmogorov-Smirnov type functional or Cramér-von Mises type functional on $\mathcal{G}_P$, one can consider asymptotic minimax tests (e.g. Ermakov (1995)).

In view of our specific interest in average partial effects, it is natural to equip the space $\mathcal{G}_P$ with the following pseudo-norm:

$$||m||_{M_P} \triangleq |M_P m| = \left| \int \frac{\partial m}{\partial x_1}(x) P(dx) \right|.$$

The distance between models is measured in terms of the deviation of their average partial effects. For each probability $P \in \mathcal{P}$, let us define

$$\mathcal{G}(P; r_n) \triangleq \{m \in \mathcal{G}_P : ||m_P - m||_{M_P} \geq r_n\}$$

for a decreasing sequence $r_n \to 0$. The space $\mathcal{G}(P; r_n)$ represents a collection of maps $m(x) = \mu(x'\theta)$ that are deviated from the conditional mean function $m_P(x)$ of Y given $X = x$ with respect to P at least by r_n. For each $P \in \mathcal{P}_1$, there are many m's with the same "distance" from m_P with respect to $|| \cdot ||_{M_P}$. Of primary interest among those m's would be those that are closest to m_P in the $L_2(P)$ norm, $|| \cdot ||_{2,P}$. Define the space of local alternatives:

$$\mathcal{P}(r_n) \triangleq \arg\min_{P \in \mathcal{P}} \inf_{m \in \mathcal{G}(P; r_n)} ||m_P - m||_{2,P}. \tag{9}$$

The space $\mathcal{P}(r_n)$ collects probabilities that attain the minimal value of $\inf_{m \in \mathcal{G}(P; r_n)} ||m_P - m||_{2,P}$ and hence are hardest to distinguish from the null among those that have the same average partial effects deviation in terms of $|| \cdot ||_{M_P}$. Our notion of optimality centers on the comparison of the local asymptotic power properties of tests at the alternatives in $\mathcal{P}(r_n)$.

We introduce a lemma that characterizes the space $\mathcal{P}(r_n)$ as a sequence of Pitman local alternatives. First, observe that under regularity conditions, the operator M_P defined in (5) is a bounded linear functional, and hence, the Riesz Representation Theorem tells us that there exists a unique $b_P \in \mathcal{D}_P$ such that

$$M_P a = \langle a, b_P \rangle \triangleq \mathbf{E}_P[a(X) b_P(X)], \tag{10}$$

for all $a \in \mathcal{D}_P$. As a matter of fact, it can be shown that the b_P satisfying (10) is given by

$$b_P(\cdot) = -\frac{1}{f_X(\cdot)} \frac{\partial f_X}{\partial x_1}(\cdot) \tag{11}$$

where f_X is the density function of X with respect to the Lebesgue measure. Hence, $P \in \mathcal{P}_M^c$ if and only if [6]

$$\mathbf{E}_P[e(Z; \beta_P) b_P(X)] \neq 0,$$

which corresponds to the misspecification of the average partial effects. In other words, the subset of *uninteresting* alternatives in (6) is represented by

$$\mathcal{P}_M = \{P \in \mathcal{P}_1 : \mathbf{E}_P[e(Z; \beta_P) b_P(X)] = 0\}. \tag{12}$$

[6] In fact, for this equivalence to hold it is not necessary to assume that $b_P \in \mathcal{D}_P$. By integration by parts it is enough to assume that r_P is continuously differentiable, with $\mathbf{E}_P[|\partial r_P / \partial x_1(X)|] < \infty$ and $\mathbf{E}_P[|r_P(X) b_P(X)|] < \infty$.

The following lemma shows that $\mathcal{P}(r_n)$ is characterized as Pitman local alternatives with direction b_P.

Lemma 1 : *$\mathcal{P}(r_n) = \{P \in \mathcal{P} : ||m_P - (\tilde{m}_P + c_n b_P)||_{2,P} = 0\}$, where $\tilde{m}_P(x) \triangleq \mu_P(x'\theta_P)$, (θ_P, μ_P) is as defined in* (2) *and* (3)*, $b_P(x)$ is as defined in* (11)*, and c_n is equal to either $r_n/\mathbf{E}_P(b_P^2(X))$ or $-r_n/\mathbf{E}_P(b_P^2(X))$.*

According to this lemma, as long as we confine our attention to $\mathcal{P}(r_n)$ as the space of alternatives, it suffices for us to consider Pitman local alternatives of a single direction $b_P(x)$. This result has two important consequences. First, the fastest possible rate r_n that gives a test a nontrivial power uniformly over $\mathcal{P}(r_n)$ is $n^{-1/2}$. Hence the rate-optimality property is trivially satisfied with $r_n = n^{-1/2}$ when we restrict the space of alternatives to $\mathcal{P}(r_n)$. Second, this enables us to resort to the notion of asymptotic optimality of tests via the semiparametric power envelope criteria (Choi, Hall, and Schick (1996)). In the next subsection, we formally define the notion of asymptotic optimality, and introduce related terminologies.

3.2 Definition of Asymptotic Optimality

By the result of Lemma 1, we confine our attention to the following space of probabilities:

$$\mathcal{P}^* \triangleq \left\{P \in \mathcal{P} : m_P(x) = \mu_P(x'\theta_P) + cb_P(x),\ P\text{-a.s., } c \in \mathbf{R}\right\} \tag{13}$$

The restriction of probabilities to $\mathcal{P}^*$ is tantamount to considering the following regression model:

$$Y = \mu_0(X'\theta_0) + cb_0(X) + \varepsilon \tag{14}$$

where ε is a random variable satisfying $\mathbf{E}[\varepsilon|X] = 0$. Then the null hypothesis and the alternatives are written as the following univariate two-sided test:

$$H_0^* : c = 0 \text{ against } H_1^* : c \neq 0. \tag{15}$$

The parameter of interest is c and the nuisance parameters in the model are given by $\eta_0 = (\theta_0', \mu_0(\cdot), f_{\varepsilon|X}(\cdot), f_X(\cdot))'$, where $f_{\varepsilon|X}(\cdot)$ is the conditional density of ε given X and $f_X(\cdot)$ denotes the density of X. We follow Choi, Hall, and Schick (1996) to define asymptotic optimality of tests in this environment.[7]

Let $\gamma_0 = (0, \eta_0)$ and $\gamma = (c, \eta)$ with $\eta = (\theta', \mu(\cdot), h_{\varepsilon|X}(\cdot), h_X)' \in \mathcal{H} \triangleq \mathcal{B} \times \mathcal{F}_{\varepsilon|X} \times \mathcal{F}_X$. Here $\mathcal{F}_{\varepsilon|X}$ is the set of all the potential conditional densities $h_{\varepsilon|X}(\cdot)$ of ε given X such that $\int \varepsilon h_{\varepsilon|X}(\varepsilon) d\varepsilon = 0$, a.s., and $\mathcal{F}_X$ is the set of all the potential densities of X. Then we can parametrize $\mathcal{P}^* = \{P_\gamma : \gamma \in \Gamma\}$ where $\Gamma \triangleq \mathbf{R} \times \mathcal{H}$.

[7] Note that although b_0 is not known, we do not have to include it in our nuisance parameters. This is because the parameter of interest is $c = 0$ and there is no role left for b_0 in constructing a tangent space onto which the score of c is projected. See the Appendix.

We consider the local deviation of $\gamma_n(h)$ in the direction h from $\gamma_0 = (0, \eta_0)$:

$$
\begin{aligned}
c(h_c) &\triangleq n^{-1/2} h_c + o(n^{-1/2}) \text{ and} \\
\eta_n(h_\eta) &\triangleq \eta_0 + n^{-1/2} h_\eta + o(n^{-1/2}).
\end{aligned} \tag{16}
$$

Note that $h = (h_c, h_\eta)$ denotes the direction in which the local parameter $\gamma_n(h) \triangleq (c_n(h_c), \eta_n(h_\eta))$ deviates from the point $(0, \eta_0)$. Fix the direction $h_1 = (h_{1c}, h_{1\eta})$ and consider testing the simple hypothesis $\gamma_{0\eta} = \gamma_n(h_0)$ with $h_0 = (0, h_{0\eta})$ against $\gamma_{1\eta} = \gamma_n(h_1)$ with $h_1 = (h_{1c}, h_{1\eta})$. When we take $h_{0\eta}$ to be finite-dimensional, the model under $\gamma = \gamma_n(h_0)$ represents a parametric submodel passing through γ_0 under the null hypothesis.

A sequence of tests ψ_n that are equal to one if and only if the null is rejected is called *asymptotically unbiased* if $\limsup_n \mathbf{E}_{\gamma_n(h_0)} \psi_n \leq \liminf_n \mathbf{E}_{\gamma_n(h_1)} \psi_n$ for every $h_0 = (0, h_{\eta_0})$ and $h_1 = (h_{1c}, h_{1\eta})$, $h_{1c} \neq 0$. A test ψ_n is said to be of *asymptotic level* α at η_0 if

$$\text{limsup}_n \mathbf{E}_{\gamma_n(h_0)} \psi_n \leq \alpha \text{ for every } h_{0\eta}.$$

The restriction of candidate tests to those of asymptotic level α plays the same role as considering only regular estimators in the definition of semiparametric efficiency in estimation. A test ψ_n is called *asymptotically uniformly most powerful and asymptotically unbiased* at η_0 among asymptotically unbiased tests (AUMPU(α, η_0)) if it is asymptotically unbiased at η_0 and is of asymptotic level α at η_0 and if for every other such test ψ_n' and each $\gamma_n(h)$ with $h_c \neq 0$,

$$\liminf_n \mathbf{E}_{\gamma_n(h)} \psi_n \geq \limsup_n \mathbf{E}_{\gamma_n(h)} \psi_n'.$$

A *semiparametric power envelope* for tests of asymptotic level α at η_0 is a function of local directions h defined to be $\liminf_n \mathbf{E}_{\gamma_n(h)} \psi_n$ where ψ_n is an AUMPU(α, η) test. When the optimal test does not depend on η_0, the test is asymptotically uniformly most powerful among asymptotically unbiased tests that are asymptotic level of α (AUMPU(α)). We discuss the construction of a test that is AUMPU(α) in the next subsection.

3.3 Construction of Asymptotically Optimal Tests

An asymptotically uniformly most powerful test can be characterized as a test that achieves a semiparametric power envelope. Given the semiparametric model $\mathcal{P}^*$ in the preceding section, the definition of a semiparametric power envelope parallels that of semiparametric efficiency bound in estimation. We first find an asymptotic power envelope for the tests of an asymptotic level α by focusing on the parametric submodels with local deviation $\gamma_{1\eta} = \gamma_n(h_1)$ that passes through $\gamma_{0\eta} = \gamma_n(h_0)$ with directions h_0 and h_1 fixed. Then, from the local asymptotic normality (LAN) of the likelihood ratio, we find that the upper bound for the local asymptotic power is increasing in the $L_2(P)$ distance between the two directions h_1 and h_0. We obtain a *least favorable direction* by choosing a parametric submodel $P_{\gamma_n(h_0)}$ in the null hypothesis that minimizes this distance in $h_{0\eta}$.

The asymptotic power envelope obtained through this least favorable direction serves as a semiparametric power envelope and a test that achieves this bound is AUMPU(α, η_0). However, such a test depends on the nuisance parameter η_0. When there exists a test statistic that is asymptotically equivalent to an AUMPU(α, η_0) test and does not depend on η_0, such a test has the asymptotic optimality property uniformly over η_0 (i.e. AUMPU(α)). Choi, Hall, and Schick (1996) (hereafter CHS) call this test statistic an *efficient test statistic.*

The central step in constructing an asymptotically optimal test is to find the least favorable direction. Paralleling the literature of semiparametric efficiency, the least favorable direction is found by projecting the score with respect to c at $c = 0$ in (13) onto the tangent space of the nuisance parameter η_0 under the null hypothesis (i.e. $c = 0$) (e.g. Begun, Hall, Huang, and Wellner (1983), or Bickel, Klassen, Ritov, and Wellner (1993) (hereafter, BKRW)). An asymptotic optimal test is constructed from a sample version of the L_2-norm of the efficient score obtained from this projection. In the appendix, we compute the efficient score as

$$\ell_1^*(z) \triangleq \frac{e(z, \beta_0) b^*(x)}{\sigma^2(x)} \tag{17}$$

where $b^*(x) \triangleq (S_\sigma b_0)(x)$ and S_σ is defined by

$$(S_\sigma a)(x) \triangleq a(x) - \frac{\mathbf{E}[\sigma^{-2}(X) a(X) | X'\theta_0 = x'\theta_0]}{\mathbf{E}[\sigma^{-2}(X) | X'\theta_0 = x'\theta_0]}, \tag{18}$$

and $\sigma^2(x) \triangleq \mathbf{E}[\varepsilon^2 | X = x]$. Note that S_σ is an orthogonal projection in the L_2-space with the inner product

$$\langle f, g \rangle_\sigma \triangleq \int f(x) g(x) \sigma^{-2}(x) P(dx). \tag{19}$$

It is worth noting that when we know θ_0, the tangent space becomes smaller, but the projection remains the same and so does the efficient score in (17). Therefore, the estimation of θ_0 is *ancillary* to the testing problem in the sense that the semiparametric power envelope does not change due to the lack of the knowledge of θ_0. As we will see later, our optimal test achieves this ancillarity by reparametrizing $\mu(X'\theta)$ into $\mu(F_\theta(X'\theta))$, where F_θ is the distribution function of $X'\theta$.

An optimal test rejects the null hypothesis for large values of

$$T_n^* \triangleq \left\{ \frac{1}{\sqrt{n}\sigma_e} \sum_{i=1}^n \ell_1^*(Z_i) \right\}^2 = \left\{ \frac{1}{\sqrt{n}\sigma_e} \sum_{i=1}^n \frac{e(Z_i; \beta_0) b^*(X_i)}{\sigma^2(X_i)} \right\}^2,$$

where $\sigma_e^2 = \mathbf{E}[\sigma^{-2}(X)(b^*)^2(X)]$. We obtain a feasible test statistic by replacing the unknown components β_0, σ_e, σ, and b^* with their appropriate estimators:

$$T_n \triangleq \left\{ \frac{1}{\sqrt{n}\hat{\sigma}_e} \sum_{i=1}^n \frac{e(Z_i; \hat{\beta}) \hat{b}^*(X_i)}{\hat{\sigma}^2(X_i)} \right\}^2,$$

where $\hat{\sigma}^2(x)$ and $\hat{\sigma}_e$ are estimators of $\sigma^2(x)$ and σ_e^2, and $\hat{b}^*(x)$ is a nonparametric estimator for $b^*(x)$ which is defined prior to (18). An estimator $e(Z_i; \hat{\beta})$ based on the sieves method is introduced in the next section. The nonparametric estimators $\hat{\sigma}$, $\hat{\sigma}_e$, and $\hat{b}^*$ can be constructed using the usual nonparametric estimation methods. A set of high-level conditions for these estimators and discussions about their lower level conditions and references are provided in the next section.

The feasible and infeasible test statistics T_n^* and T_n are asymptotically equivalent under regularity conditions as we discuss in a later section. Hence T_n is an efficient test statistic and a test of asymptotic level α based on T_n is AUMPU(α). Note also that the test is asymptotically pivotal. Indeed, under these regularity conditions,

$$T_n \to \chi_1^2 \text{ under the null hypothesis.}$$

The asymptotic pivotalness comes as a by-product of confining our attention to the interesting alternatives, discarding the omnibus approach. It is worth noting that many omnibus semiparametric tests are known to be asymptotically nonpivotal (e.g. Nikitin (1995), Stute (1997), Andrews (1997), Bierens and Ploberger (1997)). We delineate the conditions for the results discussed here in the next section.

4 Asymptotic Properties of the Tests

In this subsection, we delineate the technical conditions for the asymptotic properties of the test based on T_n. Given a random sample of size n, $(Z_i)_{i=1}^n$, a test of a single index restriction can be analyzed through the asymptotic analysis of the following function-parametric marked empirical process

$$R_n(w, \beta_0) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w(X_i) e(Z_i; \beta_0),$$

where $w(\cdot)$ denotes a member of an appropriate function space $\mathcal{W}_0 \subset L_2(P_X)$. Here P_X denotes the distribution of X under P_0 and $L_2(P_X)$, the space of L_2-bounded measurable functions with respect to $||\cdot||_{2,P_X}$ where $||f||_{2,P_X} = \{\int f^2 dP_X\}^{1/2}$.

In the omnibus test, a test statistic is constructed as a functional of $R_n(\cdot, \beta_0)$ and $\mathcal{W}_0$ is chosen to be a space of functions whose linear span is dense in $L_2(P_X)$ in weak topology (Stinchcombe and White (1998)). Examples of such function spaces $\mathcal{W}_0$ are $\mathcal{W}_0 = \{w(x) = 1(x \leq t) : t \in \mathbf{R}^{d_X}\}$ (e.g. Andrews (1997) and Stute (1997)) and $\mathcal{W}_0 = \{w(x) = \exp(it'x) : t \in \mathbf{R}^{d_X}, i = \sqrt{-1}\}$ (Bierens (1990)). See Escanciano (2006) for other interesting choices of $\mathcal{W}_0$. For a general characterization of $\mathcal{W}_0$ required for omnibus tests of conditional mean models, see Stinchcombe and White (1998). The result of Lemma 1 and our preceding development of an optimal test suggest that we choose $\mathcal{W}_0 = \{b^*/\sigma^2\}$.

The function b^*/σ^2 and the parameter $\beta_0 = (\theta_0, \mu_0)$ are in general unknown, and we assume that consistent estimators $\hat{b}^*/\hat{\sigma}^2$ and $\hat{\beta} = (\hat{\theta}, \hat{\mu})$ with a certain rate of convergence are available. To keep the exposition simple, we provide high-level conditions for $\hat{b}^*/\hat{\sigma}^2$ and $\hat{\theta}$ suppressing the details about their estimation method, but delineate the estimation procedure of $\hat{\mu}$ and the accompanying conditions.

Let us define the following feasible residual-marked empirical process,

$$R_{1,n}(w) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w(X_i) e(Z_i; \hat{\beta}). \tag{20}$$

In particular, we can obtain an estimator $\hat{\theta}$ for $\theta_0 \in \Theta$ that is $\sqrt{n}$-consistent (see Powell, Stock, and Stoker (1989)). Using this estimator we can construct an estimator $\hat{\mu}(X_i'\hat{\theta})$ for $\mu_0(X_i'\theta_0)$ where $\hat{\mu}(\cdot)$ denotes a nonparametric estimator of μ_0. Our optimal test is based on the test statistic

$$T_n = \left\{ \frac{R_{1,n}(\hat{b}^*/\hat{\sigma}^2)}{\hat{\sigma}_e} \right\}^2 .$$

Suppose that we are given with a $\sqrt{n}$-consistent estimator $\hat{\theta}$ of θ_0, and consider the following procedure to obtain $e(Z_i; \hat{\beta})$. It is convenient for our purpose to normalize the conditioning variable by taking a quantile transform of $X'\theta$. Define

$$U_{n,\theta,i} \triangleq F_{n,\theta,i}(X_i'\theta) \triangleq \frac{1}{n} \sum_{j=1, j \neq i}^{n} 1\{X_j'\theta \leq X_i'\theta\} \text{ and } U_{\theta,i} \triangleq F_\theta(X_i'\theta),$$

where $F_\theta(\cdot)$ is the cdf of $X'\theta$ and $F_{n,\theta,i}$ is the empirical cdf which is implicitly defined above. We simply write $U_i \triangleq U_{\theta_0,i}$ and $U \triangleq F_{\theta_0}(X'\theta_0)$. In this paper, we consider a series estimator as follows. First, we introduce a vector of basis functions:

$$p^K(u) \triangleq (p_{1K}(u), \cdots, p_{KK}(u))', \; u \in [0,1]. \tag{21}$$

Using these basis functions, we approximate $g(u; \theta_0) = \mathbf{E}[Y|U = u]$ by $p^K(u)'\pi$ for an appropriate vector π. Define a series estimator $\hat{g}$ as

$$\hat{g}(u; \theta) \triangleq p^K(u)'\pi_n(\theta), \tag{22}$$

where $\pi_n(\theta) = [P_n'(\theta)P_n(\theta)]^{-1}P_n'(\theta)a_n$,

$$a_n \triangleq \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \text{ and } P_n(\theta) \triangleq \begin{bmatrix} p^K(F_{n,\theta,1}(X_1'\theta))' \\ \vdots \\ p^K(F_{n,\theta,n}(X_n'\theta))' \end{bmatrix}. \tag{23}$$

Then, we obtain residuals

$$e(Z_i; \hat{\beta}) \triangleq Y_i - \hat{g}(F_{n,\hat{\theta}}(X_i'\hat{\theta}); \hat{\theta}), \; 1 \leq i \leq n. \tag{24}$$

Conditions for basis functions and others needed for the nonparametric estimation in $e(Z_i;\hat{\beta})$ are mostly subsumed into a high-level condition in Assumption 3(i)(c) below and its lower-level conditions are relegated to Appendix C. We introduce a set of regularity assumptions.

Assumption 1: (i) $\mu_0(v)$ is continuously differentiable in v with a uniformly bounded derivative $\dot{\mu}_0$. (ii) $\mathbf{E}[e^4(Z;\beta_0)] < \infty$ and $\mathbf{E}||X||^4 < \infty$. (iii) There exists a neighborhood B of θ_0 such that (a) for all $\theta \in B$, $X'\theta$ is continuous, (b)

$$\sup_{\theta \in B} \sup_{v \in \mathrm{supp}(X'\theta)} |F_\theta(v+\delta) - F_\theta(v-\delta)| < C\delta \text{ for all } \delta > 0$$

where $\mathrm{supp}(X'\theta)$ denotes the support of $X'\theta$, and (c) the conditional density function $f_\theta(y,x|u)$ of (Y,X) given $F_\theta(X'\theta) = u$ satisfies that for all (y,x) in the support of (Y,X) and for all $u \in [0,1]$,

$$\sup_{u_1 \in [0,1]:|u-u_1|<\delta} |f_u(y,x|u) - f_u(y,x|u_1)| \leq \varphi_u(y,x)\delta$$

where $\varphi_u(y,x)$ is a real valued function such that $\int y\varphi_u(y,x)dy < C$ and $\int \varphi_u(y,x)dx < Cf_Y(y)$ with $f_Y(y)$ denoting the density of Y and C denoting an absolute constant.

Assumption 2 : The conditional density $f(\varepsilon|X)$ of ε given X in (14) satisfies that (a) $\int \varepsilon f(\varepsilon|X)d\varepsilon = 0$ and $\int \varepsilon^2 f(\varepsilon|X)d\varepsilon < \infty$, P_X-a.s., (b) $f(\varepsilon|X)$ is continuously differentiable in ε with the derivative $\dot{f}(\varepsilon|X)$ satisfying $\int_{\{f(\varepsilon|X)>0\}} \{\dot{f}(\varepsilon|X)/f(\varepsilon|X)\}^2 f(\varepsilon|X)d\varepsilon < \infty$, P_X-a.s.

Conditions in Assumption 1 are used to resort to Lemma 1U, a general result that ensures Assumption 3(i)(c) below. Conditions in Assumption 2 are made to ensure the *regularity* of the parametric model indexed by $c \in \mathbf{R}$ (see Proposition 3.4.1 of BKRW).[8] For a function $w \in \mathcal{W}$ with $\mathbf{E}[|w(X)|] < \infty$, we define

$$g_w(u) \triangleq \mathbf{E}[w(X)|U=u].$$

In order to obtain the uniform behavior of an empirical process indexed by $\mathcal{W}$, we need an appropriate device to control the size of the space $\mathcal{W}$. Let $L_p(P)$, $p \geq 1$, be the space of L_p-bounded functions: $||f||_{p,P} \triangleq \{\int |f(x)|^p P(dx)\}^{1/p} < \infty$, and for a space of functions $\mathcal{F} \subset L_p(P)$ for $p \geq 1$, let $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_{p,P})$, the bracketing number of $\mathcal{F}$ with respect to the norm $||\cdot||_{p,P}$, to be the smallest number r such that there exist $f_1, \cdots, f_r$ and $\Delta_1, \cdots, \Delta_r \in L_p(P)$ such that $||\Delta_i||_{p,P} < \varepsilon$ and for all $f \in \mathcal{F}$, there exists $i \leq r$ with $||f_i - f||_{p,P} < \Delta_i$. The logarithm of the bracketing number is called bracketing entropy. We introduce additional assumptions.

Assumption 3 : (i) The function b^*/σ^2 in (18) satisfies $0 < \mathbf{E}[(b^{*4}/\sigma^8)(X)] < \infty$. Furthermore, there exists a class $\mathcal{W}$ such that (a) $b^*/\sigma^2 \in \mathcal{W}$, $P\{\hat{b}^*/\hat{\sigma}^2 \in \mathcal{W}\} \to 1$ as $n \to \infty$ and $||\hat{b}^*/\hat{\sigma}^2 -$

[8] In particular, the needed condition that the semiparametric efficiency bound for c being bounded below from zero is implied by Assumption 3(i) below.

$b^*/\sigma^2||_\infty = o_P(n^{-1/4})$, (b) for some $C > 0$ and $c \in [0, 2)$, $\log N_{[]}(\varepsilon, \mathcal{W}, ||\cdot||_\infty) \le C\varepsilon^{-c}$, (c)

$$\sup_{w \in \mathcal{W}} \left| R_{1,n}(w) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{w(X_i) - g_w(U_i)\}\{Y_i - \mathbf{E}[Y_i|U_i]\} \right| = o_P(1), \tag{25}$$

and (d) $\hat{\sigma}_e^2 \to_p \sigma_e^2 > 0$. (ii) The estimator $\hat{g}(\cdot)$ defined in (24) satisfies $||\hat{g} - g||_\infty = o_P(n^{-1/4})$ where $g(u) = \mathbf{E}[Y|U = u]$ has the first order derivative $\dot{g} \in L_2(P_X)$ satisfying $\sup_{u \in [0,1]} \dot{g}(u) < \infty$, (iii) The estimators $\hat{\theta}$ and $\hat{\sigma}_e$ satisfy $||\hat{\theta} - \theta_0|| = O_P(n^{-1/2})$ and $\sup_x |\hat{\sigma}_e(x) - \sigma_e(x)| = o_P(1)$.

The above conditions are high-level conditions. Condition (i)(a) follows from certain smoothness properties of f and its estimator $\hat{f}$ along with appropriate trimming factors (e.g. see Powell, Stock, and Stoker (1989)). Condition (i)(b) is satisfied when b^* belongs to a class of smooth functions with a certain order of differentiability. For example, when b^* has a uniformly bounded partial derivatives up to the order $\lfloor h \rfloor$, the greatest integer smaller than h, and its highest derivatives are of Lipschitz order $h - \lfloor h \rfloor$, then Assumption 3(i)(b) is satisfied with $c = d_X/h$ (e.g. see Theorem 2.7.1 in van der Vaart and Wellner (1996) and also see Andrews (1994).) Uniform consistency can be obtained, for example, by using Newey (1997) or Song (2006) in the case of series estimator, and Andrews (1995) in the case of kernel estimators. The condition $P\{\hat{b}^*/\hat{\sigma}^2 \in \mathcal{W}\} \to 1$ is weaker than the condition $P\{\hat{b}^*/\hat{\sigma}^2 \in \mathcal{W}\} = 1$ from some sufficiently large n on. This latter condition is satisfied when $\hat{b}^*/\hat{\sigma}^2$ satisfies the aforementioned smoothness conditions, as can be fulfilled by choosing kernels or basis functions satisfying the smoothness conditions. For details, see Andrews (1994).

Condition (i)(c) contains an asymptotic representation in (25) of the feasible empirical process $R_{1,n}(w)$. One can prove (25) by using a more general result established in Appendix C. In Appendix C we delineate low-level conditions for basis functions, the space $\mathcal{W}$, and other regularity conditions under which we obtain an asymptotic representation of the form in (25). This result is also of independent interest; for example it can be used to establish the asymptotic distribution of omnibus tests. In the case of kernel estimation, a similar but pointwise result (i.e. with fixed $w \in \mathcal{W}$) was obtained by Stute and Zhu (2005).

Theorem 1 : *Suppose* Assumptions 1-3 *hold. Then the following are satisfied.*
(i) *Under the null hypothesis of single index restriction,*

$$T_n \to_d \chi_1^2,$$

and under the local alternatives such that

$$Y_i = a(X_i)/\sqrt{n} + \mu_0(X_i'\theta_0) + \varepsilon_i \tag{26}$$

with $\mathbf{E}[\varepsilon_i|X_i] = 0$ *and* $\mathbf{E}[|a(X_i)|] < \infty$,

$$T_n \to_p \chi_1^2 \left(\frac{\langle a, S_\sigma b_0 \rangle_\sigma^2}{||S_\sigma b_0||_\sigma^2} \right),$$

where $\langle \cdot, \cdot \rangle_\sigma$ *is as defined in* (19) *and* $|| \cdot ||_\sigma^2$ *is defined as* $||b||_\sigma^2 \triangleq \langle b, b \rangle_\sigma$.
(ii) *For each* $\alpha \in (0,1)$, *the test* $\psi_n = 1\{T_n > c_\alpha\}$ *with* c_α *being determined to deliver an asymptotic level* α *is* AUMPU(α) *for testing* H_0^* *against* H_1^* *in* (15).

The result (i) determines the asymptotic properties of the test T_n. The result is established via the asymptotic equivalence of T_n^* and T_n both under the null hypothesis and under the alternatives. The test is asymptotically pivotal, having a χ_1^2 distribution under the null hypothesis. Under the local alternatives of the form in (26), the test statistic has a limiting noncentral χ_1^2 distribution.

Let us discuss the implications from the result of the local power properties in (i). We confine our attention to local alternatives with the directions $a(x)$ such that $||a^2||_\sigma = 1$ and $a = S_\sigma a$ as a normalization.[9] For such directions a, the noncentrality parameter becomes

$$\frac{\langle a, S_\sigma b_0 \rangle_\sigma^2}{||S_\sigma b_0||_\sigma^2} = \frac{\langle S_\sigma a, b_0 \rangle_\sigma^2}{||S_\sigma b_0||_\sigma^2} = \frac{\langle a, b_0 \rangle_\sigma^2}{||S_\sigma b_0||_\sigma^2}.$$

Therefore, the test has a maximal power when a is in the direction of b_0, and the test has no power when a is orthogonal to b_0 with respect to $\langle \cdot, \cdot \rangle_\sigma$.

Recall that the demarcation of interesting alternatives $\mathcal{P}_M^c$ and uninteresting alternatives $\mathcal{P}_M$ was made in terms of whether a is orthogonal to b_0 with respect to $\langle \cdot, \cdot \rangle$ or not. The directions a against which the test has no power are not necessarily the directions that represent uninteresting alternatives, being orthogonal to b_0 with respect to $\langle \cdot, \cdot \rangle$, but are "tilted" ones. This tilting is due to the optimal incorporation of the information in the conditional moment restriction $\mathbf{E}[\varepsilon|X] = 0$. Hence, a consequence of this tilting is that the space of alternatives against which the test has nontrivial local asymptotic power does not in general coincide with that of interesting alternatives $\mathcal{P}_M^c$. In fact, the coincidence arises only when the demarcation between the interesting and uninteresting alternatives is made in terms of a weighted average partial effect where the weight is given by $\sigma^{-2}(x)$.

We stress that our minimax-based notion of optimality crucially depends on the formulation of the null hypothesis. To illustrate this point, consider the situation in which one is interested in testing the null of $P_0 \in \mathcal{P}_M$ and against alternatives $P_0 \in \mathcal{P}_M^c$. This test is a test of whether the restricted average partial effects coincide with the unrestricted average partial effects. In this situation, the notion of optimality of tests changes accordingly. More specifically, one might consider

[9] Note that $a(x) = (S_\sigma a)(x) + \{a(X) - (S_\sigma a)(X)\}$. The second part $\{a(X) - (S_\sigma a)(X)\}$ cannot be identified separately from $\mu_0(X'\theta_0)$ in the regression formulation in (26), and hence the "effective" direction is the remaining $(S_\sigma a)(x)$.

constructing a test based on the moment restriction (e.g. Newey (1985) and Tauchen (1985))

$$\mathbf{E}[e(Z;\beta_0)b_0(X)] = 0.$$

The restriction suggests that we use the semiparametric empirical process $R_{1,n}(b_0)$ to construct a test statistic. Indeed, the test statistic can be constructed as

$$T_{n,2} \triangleq \left\{ \frac{R_{1,n}(\hat{b})}{\hat{\sigma}_b} \right\}^2 .$$

where $\hat{\sigma}_b$ is an estimator of $\mathbf{E}\left[(Sb_0)(X)e(Z;\beta_0)^2\right]$. Here S is a linear operator defined by $(Sa)(x) \triangleq a(x) - \mathbf{E}\left[a(X)|X'\theta_0 = x'\theta_0\right]$. Then we can show that under similar conditions for Theorem 1, the test statistic $T_{n,2}$ has the limiting distribution of χ_1^2 under the null hypothesis, whereas under the local alternatives of the form in (26),

$$T_{n,2} \to_d \chi_1^2 \left(\frac{\langle a, Sb_0 \rangle^2}{||Sb_0||_2^2} \right).$$

Among the alternatives such that $||a||_2 = 1$ and $a = Sa$, the maximal power is achieved when $a = b_0/||b_0||_2^2$, and the test has no power when $\langle a, b_0 \rangle = 0$. Therefore, the space of local alternatives against which the test based on $T_{n,2}$ has nontrivial local asymptotic power coincides with the space of interesting alternatives $\mathcal{P}_M^c$. Since in this situation of testing $P_0 \in \mathcal{P}_M$, the conditional moment restriction $\mathbf{E}[\varepsilon|X] = 0$ is not needed in the formulation of the null hypothesis, the optimal test should be defined differently depending on the information that is contained in the null hypothesis.

5 Further Extensions

5.1 Average Partial Effects of a Binary Covariate

The development has so far relied on the assumption that the covariate X_1 is a continuous variable. In many cases, the variable of interest is a binary variable. For example, the covariate can be a dummy variable representing the qualitative information about a certain state. In this case, we need to consider a different test statistic because the direction $b_0(x)$ computed in (11) is based on the continuity of the random variable X_1. This section is devoted to analyzing the case when the covariate X_1 of interest is a binary variable. As it turns out, the direction in the Riesz Representation of the linear functional is fully known in this case, leading to a simpler test statistic.

Suppose $X = (X_1, X_2')'$ where $X_1 \in \{0,1\}$ and $X_2 \in \mathbf{R}^{d_{X_2}}$. The average partial effect of X_1 is defined as

$$\mathbf{E}\left[\mathbf{E}[Y|X_1 = 1, X_2] - \mathbf{E}[Y|X_1 = 0, X_2]\right].$$

Under the single index restriction, the average partial effect becomes

$$\mathbf{E}\left[\mu_0(\theta_1 + X_2'\theta_2) - \mu_0(X_2'\theta_2)\right].$$

Let us define the state space of covariates to be $\mathcal{X} \triangleq \{0,1\} \times \mathbf{R}^{d_{X_2}}$. We define a linear functional M on $\mathcal{D}_{P_0}$ as follows. For any $h : \mathcal{X} \to \mathbf{R}$ in $L_2(P_X)$, define

$$Mh \triangleq \mathbf{E}\left[h(1,X_2) - h(0,X_2)\right] = \int \left[h(1,x_2) - h(0,x_2)\right] f(x_2)dx_2,$$

where $f(x_2)$ denotes the density of X_2. Then Mh is an average partial effect of $h(X_1, X_2)$ with respect to X_1. Since $\int h(x)^2 P(dx) < \infty$, the functional M is bounded. The *uninteresting* alternatives are those with $(Mr) = 0$ where $r(X_1, X_2) = \mathbf{E}\left[Y - \mu(X_1\theta_1 + X_2'\theta_2)|X_1, X_2\right]$. By the Riesz Representation Theorem, there exists $b_0 \in L_2(P_X)$ such that $Mh = \langle h, b_0 \rangle$. It is straightforward to find b_0 :

$$b_0(x_1, x_2) = (-1)^{x_1+1}.$$

Note that b is fully known and there is no need to estimate it. Therefore, the suboptimal test statistic that is analogue of $T_{n,2}$ in the previous section can be constructed as

$$T_{n,2}^B \triangleq \left(\frac{1}{\sqrt{n\hat{\sigma}_{eB}^2}} \sum_{i=1}^n (-1)^{X_{i1}+1}(Y_i - \hat{\mu}(X_{i1}\hat{\theta}_{i1} + X_{i2}'\hat{\theta}_2))\right)^2, \tag{27}$$

where $\hat{\sigma}_{eB}^2$ is computed as follows. First, note that

$$\begin{aligned}
\mathbf{E}\left[b_0(X_1,X_2)|U\right] &= P\{X_1 = 1|U\} - P\{X_1 = 0|U\} \\
&= 2P\{X_1 = 1|U\} - 1
\end{aligned}$$

so that by using the fact that $b_0(x_1, x_2) \in \{-1, 1\}$, we deduce

$$\begin{aligned}
\mathbf{E}[(b_0(X_1,X_2) - \mathbf{E}\left[b_0(X_1,X_2)|U\right])^2|U] &= 1 - \{2P\{X_1 = 1|U\} - 1\}^2 \\
&= 4P\{X_1 = 1|U\}P\{X_1 = 0|U\}.
\end{aligned}$$

Therefore, we can estimate σ_{eB}^2 as

$$\hat{\sigma}_{eB}^2 \triangleq \frac{4}{n}\sum_{i=1}^n \hat{P}\{X_{i1} = 1|U_i\}\hat{P}\{X_{i1} = 0|U_i\}.$$

where $\hat{P}\{X_{i1} = 1|U_i\}$ is a consistent estimator for $P\{X_{1i} = 1|U_i\}$. Our test statistic is finally obtained by plugging this into (27). Note that the derivation of the asymptotic properties can be performed by modifying Theorem 1. In particular, the null limiting distribution of the test can be shown to be χ_1^2. The analysis of power under the Pitman local alternatives can be performed similarly as before.

Following the previous development, we construct an asymptotically efficient test in this case of a binary covariate. The formulation of the score and a tangent space can be proceeded similarly, so that we have the efficient score

$$\ell_1^*(z) \triangleq e(z, \beta_0)\sigma^{-2}(x)b^*(x)$$

where

$$b^*(x) \triangleq (-1)^{x_1+1} - \frac{\mathbf{E}[\sigma^{-2}(X)(-1)^{X_1+1}|X'\theta_0 = x'\theta_0]}{\mathbf{E}[\sigma^{-2}(X)|X'\theta_0 = x'\theta_0]}.$$

As developed in previous sections, a nonparametric estimator for b^* can be used to construct a feasible test.

5.2 Multivariate Average Partial Effects

We can extend the framework to multivariate average partial effects. Suppose we are interested in the joint average partial effect of d_1 number of covariates and write it as a column vector

$$\mathbf{E}\left[\frac{\partial}{\partial x_1}\mathbf{E}[Y|X = x]|_{x=X}\right]$$

in $\mathbf{R}^{d_1}$, whose k-th element is $\mathbf{E}\left[(\partial/\partial x_{1,k})\mathbf{E}[Y|X = x]|_{x=X}\right]$ with the obvious individual derivative notation of $\partial/\partial x_{1,k}$. We define the functions b_k by $b_k(\cdot) = -(1/f(\cdot))(\partial/\partial x_{1,k})f(\cdot)$ as in (11) and collect these into a column vector b. Suppose that uninteresting alternatives in this setting are those that make no difference to the joint average partial effects by introducing the single index restriction. The space of these alternatives can be defined in the same way as (12).

The test becomes a J-test in the standard GMM problem (Hansen (1982)). More specifically, an analogue of $T_{n,2}$ is constructed as a quadratic form of a vector process:

$$T_{n,2}^M \triangleq \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{b}(X_i)e(Z_i;\hat{\beta})\right\}' V_n^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{b}(X_i)e(Z_i;\hat{\beta})\right\} \tag{28}$$

where V_n is a consistent weighting matrix for $V \triangleq \mathbf{E}\left[(Sb)(X_i)(Sb)(X_i)'e(Z_i;\beta_0)\right]$. Using similar arguments used to prove Theorem 1, we can show that under the null hypothesis, $T_{n,2}^M \to_d \chi_{d_1}$.

Asymptotically efficient tests can be constructed analogously as before, but are in need of further restrictions on the notion of asymptotic optimality. In view of the hypothesis testing theory in the Euclidean space, the most natural way is to confine candidate tests to those with *asymptotically* *invariant* property as done by CHS.[10] We first adopt the regression formulation as follows:

$$e(Z;\beta) = c'b_0(X) + \varepsilon, \ \mathbf{E}[\varepsilon|X] = 0,$$

and then the testing problem is mapped into testing $H_0^* : c = 0$ against $H_1^* : c \neq 0$. The tangent space for the nuisance parameters does not change when c becomes multivariate. The efficient score b^* is obtained as a coordinate-wise projection onto this tangent space. Details are omitted.

6 Conclusion

This paper considers a situation of testing a single-index restriction where uninteresting alternatives are characterized by a linear functional equation that represents the coincidence of a restricted average partial effect with its unrestricted version. A new notion of asymptotic optimality of tests suited to this situation is suggested in which the set of uninteresting alternatives are eliminated and after that, an exclusive focus is drawn on a set of least favorable alternatives. We find that the least favorable set is characterized as a single direction of Pitman local alternatives, and building on this, we define an optimal test to be one that achieves the semiparametric power envelope, following CHS.

We suggest an asymptotically distribution free test that is optimal in the sense defined previously. Based on a general result of semiparametric empirical processes involving series-based conditional mean estimators, we explore the asymptotic properties of the test, with a particular interest in the behavior of local asymptotic powers. The proposed optimal test has maximal local power against alternatives in the interesting subset of alternatives.

We want to emphasize that our framework can be applied to other linear functionals and the basic idea suggested in the paper does not confine itself to single index restrictions either. For example, one could consider a conditional moment restriction in general combined with a demarcated subspace of alternatives given by a linear functional. However, the asymptotic theory required will depend on the specific context.

Finally, the contribution of this paper can be viewed as a step toward unifying the inference procedure of specification test and estimation in a single decision theoretic framework. The main thesis of this paper is to design a specification test that envisions its eventual use in the estimation. The natural, ultimate question in this context would concern how the uncertainty due to the lack of information about the specification can be properly incorporated in the subsequent estimation procedure. We believe that this remains a very interesting research agenda, may be, a challenging

[10] A test φ_n is asymptotically invariant if the limit test φ is rotationally invariant $\varphi(u) = \varphi(Ru)$ for all $u \in \mathbf{R}^d$ for any orthogonal matrix $R : R'R = I$. Note that the limit test φ is defined to be a measurable map such that

$$\lim_{n\to\infty} \mathbf{E}_{\gamma_n(h)}\varphi_n = \int \varphi(u)d\Phi(u - B^{*1/2}h_c)$$

for all h in a certain hyperplane under the local alternatives in (16). Here Φ denotes the standard normal distribution function and B^* the efficient information for c. For details, see CHS.

one.[11]

7 Appendix

7.1 Appendix A: Semiparametric Efficient Tests

We provide a review of semiparametric efficient tests based on CHS. In order to characterize the space of alternatives, we focus on the following class of local alternatives. First for each $(h_c, h_\eta) \in \mathbf{R} \times \mathcal{H}$, define two sequences

$$\begin{aligned} c_n(h_c) &\triangleq n^{-1/2}h_c + o(n^{-1/2}) \text{ and} \\ \eta_n(h_\eta) &\triangleq \eta_0 + n^{-1/2}h_\eta + o(n^{-1/2}). \end{aligned} \tag{29}$$

The vector $h \triangleq (h_c, h_\eta)'$ denotes directions in which the local parameter $\gamma_n(h) \triangleq (c_n(h_c), \eta_n(h_\eta))$ deviates from the point $(0, \eta_0)$. Let $\mathbf{R} \times \mathcal{H}$ be a Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$. It is convenient to introduce the local asymptotic normality (LAN) of the likelihood ratio process:

$$L_n(h) \triangleq \log \frac{dP_{n,\gamma_n(h)}}{dP_{n,\gamma_0}} = S_n h - \frac{1}{2}\sigma^2(h) + r_n(h)$$

where $S_n = (S_{nc}, S_{n\eta})'$ is a random linear functional which is asymptotically centered Gaussian with kernel B under the null hypothesis, and $r_n(h) = o_P(1)$ for every h under the null hypothesis. Hence the variance $\sigma^2(h)$ of $S_n h$ is equal to $\langle h, Bh \rangle$. The LAN property follows when the local alternatives are Hellinger differentiable with respect to parameters, and is useful for investigating the asymptotic behavior of the test statistic under the local alternatives by using Le Cam's third lemma (See Begun, Hall, Huang, and Wellner (1983) or BKRW for details.)

Consider a test ψ_n taking values in $\{0, 1\}$ depending on the rejection and acceptance of the null hypothesis. For the moment, let us consider the one-sided test of

$$H_0 : c = 0 \text{ against } H_1 : c > 0.$$

Then using the LAN property, we can write

$$\begin{aligned} \mathbf{E}_{\gamma_n(h)}\psi_n &= \mathbf{E}_{\gamma_0}\psi_n \exp\left(L_n(h)\right) + o(1) \\ &= \mathbf{E}_{\gamma_0}\psi_n \exp\left(S_n h - \frac{1}{2}\sigma^2(h) + r_n(h)\right) + o(1). \end{aligned}$$

Fix $h_1 = (h_{1c}, h_{1\eta})$ and consider testing the simple hypothesis $h_0 = (0, h_{0\eta})$ against h_1. Then the Neyman-Pearson lemma gives an optimal test φ_n of asymptotic level α in the following form: $\varphi_n = 1$ if

$$S_n(h_1 - h_0) - \frac{1}{2}\{\sigma^2(h_1) - \sigma^2(h_0)\} + r_n(h_1) - r_n(h_0) > c_n$$

[11] See Leeb and Pötcher (2005) and references therein for issues of post-model selection inferences.

and $\varphi_n = 0$ otherwise. And for this test, it is a straightforward matter to obtain the following bound for the power of the test

$$\limsup \mathbf{E}_{\gamma_n(h_1)}\varphi_n \leq 1 - \Phi(z_\alpha - \sigma(h_1 - h_0)) \tag{30}$$

where z_α is the upper α-quantile of the standard normal distribution function Φ.

Now, we aim to devise a test that is uniformly most powerful at each point of $h_{0\eta} \in \mathcal{H}$. The bound for the power of the test is attained by an optimal test against a simple alternative corresponding to the *least favorable direction.* Let $(B_{ij})_{i,j=1,2}$ denote the partition of B such that B_{11} is the information for c, B_{22} is the information for η, and B_{12} and B_{21} are coinformations. Obviously from (30), the least favorable direction is obtained by minimizing $\sigma(h_1 - h_0)$ in $h_{0\eta}$ and is found to be $h_{1\eta} - h^*_{0\eta}$ where $h^*_{0\eta} = h_{1\eta} + B_{22}^{-1}B_{21}h_{1c}$. Hence the point $(0, h^*_{0\eta})$ is the projection of h_1 onto the local null space under the inner product induced by B, namely, $\langle h, g\rangle_B = \langle h, Bg\rangle,\ h, g \in \mathcal{H}$. By plugging in this least favorable direction, we obtain

$$\limsup \mathbf{E}_{\gamma_n(h_1)}\psi_n \leq 1 - \Phi(z_\alpha - \sigma(B^{*1/2}h_{1c}))$$

where $B^* = B_{11} - B_{12}B_{22}^{-1}B_{21}$ is what is called *efficient information.* Let us define the efficient score S^*_n as $S^*_n a = S_{nc}a - S_{n\eta}B_{22}^{-1}B_{21}a,\ a \in \mathbf{R}$. Since c is a scalar, so are S_{nc} and S^*_n. Note that S^*_n depends on η_0 and we write $S^*_n(\eta_0)$ explicitly. Now, an optimal test is obtained by taking $\varphi_n = 1\left\{B^{*-1/2}S^*_n(\eta_0) \geq z_\alpha\right\}$. The resulting test φ_n does not depend on $h_1 = (h_{1c}, h_{1\eta})$. Hence the test is asymptotically uniformly most powerful (AUMP(α, η_0)) at the level α and at the nuisance parameter η_0.

The procedure easily applies to a two-sided test. A test ψ_n is *asymptotically unbiased* at η_0 if $\limsup_n \mathbf{E}_{\gamma_n(h_0)}\psi_n \leq \liminf_n \mathbf{E}_{\gamma_n(h_1)}\psi_n$ for every $h_0 = (0, h_{0\eta})$ and $h_1 = (h_{1c}, h_{1\eta})$ with $h_{1c} \neq 0$. Then Theorem 2 of CHS gives the following bound for the local power:

$$\limsup \mathbf{E}_{\gamma_n(h)}\psi_n \leq \Phi(|B^{*1/2}h_c| - z_{\alpha/2}) + \Phi(-|B^{*1/2}h_c| - z_{\alpha/2})$$

for all $h = (h_c, h_\eta) \in \mathbf{R} \times \mathcal{H}$. The two-sided test that is AUMPU(α, η_0) among the asymptotically unbiased tests is given by

$$\varphi_n \triangleq 1\left\{|B^{*-1/2}S^*_n(\eta_0)| \geq z_{\alpha/2}\right\}.$$

To apply this framework to our context of testing single index restrictions, we need to compute the efficient score and the efficient information. To this end, we need to find a tangent space of the nuisance parameters. First fix $P_{\gamma_0} \in \mathcal{P}_0$ where $\gamma_0 = (0, \eta_0)$ and introduce

$$\mathcal{P}_1 \triangleq \{P_{(c,\eta_0)} \in \mathcal{P}^* : c \in \mathbf{R}\} \text{ and } \mathcal{P}_2 \triangleq \{P_{(0,\eta)} \in \mathcal{P}^*_0 : \eta \in \mathcal{H}\}.$$

The space $\mathcal{P}_1$ contains alternatives (i.e., $c \neq 0$) with the nuisance parameter η fixed at $\eta = \eta_0$. The space $\mathcal{P}_2$ contains probabilities that satisfy the null hypothesis (i.e., $c = 0$) with η running in $\mathcal{H}$.

Fix $P_0 \in \mathcal{P}_1$ and let $f_{\varepsilon|X}$ and f_X be the conditional density of ε given X and the density of X under P_0. The log-likelihood in the regression setup is given by

$$\log f_Y(y; c, \eta_0) = \log f_{\varepsilon|X}(y - \mu_0(x'\theta_0) - cb_0(x)) + \log f_X(x).$$

Therefore, its score $\dot{\ell}_1$ with respect to c at $c = 0$ is equal to

$$\dot{\ell}_1(z) = \frac{\partial}{\partial c} \log f_{\varepsilon|X}(e(z, \beta_0) - cb_0(x)|x)|_{c=0} = -b_0(x)s_0(z),$$

where $s_0(z) = \dot{f}_{\varepsilon|X}(y - \mu_0(x'\theta_0)|x)/f_{\varepsilon|X}(y - \mu_0(x'\theta_0)|x)$ and $\dot{f}_{\varepsilon|X}(\cdot|x)$ is the derivative as in Assumption 2(i). The efficient score at $P_0 \in \mathcal{P}_1 \cap \mathcal{P}_2$ is computed as the orthogonal complement from the projection of this score $\dot{\ell}_1$ onto the tangent space $\dot{\mathcal{P}}_2$ at P_0 of $\mathcal{P}_1 \cap \mathcal{P}_2$ (e.g. BKRW, p.70). The tangent space $\dot{\mathcal{P}}_2$ is the closed linear span of the tangent spaces for the regular parametric submodels in $\mathcal{P}_2$. Let us construct the parametric submodels in $\mathcal{P}_2$. Define $\eta_t = (\theta_t, \mu_t(\cdot), f_{\varepsilon|X,t}(\cdot), f_{X,t}(\cdot)),\ t \in \mathbf{R}$ and $P_t = P_{\eta_t}$, where at $t = 0$, it is satisfied that $\eta_t = \eta_0$, and hence η_0 is the parameter corresponding to P_0. Note that $\mu_t(\cdot)$ is determined by

$$\mu_t(v) \triangleq \mathbf{E}_t[Y|X'\theta_t = v] = \mu_0(v) + \xi_t(v), \text{ say}, \tag{31}$$

where the conditional expectation is with respect to P_t. We define a class of submodels $\mathcal{P}_{2,S} = \{P_t : \mathbf{E}_t[Y - \mu_t(X'\theta_t)|X] = 0,\ t \in \mathbf{R}\} \subset \mathcal{P}_2$, where the conditional expectation operator $\mathbf{E}_t(\cdot|X)$ is with respect to the conditional density $f_{\varepsilon|X,t}(\cdot)$. Then by applying the implicit function theorem to $\mathbf{E}_t[Y - \mu_t(X'\theta_t)|X] = 0,\ t \in \mathbf{R}$, and using (31), we deduce

$$\begin{aligned} 0 &= \frac{\partial \mathbf{E}_t[Y - \mu_0(X'\theta_0)|X]}{\partial t}|_{t=0} - \frac{\partial}{\partial t}\mu_t(X'\theta_t)|_{t=0} \\ &= \mathbf{E}_0[e(Z;\beta_0)S_{\eta_0}(Z)|X] - \frac{\partial}{\partial v}\mu_0(X'\theta_0)X'\dot{\theta}_0 - \frac{\partial}{\partial v}\xi_0(X'\theta_0)X'\dot{\theta}_0 - \dot{\xi}_0(X'\theta_0), \end{aligned} \tag{32}$$

where $(\partial/\partial v)\mu_0(\cdot)$ denotes the first order derivative of $\mu_0(v)$ and the functions with dots represent derivatives with respect to t at $t = 0$ and $S_{\eta_0}(z)$ is the score defined by $S_{\eta_0}(z) \triangleq \partial \log f_{\varepsilon|X,t}(y - \mu_0(x'\theta_0))/\partial t|_{t=0.}$ Therefore,

$$\mathbf{E}_0[e(Z;\beta_0)S_{\eta_0}(Z)|X] = \mathbf{E}_0\left[e(Z;\beta_0)S_{\eta_0}(Z)|X'\dot{\theta}_0, X'\theta_0\right]. \tag{33}$$

By defining $g_0(x) \triangleq f_{X|X'\dot{\theta}_0, X'\theta_0}(x|x'\dot{\theta}_0, x'\theta_0)/f_{X|X'\theta_0}(x|x'\theta_0)$, we write the above equality as

$$\mathbf{E}[e(Z;\beta_0)s_0(Z)|X] = \mathbf{E}[e(Z;\beta_0)s_0(Z)g_0(X)|X'\theta_0]. \tag{34}$$

Hence we conjecture that the tangent space $\dot{\mathcal{P}}_2$ at $P_0 \in \mathcal{P}_1 \cap \mathcal{P}_2$ is given by

$$\dot{\mathcal{P}}_2 = \{s \in \tilde{L}_2(P_0) : \mathbf{E}[e(Z;\beta_0)s(Z)|X] = \mathbf{E}[e(Z;\beta_0)s(Z)g(X)|X'\theta_0] \text{ a.s., for some } g \in \breve{L}_2(P_0)\}$$

where $\tilde{L}_2(P_0) = \{s \in L_2(P_0) : \mathbf{E}[s(Z)] = 0\}$ and $\breve{L}_2(P_0) = \{g \in L_2(P_0) : \mathbf{E}[g(X)|X'\theta_0] = 1\}$. In the following we show that $\dot{\mathcal{P}}_2$ is indeed the tangent space and compute the projection $\Pi[\dot{\ell}_1|\dot{\mathcal{P}}_2]$. Let us define

$$R_h(X) \triangleq \mathbf{E}[e(Z;\beta_0)h(Z)|X] - \frac{\mathbf{E}[\sigma^{-2}(X)e(Z;\beta_0)h(Z)|X'\theta_0]}{\mathbf{E}[\sigma^{-2}(X)|X'\theta_0]}.$$

Then we have the following.

Lemma A1 : Suppose Assumptions 1(iii) and 3 hold. Then
(i) $\dot{\mathcal{P}}_2$ is the tangent space.
(ii) $\Pi[h|\dot{\mathcal{P}}_2](Z) = h(Z) - \mathbf{E}[h(Z)] - e(Z;\beta_0)\sigma^{-2}(X)R_h(X)$ for $h \in L_2(P_0)$.

Proof of Lemma A1: (i) We follow the procedure in Example 3.2.3 of BKRW. Let $\dot{T}$ be the tangent space. Since any $s \in \dot{T}$ satisfies (33), we have $\dot{T} \subset \dot{\mathcal{P}}_2$.

It suffices to show that for any $s \in \dot{\mathcal{P}}_2$, we can construct a parametric submodel in $\mathcal{P}_2$ with a score s. We fix an arbitrary $s \in \dot{\mathcal{P}}_2$ that satisfies the equation (34) for some g such that $\mathbf{E}[g(X)|X'\theta_0] = 1$. Define

$$\psi(Z) \triangleq e(Z;\beta_0)\left[1 - \{g(X)f_{z|X'\theta_0}(Z|X'\theta_0)/f_{z|X}(Z|X)\}\right].$$

The function ψ satisfies $\mathbf{E}[s(Z)\psi(Z)|X] = 0$, since $s \in \dot{\mathcal{P}}_2$. We take $\tau(Z) \triangleq \mathbf{E}\left[\psi(Z)^2|X\right]^{-1}\psi(Z)$ so that $\mathbf{E}[\psi(Z)\tau(Z)|X] = 1$ and $\mathbf{E}[s(Z)\tau(Z)|X] = 0$. Define

$$f_{t,\delta}(z|x) \triangleq \frac{f_0(z|x)\xi\left(\delta\tau(z) + ts(z)\right)}{\int f_0(z|x)\xi\left(\delta\tau(z) + ts(z)\right)d\lambda(z)}$$

with $\xi(u) = 2(1+e^{-2u})^{-1}$ and f_0 is the density of the distribution P_0 in $\mathcal{P}_2$ so that $\mathbf{E}_{P_0}[e(Z;\beta_0)|X = x] = 0$. Here λ denotes the dominating measure of P_Z, the distribution of Z. Then we can easily check that

$$\begin{aligned} \frac{\partial}{\partial t}\log f_{t,\delta}(z|x) &= s(z),\ \frac{\partial}{\partial \delta}\log f_{t,\delta}(z|x) = \tau(z) \text{ and} \\ \sqrt{f_{t,\delta}(z|x)} - \sqrt{f_0(z|x)} &= S_{t,\delta}(z) + o(\delta) + o(t), \end{aligned}$$

where $S_{t,\delta}(z) \triangleq (\delta\tau(z) + ts(z))/2$, and that

$$\begin{aligned} \mathbf{E}_{t,\delta}[e(Z;\beta_0)|X] - \mathbf{E}_{P_0}[e(Z;\beta_0)|X] &= \mathbf{E}_{P_0}\left[e(Z;\beta_0)\{f_{t,\delta}(Z|X) - f_0(Z|X)\}/f_0(Z|X)|X\right] \\ &= \mathbf{E}_{P_0}\left[e(Z;\beta_0)S_{t,\delta}(z)|X\right] + o(\delta) + o(t). \end{aligned}$$

The notation $\mathbf{E}_{t,\delta}(\cdot|X)$ indicates the conditional expectation with respect to $f_{t,\delta}(z|x)$. Choose a path $\beta_t = (\theta_t, \mu_t)$ such that

$$\mu_t(X'\theta_t) = \mathbf{E}_{P_0}\left[S_{t,\delta}(Z)e(Z;\beta_0)g(X)|X'\theta_0\right],$$

so that we have

$$\begin{aligned}\mathbf{E}_{P_0}[e(Z;\beta_t)|X]-\mathbf{E}_{P_0}[e(Z;\beta_0)|X] &= -\mathbf{E}_{P_0}\left[S_{t,\delta}(Z)e(Z;\beta_0)g(X)|X'\theta_0\right]\\ &= \mathbf{E}_{P_0}\left[S_{t,\delta}(Z)\left(\psi(Z)-e(Z;\beta_0)\right)|X\right].\end{aligned}$$

Then, from this choice, it follows that $\mathbf{E}_{t,\delta}[e(Z;\beta_t)|X]$ is equal to

$$\begin{aligned}&\mathbf{E}_{P_0}[e(Z;\beta_0)|X]+\mathbf{E}_{t,\delta}[e(Z;\beta_t)|X]-\mathbf{E}_{P_0}[e(Z;\beta_0)|X]\\ =\;&\mathbf{E}_{t,\delta}[e(Z;\beta_0)|X]-\mathbf{E}_{P_0}[e(Z;\beta_0)|X]+\mathbf{E}_{P_0}[e(Z;\beta_t)|X]-\mathbf{E}_{P_0}[e(Z;\beta_0)|X]+o(t)\\ =\;&\mathbf{E}_{P_0}[e(Z;\beta_0)S_{t,\delta}(Z)|X]+\mathbf{E}_{P_0}[S_{t,\delta}(Z)\left(\psi(Z)-e(Z;\beta_0)\right)|X]+o(\delta)+o(t)\\ =\;&\mathbf{E}_{P_0}[\psi(Z)S_{t,\delta}|X]+o(\delta)+o(t)=(1/2)\mathbf{E}_{P_0}[\psi(Z)(\delta\tau(Z)+ts(Z))|X]+o(\delta)+o(t).\end{aligned}$$

However, by the choice of ψ,

$$\mathbf{E}_{P_0}[\psi(Z)(\delta\tau(Z)+ts(Z))|X]=(1/2)\delta+o(\delta)+o(t).$$

The equation $\mathbf{E}_{t,\delta}[e(Z;\beta_t)|X]=0$ has a root at $\delta=\delta(t)$ with $\delta=o(t)$. This implies that

$$\mathbf{E}_{P_0}\left[S_{t,\delta(t)}(Z)e(Z;\beta_0)|X\right]=\mathbf{E}_{P_0}\left[S_{t,\delta(t)}(Z)e(Z;\beta_0)g(X)|X'\theta_0\right]+o(t).$$

Therefore, for small $|t|$, $\{f_{t,\delta(t)}\}$ is a submodel in $\mathcal{P}_2$ with the required tangent s. Since this implies $\overline{\dot{\mathcal{P}}}_2\subset\dot{T}$, we conclude $\dot{T}=\dot{\mathcal{P}}_2=\overline{\dot{\mathcal{P}}}_2$.

(ii) It suffices to show that (a) $\Pi[h|\dot{\mathcal{P}}_2]\in\dot{\mathcal{P}}_2$ and (b) $h-\Pi[h|\dot{\mathcal{P}}_2]\perp\dot{\mathcal{P}}_2$. To show (a), notice that using the null restriction $\mathbf{E}[e(Z;\beta_0)|X]=0$, we have $\mathbf{E}(\Pi[h|\dot{\mathcal{P}}_2])=0$, and also by the definition of $R_h(X)$,

$$\begin{aligned}\mathbf{E}[e(Z;\beta_0)\Pi[h|\dot{\mathcal{P}}_2]|X] &= \mathbf{E}[e(Z;\beta_0)h(Z)|X]-R_h(X)\\ &= \frac{\mathbf{E}[\sigma^{-2}(X)e(Z;\beta_0)h(Z)|X'\theta_0]}{\mathbf{E}[\sigma^{-2}(X)|X'\theta_0]}\end{aligned}$$

whereas

$$\begin{aligned}\mathbf{E}[e(Z;\beta_0)\Pi[h|\dot{\mathcal{P}}_2]g(X)|X'\theta_0] &= \mathbf{E}[e(Z;\beta_0)h(Z)g(X)|X'\theta_0]-\mathbf{E}[R_h(X)g(X)|X'\theta_0]\\ &= \frac{\mathbf{E}[g(X)|X'\theta_0]\mathbf{E}[\sigma^{-2}(X)e(Z;\beta_0)h(Z)|X'\theta_0]}{\mathbf{E}[\sigma^{-2}(X)|X'\theta_0]}.\end{aligned}$$

Hence using $\mathbf{E}[g(X)|X'\theta_0]=1$, we obtain (a).

To show (b), notice that $\mathbf{E}[\sigma^{-2}(X)R_h(X)|X'\theta_0]=0$. Therefore, for $s\in\dot{\mathcal{P}}_2$

$$\begin{aligned}&\mathbf{E}[(h-\Pi[h|\dot{\mathcal{P}}_2])s(Z)]=\mathbf{E}[e(Z;\beta_0)\sigma^{-2}(X)R_h(X)s(Z)]=\mathbf{E}[\sigma^{-2}(X)R_h(X)\mathbf{E}\left[e(Z;\beta_0)s(Z)|X\right]]\\ =\;&\mathbf{E}\left[\sigma^{-2}(X)R_h(X)\mathbf{E}[e(Z;\beta_0)s(Z)g(X)|X'\theta_0]\right]=0.\end{aligned}$$

■

Notice that since for each $x \in \mathbf{R}^{d_X}$, $\lim_{|\varepsilon|\to\infty} \varepsilon f_{\varepsilon|X}(\varepsilon|x) = 0$, $\mathbf{E}[e(Z;\beta_0)s_0(Z)|X] = -1$. This provides the expression for the efficient score as

$$\ell_1^*(z) = \dot{\ell}_1 - \Pi[\dot{\ell}_1|\dot{\mathcal{P}}_2] = e(z,\beta_0)\sigma^{-2}(x)b^*(x)$$

where $b^*(x) = R_{b_0}(x)$.

It is worth noting that when we know θ_0, the tangent space becomes

$$\dot{\mathcal{P}}_2 = \{s \in \tilde{L}_2(P) : \mathbf{E}[e(Z;\beta_0)s(Z)|X] = \mathbf{E}[e(Z;\beta_0)s(Z)|X'\theta_0] \text{ a.s.}\},$$

which is smaller than the previous one when we do not know θ_0. However, the projection remains the same. Therefore, the estimation of θ_0 is ancillary to the testing problem in the sense that the semiparametric efficiency bound for $c = 0$ does not change due to the lack of the knowledge of θ_0. As mentioned in the main text, our test statistic achieves this efficiency bound via the reparametrization of $\mu(X'\theta)$ into $\mu(F_\theta(X'\theta))$. (See Bickel (1982) and Cox and Reid (1987).)

7.2 Appendix B: Mathematical Proofs of the Main Results

In this subsection, we provide the proofs of the main results. The notations are as in the main text. For this section and Appendix C below, we use the notation C to denote an absolute constant which can take different values in different places. Recall that $Sb(x) = b(x) - \mathbf{E}[b(X)|X'\theta_0 = x'\theta_0]$. For an estimator $\hat{b}$ of b, the notation $S\hat{b}(x)$ means $Sb(x)$ with b replaced by $\hat{b}$. Hence the randomness of $\hat{b}(\cdot)$ does not interfere with the conditional expectation in the operator of S.

Proof of Lemma 1: By (2) and (3), it suffices to consider $\tilde{m}_P(x) \triangleq \mu_P(x'\theta_P)$ for m's in $\mathcal{G}(P;r_n)$. Consider the following alternatives $P_n \in \mathcal{P}(r_n)$ such that

$$m_{P_n}(x) = \tilde{m}_{P_n}(x) + w_{P_n}(x)$$

where $||w_{P_n}||_{M_{P_n}} = r_n$ and hence $m_{P_n} \in \mathcal{G}(P_n;r_n)$, and we can decompose

$$w_{P_n}(x) = c_{1n}b_{P_n}(x) + c_{2n}w_{1P_n}(x)$$

where $\mathbf{E}_{P_n}[b_{P_n}(X)w_{1P_n}(X)] = 0$ and $\mathbf{E}_{P_n}[w_{1P_n}^2(X)] > 0$ and c_{1n} and c_{2n} are constants. Note that by (10),

$$||w_{P_n}||_{M_{P_n}} = |c_{1n}||M_{P_n}b_{P_n}| = |c_{1n}|\mathbf{E}_{P_n}[b_{P_n}^2(X)].$$

This implies that $|c_{1n}| = r_n/\mathbf{E}_{P_n}[b_{P_n}^2(X)]$ and that $||m_{P_n} - \tilde{m}_{P_n}||_{M_{P_n}} = |c_{1n}|\mathbf{E}_{P_n}[b_{P_n}^2(X)]$.

For the proof of Lemma 1, it suffices to show that $c_{2n} = 0$. To the contrary, suppose that $|c_{2n}| > 0$. Choose P_n' such that $m_{P_n'}(x) = \tilde{m}_{P_n'}(x) + c_{1n}b_{P_n'}(x)$ where $\tilde{m}_{P_n'}(x) = \mu_{P_n'}(x'\theta_{P_n'})$ and

$\mathbf{E}_{P_n'}[b_{P_n'}^2(X)] = \mathbf{E}_{P_n}[b_{P_n}^2(X)]$. For example, we can choose P_n' under which it holds that

$$Y = \mu_{P_n'}(X'\theta_{P_n'}) + c_{1n} b_{P_n'}(X) + \varepsilon$$

where $\mathbf{E}_{P_n'}[\varepsilon|X] = 0$ and the marginal distribution of X under P_n' is the same as that under P_n. Then $\tilde{m}_{P_n'} \in \mathcal{G}(P_n'; r_n)$ because

$$||m_{P_n'} - \tilde{m}_{P_n'}||_{M_{P_n'}} = |c_{1n}||M_{P_n'} b_{P_n'}| = |c_{1n}|\mathbf{E}_{P_n'}[b_{P_n'}^2(X)] = |c_{1n}|\mathbf{E}_{P_n}[b_{P_n}^2(X)] = r_n.$$

Hence

$$\begin{aligned} ||m_{P_n'} - \tilde{m}_{P_n'}||_{2,P_n'} &= |c_{1n}|\mathbf{E}_{P_n'}[b_{P_n'}^2(X)] = |c_{1n}|\mathbf{E}_{P_n}[b_{P_n}^2(X)] \\ &< |c_{1n}|\mathbf{E}_{P_n}[b_{P_n}^2(X)] + |c_{2n}|\mathbf{E}_{P_n}[w_{1P_n}^2(X)] = ||m_{P_n} - \tilde{m}_{P_n}||_{2,P_n}. \end{aligned}$$

Therefore, $P_n \notin \mathcal{P}(r_n)$ leading to a contradiction.■

Proof of Theorem 1 : (i) Define $\mathcal{W}_n$ to be a shrinking neighborhood of b^*/σ^2 in $\mathcal{W}$ such that

$$\mathcal{W}_n = \left\{ \tilde{b} \in \mathcal{W} : \sup_{x \in \mathbf{R}^{d_X}} \left| \frac{b^*(x)}{\sigma^2(x)} - \tilde{b}^*(x) \right| < \delta_n \right\},$$

with $\delta_n = Cn^{-1/4} \to 0$ and $P\left\{ \hat{b}^*/\hat{\sigma}^2 \in \mathcal{W}_n \right\} \to 1$ as $n \to \infty$. By Assumption 3(i), it follows that

$$\begin{aligned} & P\left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{e(Z_i;\hat{\beta})\hat{b}^*(X_i)}{\hat{\sigma}^2(X_i)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n e(Z_i;\beta_0) S\left(\frac{\hat{b}^*}{\hat{\sigma}^2}\right)(X_i) \right| > \varepsilon \right\} \\ \leq\ & P\left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{e(Z_i;\hat{\beta})\hat{b}^*(X_i)}{\hat{\sigma}^2(X_i)} - \frac{1}{\sqrt{n}} \sum_{i=1}^n e(Z_i;\beta_0) S\left(\frac{\hat{b}^*}{\hat{\sigma}^2}\right)(X_i) \right| > \varepsilon,\ \frac{\hat{b}^*}{\hat{\sigma}^2} \in \mathcal{W}_n \right\} + P\left\{ \frac{\hat{b}^*}{\hat{\sigma}^2} \notin \mathcal{W}_n \right\} \\ \leq\ & P\left\{ \sup_{\tilde{b} \in \mathcal{W}_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e(Z_i;\hat{\beta})\tilde{b}(X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n e(Z_i;\beta_0) S\tilde{b}(X_i) \right| > \varepsilon \right\} + o(1). \end{aligned}$$

Now, by Lemma 1U below, the first probability is $o(1)$. On the other hand, the process

$$V_n(\tilde{b}) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n e(Z_i;\beta_0) S\tilde{b}(X_i)$$

is mean-zero under the null hypothesis since $\mathbf{E}\left[e(Z_i;\beta_0)|X_i\right] = 0$. We claim that the process $V_n(\tilde{b})$ is stochastically equicontinuous in $\tilde{b} \in \mathcal{W}_n$. (See e.g. Andrews (1994)). In order to see this, we need only to observe that the class $S\mathcal{W}_n \triangleq \left\{ S\tilde{b} : \tilde{b} \in \mathcal{W}_n \right\}$ has a finite bracketing integral entropy with a square integrable envelope. This latter condition follows due to the bracketing integral entropy condition for the class $\mathcal{W}_n$ because the operator S is a linear operator and for the envelope B_n of $\mathcal{W}_n$, $2B_n$ is an envelope of $S\mathcal{W}_n$. Therefore, we have

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e(Z_i;\beta_0) \left\{ S\left((\hat{b}^*/\hat{\sigma}^2)(X_i) \right) - S\left((b^*/\sigma^2)(X_i) \right) \right\} \right| = o_P(1).$$

By the central limit theorem, the process $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e(Z_i;\beta_0) S((b^*/\sigma^2)(X_i))$ converges in distribution to a centered normal variable with variance σ_{eb}^2. The result of (i) under the null hypothesis now follows by noting that $\hat{\sigma}_{eb}^2 = \sigma_{eb}^2 + o_P(1)$.

Consider the case of local alternatives. From the previous development, we have

$$T_n = \left(\frac{1}{\sqrt{n\sigma_{eb}^2}} \sum_{i=1}^{n} e(Z_i;\beta_0) S_\sigma \left(b^*(X_i) \right) \right)^2 + o_P(1).$$

Note that the derivation of the above did not rest on whether we are under the null hypothesis or not. Now, under the local alternatives, we have

$$\begin{aligned} \frac{1}{\sqrt{n\sigma_{eb}^2}} \sum_{i=1}^{n} e(Z_i;\beta_0)(S_\sigma b^*)(X_i) &= \frac{1}{\sqrt{n\sigma_{eb}^2}} \sum_{i=1}^{n} (Y_i - \mu_0(X_i'\theta_0))(S_\sigma b^*)(X_i) \\ &= \frac{1}{n\sigma_{eb}} \sum_{i=1}^{n} a(X_i)(S_\sigma b^*)(X_i) + \frac{1}{\sqrt{n\sigma_{eb}^2}} \sum_{i=1}^{n} \varepsilon_i (S_\sigma b^*)(X_i). \end{aligned}$$

Since we can apply the law of large numbers and the central limit theorem, the above is asymptotically normal with mean equal to $\mathbf{E}\left[a(X) S_\sigma b^*(X)\right]/\sigma_{eb}$ and variance one. The expectation in the mean is equal to $\langle S_\sigma a, b_0 \rangle_\sigma$.[12] By the assumption of the consistency of $\hat{\sigma}_{eb}^2$ and Slutsky's lemma, we obtain the wanted result.

(ii) The result follows by Lemma A1 above and Corollary 2 of CHS.■

7.3 Appendix C: A Uniform Representation of Empirical Processes involving a Conditional Mean Estimator

In this subsection, we present a general uniform representation of empirical processes that contain a series-based conditional mean estimator. The result immediately implies the uniform representation in Assumption 3(i)(c). Notations introduced here are self-contained for this subsection and have no association with those in the main text unless otherwise stated. Let $(S_i)_{i=1}^n \triangleq (Y_i, X_i, Z_i)_{i=1}^n$ be an i.i.d. random sample of (possibly overlapping) vectors from P. Let Λ be a class of real-valued functions whose generic element we denote by λ. Let $F_\lambda(\cdot)$ and $F_0(\cdot)$ be the distribution functions of $\lambda(X_i)$ and $\lambda_0(X_i)$, and let $F_{n,\lambda,i}$ and $F_{n,i}$ be the empirical distribution functions

[12] To see this, observe that

$$\langle a, S(b^*/\sigma^2) \rangle = \langle Sa, b^*/\sigma^2 \rangle = \langle Sa, b^* \rangle_\sigma = \langle Sa, S_\sigma b_0 \rangle_\sigma = \langle S_\sigma S a, b_0 \rangle_\sigma = \langle S_\sigma a, b_0 \rangle_\sigma.$$

For the above derivation, we used the fact that S_σ and S are self-adjoint with respect to $\langle \cdot, \cdot \rangle_\sigma$ and $\langle \cdot, \cdot \rangle$ respectively.

of $\{\lambda(X_j)\}_{j=1,j\neq i}^n$ and $\{\lambda_0(X_j)\}_{j=1,j\neq i}^n$, i.e., $F_{n,\lambda,i}(\bar{\lambda}) \triangleq \frac{1}{n}\sum_{j=1,j\neq i}^n 1\{\lambda(X_j) \leq \bar{\lambda}\}$ and $F_{n,i}(\bar{\lambda}) \triangleq \frac{1}{n}\sum_{j=1,j\neq i}^n 1\{\lambda_0(X_j) \leq \bar{\lambda}\}$.

We introduce quantile transforms

$$U_i \triangleq F_0(\lambda_0(X_i)), U_{n,i} \triangleq F_{n,i}(\lambda_0(X_i)), \text{ and } U_{n,\lambda,i} \triangleq F_{n,\lambda,i}(\lambda(X_i)),\ \lambda \in \Lambda,$$

and define

$$g_\kappa(u) \triangleq \mathbf{E}(\kappa(Y_i)|U_i = u) \text{ and } g_w(u) \triangleq \mathbf{E}(w(X_i)|U_i = u)$$

where κ and w belong to sets $\mathcal{K}$ and $\mathcal{W}$ of real-valued functions on $\mathbf{R}^{d_Y}$ and $\mathbf{R}^{d_X}$ respectively. For a vector ν of nonnegative integers, we define $|\cdot|_\nu : |g|_\nu = \sup_{\mu\leq\nu}\sup_z |D^\mu g(z)|$, where $D^\mu g(z) = (\partial^{|\mu|}/\partial z_1^{\mu_1}\cdots\partial z_d^{\mu_{d_z}})g(z)$ with d_z denoting the dimension of z. We approximate $g_\kappa(u)$ by $p^K(u)'\pi_\kappa$ using certain vectors $p^K(u)$ and π_κ. We have in mind the situation where λ_0 is not observed and is replaced by a uniformly consistent estimator $\hat{\lambda}$ such that $P\{\hat{\lambda} \in \Lambda\} \to 1$. For this, we introduce a series-based estimator indexed by $\lambda \in \Lambda$ as follows.

$$\hat{g}_{\kappa,\lambda}(u) \triangleq p^K(u)'\hat{\pi}_{\kappa,\lambda},\ \kappa \in \mathcal{K},\ \lambda \in \Lambda,$$

where $\hat{\pi}_{\kappa,\lambda} = [P_\lambda' P_\lambda]^{-1}P_\lambda' y_{\kappa,n}$ with $y_{\kappa,n}$ and P_λ defined by

$$y_{\kappa,n} \triangleq \begin{bmatrix} \kappa(Y_1) \\ \vdots \\ \kappa(Y_n) \end{bmatrix} \text{ and } P_\lambda \triangleq \begin{bmatrix} p^K(U_{n,\lambda,1})' \\ \vdots \\ p^K(U_{n,\lambda,n})' \end{bmatrix}.$$

Let $\hat{g}_{\kappa,\lambda,i}(u)$ be $\hat{g}_{\kappa,\lambda}(u)$ constructed without using the i-th data, $(\kappa(Y_i), U_{n,\lambda,i})$. We are interested in the asymptotic representation of the process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n w(Z_i)\left\{\kappa(Y_i) - \hat{g}_{\kappa,\lambda,i}(U_{n,\lambda,i})\right\}, (w,\kappa,\lambda) \in \mathcal{W}\times\mathcal{K}\times\Lambda_n$$

that is uniform over $(w,\kappa,\lambda) \in \mathcal{W}\times\mathcal{K}\times\Lambda_n$, where Λ_n is specified below.

Without loss of generality, we assume that $\lambda(x) \in [0,1]$, $\lambda \in \Lambda$.[13] Let $l_\infty(\mathbf{R}^{d_X})$ be the space of uniformly bounded real functions on $\mathbf{R}^{d_X}$. For a given basis function vector p^K, we define $\zeta_{\nu,K} \triangleq |p^K|_\nu$. We introduce a sup norm $||\cdot||_\infty$ on Λ: $||\lambda - \lambda_0||_\infty = \sup_x |\lambda(x) - \lambda_0(x)|$, and choose a neighborhood Λ_n of λ_0 by $\Lambda_n = \{\lambda \in \Lambda : ||\lambda - \lambda_0||_\infty \leq Cn^{-b}\}$ for $b \in (1/4, 1/2]$ and for some constant $C > 0$. In applications, we may consider $\hat{\lambda} \in \Lambda_n$ with probability approaching one. The neighborhood Λ_n is allowed to shrink at a rate slower than $n^{-1/2}$, and hence it allows for the case when $\hat{\lambda}$ is a nonparametric estimator.

Assumption 1U : (i) $(Y_i, X_i, Z_i)_{i=1}^n$ *is a random sample from* P. (ii) *For classes* $\mathcal{K}$, Λ, *and* $\mathcal{W}$,

[13] Otherwise, this is fullfilled by redefining $\lambda_H(X) \triangleq H(\lambda(X))$ for a strictly increasing transform $H : \mathbf{R} \to [0,1]$ and following the proof in the same manner. Note that $U_{n,\lambda,i}$ and U_i remains intact after this transform.

there exist b_1, b_2 *and* b_3 *such that* $b_1, b_3 \in [0,2)$, $b_2 \in [0,1)$, *and* $b_1(1-1/p) < 1$, $p \geq 4$,

$$\begin{aligned}\log N_{[]}(\varepsilon, \mathcal{K}, ||\cdot||_{p,P}) &< C\varepsilon^{-b_1},\ \log N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty) < C\varepsilon^{-b_2},\ \textit{and} \\ \log N_{[]}(\varepsilon, \mathcal{W}, ||\cdot||_{p,P}) &< C\varepsilon^{-b_3},\end{aligned}$$

and envelopes $\tilde{\kappa}$ *and* $\tilde{w}$ *for* $\mathcal{K}$ *and* $\mathcal{W}$ *satisfy that* $\mathbf{E}[|\tilde{\kappa}(Y)|^p|X] < \infty$ *and* $\mathbf{E}[|\tilde{w}(Z)|^p|X] < \infty$, $\mathbf{P}_X$-a.s., *for some* $\varepsilon > 0$.
(ii)(a) *For each* $\lambda \in \Lambda_n$, $\lambda(X)$ *is a continuous random variable and* (b) *for some* $C > 0$,

$$\sup_{\lambda\in\Lambda_n}\sup_{\bar{\lambda}\in[0,1]}\left|F_\lambda(\bar{\lambda}+\delta) - F_\lambda(\bar{\lambda}-\delta)\right| < C\delta,\ \textit{for all}\ \delta > 0.$$

(iii) *There exists* $C > 0$ *such that for each* $u \in \mathcal{U} \triangleq \{F_\lambda \circ \lambda : \lambda \in \Lambda_n\}$, *the conditional density function* $f_u(y,x|\bar{u})$ *of* (Y,X) *given* $u(X) = \bar{u}$ *satisfies that for all* $(y,x) \in \mathbf{R}^{d_Y+d_X}$ *and for all* $\bar{u} \in [0,1]$,

$$\sup_{\bar{u}_1\in[0,1]:|\bar{u}-\bar{u}_1|<\delta} |f_u(y,x|\bar{u}) - f_u(y,x|\bar{u}_1)| \leq \varphi_u(y,x)\delta,$$

where $\varphi_u(\cdot,\cdot)$ *is a real function that satisfies* $\sup_{x\in\mathcal{S}_X}\int|\tilde{\kappa}(y)|\varphi_u(y,x)dy < C$ *and* $\int\varphi_u(y,x)dx < Cf_Y(y)$ *with* $f_Y(\cdot)$ *denoting the density of* Y.

Assumption 2U : *For every* K, *there is a nonsingular constant matrix* B *such that for* $P^K(u) \triangleq Bp^K(u)$ *the following is satisfied.*
(i) *There exists* $C_1 > 0$ *such that from a sufficiently large* n *on,*

$$0 < C_1 < \lambda_{\min}\left(\int_0^1 P^K(u)P^K(u)'du\right).$$

(ii) *There exist* d_1 *and* $d_2 > 0$ *such that* (a) *there exist classes of vectors in* $\mathbf{R}^K$, $\{\pi_\kappa : \kappa \in \mathcal{K}\}$ *and* $\{\pi_w : w \in \mathcal{W}\}$, *such that for each* $(w,\kappa) \in \mathcal{W}\times\mathcal{K}$,

$$\begin{aligned}\sup_{(\kappa,\bar{u})\in\mathcal{K}\times[0,1]}\left|P^K(\bar{u})'\pi_\kappa - g_\kappa(\bar{u})\right| &= O(K^{-d_1})\ \textit{and} \\ \sup_{(w,\bar{u})\in\mathcal{W}\times[0,1]}\left|P^K(\bar{u})'\pi_w - g_w(\bar{u})\right| &= O(K^{-d_2}),\end{aligned}$$

and (b) *for each* $\bar{u} \in [0,1]$ *there exist classes of vectors in* $\mathbf{R}^K$, $\{\pi_{g,\bar{u}} : g \in \mathcal{G}_{\bar{u}}\}$, $\mathcal{G}_{\bar{u}} \triangleq \{D^1 g_\kappa(\cdot)1\{\bar{u} \leq \cdot\} : \kappa \in \mathcal{K}\}$, *such that*

$$\sup_{(\kappa,\bar{u})\in\mathcal{K}\times[0,1]}\ \sup_{g\in\mathcal{G}_{\kappa,\bar{u}}}\left(\int_0^1\left|P^K(u)'\pi_{g,\bar{u}} - g(u)\right|^p du\right)^{1/p} = O(K^{-d_1}).$$

(iii) *For* d_1 *and* d_2 *in* (ii), $\sqrt{n}\zeta_{0,K}^2 K^{-d_1} = o(1)$ *and* $\sqrt{n}K^{-d_2} = o(1)$.

(iv) *For b in the definition of Λ_n and p in* Assumption 1U(ii),

$$\begin{aligned} n^{1/2-2b}\zeta_{0,K}^3\zeta_{2,K} &= o(1),\ n^{-1/2+1/p}K^{1-1/p}\zeta_{0,K}^2 = o(1) \text{ and} \\ n^{-b}\zeta_{0,K}\{\sqrt{\zeta_{0,K}\zeta_{2,K}}+\zeta_{1,K}\} &= o(1). \end{aligned}$$

We define processes $\hat{\Delta}_{1n}(w,\kappa,\lambda)$ and $\Delta_{2n}(w,\kappa)$ as follows.

$$\begin{aligned} \hat{\Delta}_{1n}(w,\kappa,\lambda) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n w(Z_i)\left\{\kappa(Y_i)-\hat{g}_{\kappa,\lambda,i}(U_{n,\lambda,i})\right\}, \text{ and} \\ \Delta_{2n}(w,\kappa) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n\{w(Z_i)-g_w(U_i)\}\{\kappa(Y_i)-g_\kappa(U_i)\}. \end{aligned}$$

The following lemma establishes the uniform asymptotic equivalence of $\hat{\Delta}_{1n}(w,\kappa,\lambda)$ and $\Delta_{2n}(w,\kappa)$.

Lemma 1U : *Suppose that* Assumptions 1U-2U. *Then we have*

$$\sup_{(w,\kappa,\lambda)\in\mathcal{W}\times\mathcal{K}\times\Lambda_n}|\hat{\Delta}_{1n}(w,\kappa,\lambda)-\Delta_{2n}(w,\kappa)| = o_P(1).$$

It is worth noting that when we replace λ in $\hat{\Delta}_{1n}(w,\kappa,\lambda)$ by λ_0 so that the supremum is only over $(w,\kappa)\in\mathcal{W}\times\mathcal{K}$, we obtain the same result. This implies that the estimation error in $\hat{\lambda}$ plays no role in determining the uniform representation. This is because we use $F_\lambda(\lambda(X))$ as a conditioning variable, rather than $\lambda(X)$. By doing so, the estimation error of $\hat{\lambda}$ is cancelled out by the estimation error additionally introduced by the normalization of $\lambda(X)$ by $F_\lambda(\cdot)$. This is a generalization of a point made by Stute and Zhu (2005) who found this phenomenon in the context of kernel estimation.

Now, we show how (25) in Assumption 3(i)(c) can be derived from this result. The notations X_i and $X_i'\theta_0$ there correspond to Z_i and $\lambda_0(X_i)$ in Lemma 1U respectively. Under the lower-level assumptions in Lemma 1U translated into the environment in Assumption 3(i)(c), we can write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n w(X_i)(Y_i-\hat{g}(F_{n,\theta}(X_i'\hat{\theta});\hat{\theta})) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(w(X_i)-\mathbf{E}[w(X_i)|U_i]\right)\left(Y_i-\mathbf{E}[Y_i|U_i]\right)+o_P(1)$$

uniformly over θ such that $||\theta-\theta_0|| = O(n^{-1/2})$ and over $w\in\mathcal{W}$ in Assumption 1U(i).

Note that under the local alternatives, we replace Y_i with $a(X_i)/\sqrt{n}+\mu_0(X_i'\theta_0)+\varepsilon_i$. From $\mu_0(X_i'\theta_0)-\mathbf{E}\left[\mu_0(X_i'\theta_0)|U_i\right]=0$, we find that the statement follows once we show that the sum

$$\frac{1}{n}\sum_{i=1}^n\left(w(X_i)-\mathbf{E}[w(X_i)|U_i]\right)\mathbf{E}[a(X_i)|U_i]$$

is uniformly $o_P(1)$ over $w\in\mathcal{W}$. Since the sum is a mean-zero process, it suffices to show that the functions in the summand as indexed by $w\in\mathcal{W}$ belong to a Glivenko-Cantelli class. This latter fact immediately follows from the bracketing entropy condition for the class $\mathcal{W}$.

Proof of Lemma 1U : Fix an arbitrarily small number $\sigma > 0$ and define $M_\sigma \triangleq 1/\sigma$. As in Song (2006), we rotate the vector p^K by a matrix. Define the matrix $B_K(b) \triangleq p^K(u_b)p^K(u_b)'$ where u_b achieves the supremum of $\{b'\{p^K(\bar{u})p^K(\bar{u})'\}^{M_\sigma} b : \bar{u} \in [0,1]\}$ and let b^* be a maximizer of $b' B_K^{M_\sigma}(b) b$ over $b \in S_K$, where $S_K \triangleq \{b \in \mathbf{R}^K : ||b|| = 1\}$, and apply the spectral decomposition to the matrix

$$B_K^{M_\sigma}(b^*) + C_1 I = \tilde{B}_K \Omega \tilde{B}_K',$$

for some $C_1 > 0$ and finally rotate p^K to obtain $P^K \triangleq \tilde{B}_K' p^K / \{\lambda_{\max}(\Omega)\}^{\frac{1}{2M_\sigma}}$, but we use the same notation p^K for this rotated vector P^K.

We introduce some notations. Let

$$\begin{aligned}
\hat{Q}_{n,\lambda} &\triangleq P_{n,\lambda}' P_{n,\lambda}/n, \;\; Q \triangleq \int_0^1 p^K(u)p^K(u)'du, \\
\hat{Q}_n &\triangleq \frac{1}{n}\sum_{i=1}^n p^K(U_{n,i})p^K(U_{n,i})', \text{and } \hat{Q} \triangleq \frac{1}{n}\sum_{i=1}^n p^K(U_i)p^K(U_i)'.
\end{aligned}$$

For brevity, define

$$\begin{aligned}
p_i^K &\triangleq p^K(U_i), \; p_{n,i}^K \triangleq p^K(U_{n,i}), \; p_{n,\lambda,i}^K \triangleq p^K(U_{n,\lambda,i}), \; w_i \triangleq w(Z_i), \; \kappa_i \triangleq \kappa(Y_i), \\
g_{w,i} &\triangleq g_w(U_i), \; g_{w,n,i} \triangleq g_{w,n}(U_{n,i}), \; g_{\kappa,i} \triangleq g_\kappa(U_i), \text{ and } g_{\kappa,n,i} \triangleq g_{\kappa,n}(U_{n,i}),
\end{aligned}$$

where $g_{\kappa,n}(u) \triangleq \mathbf{E}\left[\kappa(Y_i)|U_{n,i} = u\right]$ and $g_{w,n}(u) \triangleq \mathbf{E}\left[w(Z_i)|U_{n,i} = u\right]$.

Note that $||F_\lambda \circ \lambda - F_{\lambda_0} \circ \lambda_0||_\infty$ is bounded by

$$F_{\lambda_0}(\lambda_0(x) + 2||\lambda - \lambda_0||_\infty) - F_{\lambda_0}(\lambda_0(x) - 2||\lambda - \lambda_0||_\infty) \le C||\lambda - \lambda_0||_\infty \tag{35}$$

by Assumption 1U(ii)(a). This implies that

$$\log N_{[]}(C\varepsilon, \mathcal{U}, ||\cdot||_\infty) \le \log N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty). \tag{36}$$

where $\mathcal{U}$ is as defined in Assumption 1U(iii). Also observe that $\sup_{\lambda\in\Lambda} ||F_{n,\lambda} - F_\lambda||_\infty = O_P(n^{-1/2})$. The last statement follows because Λ is P-Donsker by Assumption 1U(i).[14] Furthermore, observe that[15]

$$P\left\{\inf_{\lambda\in\Lambda_n}\lambda_{\min}(\hat{Q}_{n,\lambda})>C_1/2\right\}\to 1,$$

which makes it suffice to deal with every term multiplied by $1_n \triangleq 1\{\inf_{\lambda\in\Lambda_n}\lambda_{\min}(\hat{Q}_{n,\lambda})>C_1/2\}$. For simplicity, we suppress this from the notations.

Write $\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i\hat{g}_{\kappa,\lambda,i}(U_{n,\lambda,i})$ as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_i p_{n,\lambda,i}^{K\prime}\hat{Q}_{n,\lambda}^{-1}p_{n,\lambda,j}^K\kappa_j = tr\left\{\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_i\kappa_j p_{n,\lambda,j}^K p_{n,\lambda,i}^{K\prime}\hat{Q}_{n,\lambda}^{-1}\right\}.$$

For the term on the right-hand side, we show the following:

(UA) : $\sup_{w,\kappa,\lambda}\left|\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_i\kappa_j tr(\{p_{n,\lambda,j}^K p_{n,\lambda,i}^{K\prime}\hat{Q}_{n,\lambda}^{-1}-p_{n,j}^K p_{n,i}^{K\prime}\hat{Q}_n^{-1}\})\right| = o_P(1).$

Then, observe that $\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i(\hat{g}_{\kappa,\lambda,i}(U_{n,\lambda,i})-g_{\kappa,i})$ is equal to

$$\begin{aligned}
& \frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_i p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^K\kappa_j - \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i g_{\kappa,i}+o_P(1)\\
= & \frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_i p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^K(\kappa_j-g_{\kappa,j})\\
& +\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i\left\{p_{n,i}^{K\prime}\frac{1}{n}\sum_{j=1,j\neq i}^n\hat{Q}_n^{-1}p_{n,j}^K g_{\kappa,j}-g_{\kappa,i}\right\}+o_P(1)\\
= & A_{1n}(w,\kappa,\lambda)+A_{2n}(w,\kappa,\lambda),\ \text{say.}
\end{aligned}$$

For the second term, we show the following:

(UB) : $\sup_{w,\kappa,\lambda}|A_{2n}(w,\kappa,\lambda)| = o_P(1),$

[14] This immediately follows from the fact that the class $\mathcal{I}\triangleq\{1\{\lambda(\cdot)\leq\bar{\lambda}\}:(\lambda,\bar{\lambda})\in\Lambda\times[0,1]\}$ is P-Donsker. This latter fact follows from the local uniform L_2-continuity of the functions in the class, i.e.,

$$\left\{\mathbf{E}\left[\sup_{\lambda,\bar{\lambda}:||\lambda-\lambda_1||_\infty<\delta,|\bar{\lambda}-\bar{\lambda}_1|<\delta}\left|1\{\lambda_1(\cdot)\leq\bar{\lambda}_1\}-1\{\lambda(\cdot)\leq\bar{\lambda}\}\right|^2\right]\right\}^{1/2}\leq C\delta^{1/2}.$$

The above implies that for all $\varepsilon>0$,

$$\log N_{[]}(\varepsilon,\mathcal{I},||\cdot||_2)\leq\log N_{[]}((C\varepsilon)^2,\Lambda,||\cdot||_\infty)+\log N_{[]}((C\varepsilon)^2,[0,1],||\cdot||)\leq C\varepsilon^{-2b_2}.$$

That $\mathcal{I}$ is P-Donsker follows by the condition $b_2<1$ in Assumption 1U(i).

[15] Note that

$$\left|\lambda_{\min}(\hat{Q}_{n,\lambda})-\lambda_{\min}(Q)\right|\leq\sup_{b\in S_K}\left|b'(\hat{Q}_{n,\lambda}-Q)b\right|\leq\zeta_{0,K}\zeta_{1,K}||F_{n,\lambda}-F_\lambda||_\infty=O_P(n^{-1/2}\zeta_{0,K}\zeta_{1,K}).$$

The last term is $o_P(1)$ by Assumption 2U(iv).

leaving us with $A_{1n}(w,\kappa,\lambda)$ to deal with. For this term, we show the following.

(UC) : $\sup_{w,\kappa,\lambda}\left|A_{1n}(w,\kappa,\lambda)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{w,i}(\kappa_i-g_{\kappa,i})\right|=o_P(1)$.

By collecting the results of (UB) and (UC), we conclude that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i(\hat{g}_{\kappa,\lambda}(U_{n,\lambda,i})-g_{\kappa,i})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{w,i}(\kappa_i-g_{\kappa,i})+o_P(1).$$

This result implies that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i\left\{\kappa_i-\hat{g}_{\kappa,\lambda}(U_{n,\lambda,i})\right\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i\left\{\kappa_i-g_{\kappa,i}\right\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i\left\{g_{\kappa,i}-\hat{g}_{\kappa,\lambda}(U_{n,\lambda,i})\right\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i\left\{\kappa_i-g_{\kappa,i}\right\}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{w,i}(\kappa_i-g_{\kappa,i})+o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{w_i-g_{w,i}\}\left\{\kappa_i-g_{\kappa,i}\right\}+o_P(1),\end{aligned}$$

completing the proof.

(Proof of UA) : Write $tr\left\{\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_i\kappa_j\left(p_{n,\lambda,j}^{K}p_{n,\lambda,i}^{K\prime}\hat{Q}_{n,\lambda}^{-1}-p_{n,j}^{K}p_{n,i}^{K\prime}\hat{Q}_{n}^{-1}\right)\right\}$ as

$$\begin{aligned}& tr\left\{\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_i\kappa_j p_{n,\lambda,j}^{K}p_{n,\lambda,i}^{K\prime}(\hat{Q}_{n,\lambda}^{-1}-\hat{Q}_{n}^{-1})\right\} \qquad (37)\\ & +tr\left\{\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_i\kappa_j\left(p_{n,\lambda,j}^{K}p_{n,\lambda,i}^{K\prime}-p_{n,j}^{K}p_{n,i}^{K\prime}\right)\hat{Q}_{n}^{-1}\right\}.\end{aligned}$$

Consider the leading term in the above, which we write as

$$\begin{aligned}& \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_i\kappa_j p_{n,\lambda,i}^{K\prime}\hat{Q}_{n,\lambda}^{-1}(\hat{Q}_n-\hat{Q}_{n,\lambda})\hat{Q}_{n}^{-1}p_{n,\lambda,j}^{K}\\ \leq\ & \frac{\zeta_{0,K}^2}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|w_i||\kappa_j|b_{n,\lambda,i}'(\hat{Q}_n-\hat{Q}_{n,\lambda})b_{n,\lambda,j},\ \text{ where } b_{n,\lambda,j}\triangleq\hat{Q}_{n}^{-1}p_{n,\lambda,j}^{K}/\left\|\hat{Q}_{n}^{-1}p_{n,\lambda,j}^{K}\right\|,\\ \leq\ & \frac{\zeta_{0,K}^2}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|w_i||\kappa_j|\sup_{(b,\lambda)\in S_K\times\Lambda_n}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{(b'p^K(U_{n,k}))^2-(b'p^K(U_{n,\lambda,k}))^2\right\}\right|.\end{aligned}$$

Then we show that

(UA1) : $\sup_{(b,\lambda)\in S_K\times\Lambda_n}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{(b'p^K(U_{n,k}))^2-(b'p^K(U_{n,\lambda,k}))^2\right\}\right|=o_P(\zeta_{0,K}^{-2})$

Since $\sup_{w,\kappa}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|w_i||\kappa_j|\leq\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|\tilde{w}(Z_i)||\tilde{\kappa}(Y_j)|=O_P(1)$, by the law of large numbers, we deduce that the first term in (37) is equal to $o_P(1)$.

Now, let us deal with the second term in (37), which we bound by

$$\sup_{b\in S_K}\left(\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j b'\left(p_{n,\lambda,j}^K p_{n,\lambda,i}^{K\prime} - p_{n,j}^K p_{n,i}^{K\prime}\right)b\right) tr\left(\hat{Q}_n^{-1}\right).$$

The double sum in the above is written as

$$\begin{aligned}
& \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j b'\left(\{p_{n,\lambda,j}^K - p_{n,j}^K\}p_{n,\lambda,i}^{K\prime}\right)b + \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j b'\left(p_{n,j}^K\{p_{n,\lambda,i}^{K\prime} - p_{n,i}^{K\prime}\}\right)b \\
\triangleq\ & B_{1n}(w,\kappa,\lambda) + B_{2n}(w,\kappa,\lambda) \text{ say.}
\end{aligned}$$

Define $\psi_{1b}(u_1,u_2)\triangleq b'p^K(u_1)D^1p^K(u_2)'b$ and $\psi_{2b}(u_1,u_2)\triangleq b'p^K(u_1)D^2p^K(u_2)'b$. The first term is expanded as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j\psi_{1b}(U_i,U_j)\{U_{n,\lambda,j}-U_{n,j}\} + r_{0n}(w,\kappa,\lambda) + r_{1n}(w,\kappa,\lambda) \tag{38}$$

where

$$\begin{aligned}
r_{0n}(w,\kappa,\lambda) &\triangleq \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j\left\{\psi_{1b}(U_{n,\lambda,i},U_{n,j}) - \psi_{1b}(U_i,U_j)\right\}\left\{U_{n,\lambda,j}-U_{n,j}\right\} \\
r_{1n}(w,\kappa,\lambda) &\triangleq \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\kappa_j\psi_{2b}(U_{n,\lambda,i},U_{n,\lambda,j}^*)\{U_{n,\lambda,j}-U_{n,j}\}^2, \text{ and}
\end{aligned}$$

$U_{n,\lambda,j}^*$ lies on the line segment between $U_{n,j}$ and $U_{n,\lambda,j}$. Let us investigate $r_{1n}(w,\kappa,\lambda)$. We can deal with $r_{0n}(w,\kappa,\lambda)$ similarly by expanding $\psi_{1b}(U_{n,\lambda,i},U_{n,j})-\psi_{1b}(U_i,U_j)$ around (U_i,U_j). Note that

$$\begin{aligned}
|r_{1n}(w,\kappa,\lambda)| &\leq \frac{\zeta_{2,K}\zeta_{0,K}}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|\tilde{w}(Z_i)||\tilde{\kappa}(Y_j)|\left(U_{n,\lambda,j}-U_{\lambda,j}-(U_{n,j}-U_j)+U_{\lambda,j}-U_j\right)^2 \\
&= O_P(n^{1/2-2b}\zeta_{0,K}\zeta_{2,K}) = o_P(\zeta_{0,K}^{-2})
\end{aligned}$$

by (35) and Assumption 2U(iv).

By Lemma UA below, the first term in (38) is written as

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i b'p_i^K\mathbf{E}\left[\{\kappa_j - g_{\kappa,j}\}D^1p_j^K\{U_{\lambda,j}-U_j\}\right] + o_P(K^{-1}) \qquad (39) \\
=\ & \left(\frac{1}{n}\sum_{i=1}^{n} w_i b'p_i^K\right)\mathbf{E}\left[\{\kappa_j - g_{\kappa,j}\}D^1p_j^K U_{\lambda,j}\right] + o_P(K^{-1})
\end{aligned}$$

Now, by Lemma A2(ii) of Song (2006),

$$\sup_{x\in\mathcal{S}_X}\left|\mathbf{E}\left[D^1p_j^K\kappa_j|U_j=F_{\lambda_0}(\lambda_0(x))\right]-\mathbf{E}\left[D^1p_j^K\kappa_j|U_{\lambda,j}=F_\lambda(\lambda(x))\right]\right|\leq C\zeta_{1,K}||F_\lambda\circ\lambda-F_{\lambda_0}\circ\lambda_0||_\infty, \tag{40}$$

Hence $\left|\mathbf{E}\left[\{\kappa_j-g_{\kappa,j}\}D^1p_j^KU_{\lambda,j}\right]\right|$ is equal to

$$\begin{aligned}&\left|\mathbf{E}\left[\left\{\mathbf{E}\left[D^1p_j^K\kappa_j|U_{\lambda,j}\right]-\mathbf{E}\left[\mathbf{E}\left[D^1p_j^K\kappa_j|U_j\right]|U_{\lambda,j}\right]\right\}U_{\lambda,j}\right]\right|\\ =\;&\left|\mathbf{E}\left[\left\{\mathbf{E}\left[\mathbf{E}\left[D^1p_j^K\kappa_j|U_{\lambda,j}\right]-\mathbf{E}\left[D^1p_j^K\kappa_j|U_j\right]|U_{\lambda,j}\right]\right\}U_{\lambda,j}\right]\right|\leq C\zeta_{1,K}||\lambda-\lambda_0||_\infty=O_P(n^{-b}\zeta_{1,K}),\end{aligned}$$

using (40) and (35). The last term is $o_P(K^{-1})$ by Assumption 2U(iv). As for the term $\frac{1}{n}\sum_{i=1}^n w_ib'p_i^K$ in (39), let $\mathcal{J}_1\triangleq\{w(\cdot)b'p^K(\cdot):(w,b)\in\mathcal{W}\times S_K\}$ and define

$$J_1(z,u)\triangleq\tilde{w}(z)\left[tr\left(\{p^K(u)p^K(u)\}^{M_\sigma/2}\right)\right]^{1/M_\sigma}.$$

Then as in the proof of Theorem 1L in Song (2006) (see the proof of (B)), we have $||J_1||_{2,P}\leq CK^{\sigma/2}$ and hence by using the maximal inequality (e.g. Pollard (1989)[16]) and using Lemma UB below.

$$\begin{aligned}\mathbf{E}\left[\sup_{(w,b)\in\mathcal{W}\times S_K}\left|\frac{1}{n}\sum_{i=1}^n w_ib'p_i^K\right|\right]&\leq\frac{1}{\sqrt{n}}\int_0^{CK^{\sigma/2}}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{J}_1,||\cdot||_{2,P})}d\varepsilon\\&\leq n^{-1/2}\left\{K^{(1-b_3/2)(\sigma/2)}+\sqrt{K\log(K^{\sigma/2})}\right\}=o_P(1).\end{aligned}$$

Hence we conclude $\sup_{w,\kappa,\lambda}|B_{1n}(w,\kappa,\lambda)|=o_P(1)$. We can similarly write $B_{2n}(w,\kappa,\lambda)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\kappa_ib'p_j^K\mathbf{E}\left[\{w_i-g_{w,i}\}D^1p_i^K\{U_{\lambda,i}-U_i\}\right]+o_P(1)$$

and show that it is $o_P(1)$ exactly in the same manner as before.

(Proof of UB) : First write $\frac{1}{n\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_ip_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^Kg_{\kappa,j}-\frac{1}{\sqrt{n}}\sum_{i=1}^n w_ig_{\kappa,i}$ as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n w_ip_{n,i}^{K\prime}\hat{Q}_n^{-1}\left\{\frac{1}{n}\sum_{j=1,j\neq i}^n p_{n,j}^Kg_{\kappa,j}-\frac{1}{n}\sum_{j=1,j\neq i}^n p_{n,j}^{K\prime}p_{n,j}^{K\prime}\pi_\kappa\right\}\\&+\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i\left\{p_{n,i}^{K\prime}\pi_\kappa-g_{\kappa,i}\right\}+o_P(1).\end{aligned} \tag{41}$$

[16] The maximal inequality and its proof there are replicated in van der Vaart (1996), Theorem A.2, p.2136.

By the mean-value expansion, the first term is written as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^{K}D^1 g_\kappa(U_{n,j})(U_j-U_{n,j}) \tag{42}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i p_{n,i}^{K\prime}\hat{Q}_n^{-1}\left\{\frac{1}{n}\sum_{j=1,j\neq i}^{n} p_{n,j}^{K}\{g_\kappa(U_{n,j})-p_{n,j}^{K\prime}\pi_\kappa\}\right\}+o_P(1),$$

following steps in the proof of (UA2). The second term in the above is bounded by

$$\left(\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}|\tilde{w}(Z_i)|\left|p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^{K}\right|\right)\left\|g_\kappa-p^{K\prime}\pi_\kappa\right\|_\infty.$$

It is not hard to show that the double sum in the parenthesis is $O_P(\sqrt{n}\zeta_{0,K}^2)$. Hence the second term in (42) is $O_P(\sqrt{n}\zeta_{0,K}^2K^{-d_1})=o_P(1)$ by Assumption 2U(iii). Similarly the second term in (41) is equal to

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i D^1 g_\kappa(U_{n,i})(U_i-U_{n,i})+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\{p_{n,i}^{K\prime}\pi_\kappa-g_\kappa(U_{n,i})\}+o_P(1)$$

and again, the second term is $o_P(1)$.

Hence the first two sums in (41) is equal to

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^{K}D^1 g_\kappa(U_{n,j})(U_j-U_{n,j}) \tag{43}$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i D^1 g_\kappa(U_{n,i})(U_i-U_{n,i})+o_P(1)$$
$$=\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i p_i^{K\prime}Q^{-1}p_j^{K}D^1 g_\kappa(U_j)(U_j-U_{n,j})-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i D^1 g_\kappa(U_i)(U_i-U_{n,i})+o_P(1).$$

The last equality is obtained by expanding terms around U_i and U_j and applying Assumption 2U(iv). We define $q_K(U_j,U_k;\kappa,b)\triangleq b'p^K(U_j)D^1g_\kappa(U_j)\{U_j-1\{U_k\leq U_j\}\}$ and bound the absolute value of the leading term by

$$\frac{1}{n}\sum_{i=1}^{n} w_i||p_i^{K\prime}Q^{-1}||\left(\sup_{(\kappa,b)\in\mathcal{K}\times S_K}\left|\frac{1}{n\sqrt{n}}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq j,i}^{n} q_K(U_j,U_k;\kappa,b)\right|\right).$$

We analyze the double sum in the parenthesis. First note that $\mathbf{E}\left[q_K(U_j,U_k;\kappa,b)|U_j\right]=0$. Applying

Hoeffding's decomposition to the double sum above, we write it as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1,k\neq j,i}^{n}\mathbf{E}\left[q_K(U_j,U_k;\kappa,b)|U_k\right]+r_{2n}(w,\kappa)$$

where $r_{2n}(w,\kappa)\triangleq\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1,k\neq j,i}^{n}\left(q_K(U_j,U_k;\kappa,b)-\mathbf{E}\left[q_K(U_j,U_k;\kappa,b)|U_k\right]\right)$ is a degenerate U-process. For this we show the following:

(UB1) : $\sup_{w,\kappa}|r_{2n}(w,\kappa)|=o_P(\zeta_{0,K}^{-1})$.

We can deal with the last term in (43) similarly, so that the terms in (43) are written as

$$\begin{aligned}&\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1,k\neq j,i}^{n}w_i\left\{p_i^{K\prime}Q^{-1}\mathbf{E}\left[p_j^K\left(D^1g_\kappa(U_j)\{1\{U_k\leq U_j\}-U_j\}-p_j^{K\prime}\pi(U_k)\right)|U_k\right]\right\} \quad (44)\\&+\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1,k\neq j,i}^{n}\mathbf{E}\left[w_i\left(p_i^{K\prime}\pi(U_k)-D^1g_\kappa(U_i)\{1\{U_k\leq U_i\}-U_i\}\right)|U_k\right]+o_P(1).\end{aligned}$$

By Cauchy-Schwarz inequality, $\left|\mathbf{E}\left[b'p_j^K\left(D^1g_\kappa(U_j)\{1\{U_k\leq U_j\}-U_j\}-p_j^{K\prime}\pi(U_k)\right)|U_k\right]\right|$ is bounded by

$$\sqrt{\mathbf{E}\left[b'p_j^Kp_j^{K\prime}b|U_k\right]}\sqrt{\mathbf{E}\left(\left|D^1g_\kappa(U_j)\{1\{U_k\leq U_j\}-U_j\}-p_j^{K\prime}\pi(U_k)\right|^2|U_k\right)}=O(K^{-d_1})$$

Hence the first term of (44) is $O_P(\sqrt{n}\zeta_{0,K}K^{-d_1})=o_P(1)$. Similarly, we can show that the second term of (44) is $O_P(\sqrt{n}K^{-d_1})=o_P(1)$. The proof is complete.

(Proof of UC) : First we show that

(UC1) : $\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_i\{p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^K-p_i^{K\prime}Q^{-1}p_j^K\}(\kappa_j-g_{\kappa,j})=o_P(1)$.

Then, note that

$$\begin{aligned}\mathbf{E}\left[w_ip_i^{K\prime}Q^{-1}p_j^K(\kappa_j-g_{\kappa,j})|S_i\right]&=w_ip_i^{K\prime}Q^{-1}\mathbf{E}\left[p_j^K(\kappa_j-g_{\kappa,j})|S_i\right]=w_ip_i^{K\prime}Q^{-1}\mathbf{E}\left[p_j^K(\kappa_j-g_{\kappa,j})\right]\\&=w_ip_i^{K\prime}Q^{-1}\mathbf{E}\left[\mathbf{E}\left[p_j^K(\kappa_j-g_{\kappa,j})|U_j\right]\right]=0.\end{aligned}$$

Therefore, by Hoeffding's decomposition

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_ip_i^{K\prime}Q^{-1}p_j^K(\kappa_j-g_{\kappa,j})=\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\mathbf{E}\left[w_ip_i^{K\prime}Q^{-1}\right]p_j^K(\kappa_j-g_{\kappa,j})+r_{3n}(\kappa), \quad (45)$$

where $r_{3n}(\kappa)\triangleq\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}w_ip_i^{K\prime}Q^{-1}p_j^K(\kappa_j-g_{\kappa,j})-\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\mathbf{E}\left[w_ip_i^{K\prime}Q^{-1}\right]p_j^K(\kappa_j-g_{\kappa,j})$ is a degenerate U-process. Following steps in the proof of UB1, we can show that $\sup_\kappa|r_{3n}(\kappa)|=o_P(1)$.

As for the leading term on the right-hand side of (45), observe that it is equal to

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\pi_w'\mathbf{E}\left[p_i^K p_i^{K\prime}Q^{-1}\right]p_j^K(\kappa_j-g_{\kappa,j})+o_P(1) &= \frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\pi_w' p_j^K(\kappa_j-g_{\kappa,j})+o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n} g_{w,j}(\kappa_j-g_{\kappa,j})+o_P(1),\end{aligned}$$

by a repeated application of Assumptions 2U(ii)(a) and 2U(iii). This completes the proof.

Proof of (UA1) : We can expand $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{(b'p^K(U_{n,k}))^2-(b'p^K(U_{n,\lambda,k}))^2\right\}$ as

$$\begin{aligned}&\frac{2}{\sqrt{n}}\sum_{k=1}^{n} b'D^1p^K(U_{n,\lambda,k}^*)p^K(U_{n,\lambda,k}^*)'b(U_{n,k}-U_{n,\lambda,k})\\ =\;&\frac{2}{\sqrt{n}}\sum_{k=1}^{n} b'D^1p^K(U_k)p^K(U_k)'b(U_{n,k}-U_{n,\lambda,k})\\ &+\frac{2}{\sqrt{n}}\sum_{k=1}^{n}\left\{b'D^1p^K(U_{n,\lambda,k}^*)p^K(U_{n,\lambda,k}^*)'b-b'D^1p^K(U_k)p^K(U_k)'b\right\}(U_{n,k}-U_{n,\lambda,k}).\end{aligned}$$

We can show that the second term is $O_P(n^{-2b}\{\zeta_{0,K}\zeta_{2,K}+\zeta_{1,K}^2\})$. By Lemma UA below, the leading term is $O_P(n^{1/2-2b}\zeta_{0,K}\zeta_{1,K})$. Therefore, the result follows by Assumption 2U(iv).

Proof of (UB1) : Let $\mathcal{J}_2\triangleq\{q_K(\cdot,\cdot;\kappa,b)/(\zeta_{0,K}|D^1g_{\tilde{\kappa}}(\cdot)|):(\kappa,b)\in\mathcal{K}\times S_K\}$. Then $r_{2n}(w,\kappa)$ is a degenerate U-process indexed by uniformly bounded $\mathcal{J}_2$. We apply Proposition 1 of Turki-Moalla (p.877) to obtain that (using Lemma UB below)

$$\begin{aligned}&\mathbf{E}\left[\sup_{\varphi\in\mathcal{J}_2}\left|\frac{n^{1-\frac{1}{p}}}{n^2}\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}\{\varphi(U_j,U_k)-\mathbf{E}[\varphi(U_j,U_k)|U_k]\}\right|\right]\\ \leq\;& C\int\left\{1+\log N_{[]}(\varepsilon,\mathcal{J}_2,||\cdot||_p\right\}^{1-\frac{1}{p}}d\varepsilon\leq C\int\{\varepsilon^{-b_1}+K\log(\varepsilon)\}^{1-\frac{1}{p}}d\varepsilon\leq CK^{1-\frac{1}{p}},\end{aligned}$$

since $b_1(1-1/p)<1$. Observe that

$$\begin{aligned}\mathbf{E}\left[\sup_{w,\kappa}|r_{2n}(w,\kappa)|\right] &\leq C\mathbf{E}\left[\sup_{\varphi\in\mathcal{J}_2}\left|\frac{\zeta_{0,K}}{n\sqrt{n}}\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}\{\varphi(U_j,U_k)-\mathbf{E}[\varphi(U_j,U_k)|U_k]\}\right|\right]\\ &= O(n^{-\frac{1}{2}+\frac{1}{p}}K^{1-\frac{1}{p}}\zeta_{0,K})=o(\zeta_{0,K}^{-1})\end{aligned}$$

by Assumption 2U(iv).

Proof of (UC1) : Write $\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\{p_{n,i}^{K\prime}\hat{Q}_n^{-1}p_{n,j}^K - p_i^{K\prime}Q^{-1}p_j^K\}(\kappa_j - g_{\kappa,j})$ as

$$\begin{aligned} & tr\left\{\hat{Q}_n^{-1}\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i(p_{n,j}^K p_{n,i}^{K\prime} - p_j^K p_i^{K\prime})(\kappa_j - g_{\kappa,j})\right\} \\ & +tr\left\{(\hat{Q}_n^{-1} - Q^{-1})\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i p_j^K p_i^{K\prime}(\kappa_j - g_{\kappa,j})\right\}. \end{aligned}$$

By applying the mean-value expansion to the terms in the double sum of the leading term, we write the double sum as

$$\begin{aligned} & \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\left\{D^1p^K(U_j)p^K(U_i)(U_{n,j} - U_j)(\kappa_j - g_{\kappa,j})\right\} \\ & +\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_i\left\{D^1p^K(U_i)p^K(U_j)(U_{n,i} - U_i)(\kappa_j - g_{\kappa,j})\right\} \\ & +o_P(1). \end{aligned} \tag{46}$$

The first term on the right-hand side is equal to

$$\frac{1}{n}\sum_{i=1}^{n} w_i p^K(U_i)'\left(\frac{1}{n\sqrt{n}}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq j,i}^{n}\left\{D^1p^K(U_j)(1\{U_k \leq U_j\} - U_j)(\kappa_j - g_{\kappa,j})\right\}\right). \tag{47}$$

However, note that $\mathbf{E}\left[D^1p^K(U_j)(1\{U_k \leq U_j\} - U_j)(\kappa_j - g_{\kappa,j})|S_j\right] = 0$ and

$$\begin{aligned} & \mathbf{E}\left[D^1p^K(U_j)(1\{U_k \leq U_j\} - U_j)(\kappa_j - g_{\kappa,j})|S_k\right] \\ = & \mathbf{E}\left[D^1p^K(U_j)(1\{U_k \leq U_j\} - U_j)\mathbf{E}\left[\kappa_j - g_{\kappa,j}|S_k, U_j\right]|S_k\right] \\ = & \mathbf{E}\left[D^1p^K(U_j)(1\{U_k \leq U_j\} - U_j)\mathbf{E}\left[\kappa_j - g_{\kappa,j}|U_j\right]|S_k\right] = 0. \end{aligned}$$

Hence the double sum in (47) is a degenerate U-process and similarly as before, we can show that it is $o_P(\zeta_{0,K}^{-1})$. The last term in (46) is written as

$$\frac{1}{\sqrt{n}}\left(\frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n} w_i D^1p^K(U_i)(1\{U_k \leq U_i\} - U_i)\right)\left(\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n} p^K(U_j)'(\kappa_j - g_{\kappa,j})\right) + o_P(1). \tag{48}$$

Then,

$$\sup_{\kappa\in\mathcal{K}}\left\|\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n} p^K(U_j)'(\kappa_j - g_{\kappa,j})\right\| \leq \sup_{(\kappa,b)\in\mathcal{K}\times S_K}\zeta_{0,K}\left\|\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n} p^K(U_j)'b(\kappa_j - g_{\kappa,j})\right\|.$$

It is not hard to see that the last term is $O_P(\zeta_{0,K})$, using the maximal inequality and Lemma A1(ii) of Song (2006). For the double sum in the first parenthesis of (48), note that

$$\mathbf{E}\left[w_i D^1 p^K(U_i)(1\{U_k \leq U_i\} - U_i)|S_i\right] = 0.$$

Hence we write the double sum as

$$\frac{1}{n\sqrt{n}}\sum_{i=1}^n \sum_{k=1,k\neq i}^n \mathbf{E}\left[w_i D^1 p^K(U_i)(1\{U_k \leq U_i\} - U_i)|U_k\right] + r_{4n}(w)$$

where

$$r_{4n}(w) \triangleq \frac{1}{n\sqrt{n}}\sum_{i=1}^n \sum_{k=1,k\neq i}^n \left\{w_i D^1 p^K(U_i)(1\{U_k \leq U_i\} - U_i) - \mathbf{E}\left[w_i D^1 p^K(U_i)(1\{U_k \leq U_i\} - U_i)|U_k\right]\right\}.$$

Similarly as before, we can show that $\sup_w |r_{4n}(w)| = o_P(1)$. Hence the leading term in (48) is equal to $O_P(n^{-1/2}\zeta_{0,K}\zeta_{1,K}) = o_P(1)$ by Assumption 2U(iv).■

Lemma UA : Let $\{\psi_{n,\kappa}(y,x)\}_{n\geq 1}$ be a sequence of real-valued functions on $\mathbf{R}^{d_S}$ indexed by $\kappa \in \mathcal{K}$, where the class $\mathcal{J}_n \triangleq \{\psi_{n,\kappa}(\cdot) : \kappa \in \mathcal{K}\}$ has an envelope $\bar{\psi}_n$ with $||\bar{\psi}_n||_{2,P} < \infty$ and satisfies that $\log N_{[]}(\varepsilon, \mathcal{J}_n, ||\cdot||_{2,P}) \leq C\varepsilon^{-c}$ with $c \in [0,2)$. Suppose further that the conditional density $f(s|u)$ of S_i given $U_i = u$ satisfies Assumption 1U(iii). Then we have

$$\begin{aligned}
&\left\{\mathbf{E}\left[\sup_{(\kappa,\lambda)\in\mathcal{K}\times\Lambda_n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^n \psi_{n,\kappa}(Y_j,X_j)(U_{n,\lambda,j} - U_{n,j}) - \sqrt{n}\mathbf{E}\left[\Delta_{n,\kappa}(Y_i,X_i)(U_{\lambda,i} - U_i)\right]\right|^2\right]\right\}^{1/2} \\
&= O(n^{1/2-2b}||\bar{\psi}_n||_{2,P}),
\end{aligned}$$

where $\Delta_{n,\kappa}(Y_i,X_i) \triangleq \psi_{n,\kappa}(Y_i,X_i) - \mathbf{E}\left[\psi_{n,\kappa}(Y_i,X_i)|U_i\right]$.

Proof of Lemma UA : Write $W_i \triangleq (Y_i, X_i)$. Note that $\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^n \psi_{n,\kappa}(W_j)(U_{n,\lambda,j} - U_{n,j})$ is equal to

$$\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^n \left\{\psi_{n,\kappa}(W_j)(U_{n,\lambda,j} - U_{n,j}) - \mathbf{E}\left[\psi_{n,\kappa}(W_j)(U_{n,\lambda,j} - U_{n,j})\right]\right\} + \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(U_{n,\lambda,j} - U_{n,j})\right] \tag{49}$$

Write the first sum as

$$\frac{1}{n\sqrt{n}}\sum_{j=1,j\neq i}^n \sum_{k=1}^n \left\{r_{\kappa,\lambda}(W_k,W_j) - \mathbf{E}\left[r_{\kappa,\lambda}(W_k,W_j)\right]\right\}, \tag{50}$$

where $r_{\kappa,\lambda}(W_k,W_j) \triangleq \psi_{n,\kappa}(W_j)(1\{\lambda(X_k) \leq \lambda(X_j)\} - 1\{\lambda_0(X_k) \leq \lambda_0(X_j)\})$. Now, the class $\mathcal{V}' \triangleq \{v_\lambda(\cdot,\cdot) : \lambda \in \Lambda_n\}$, $v_\lambda(x_1,x_2) \triangleq 1\{\lambda(x_1) \leq \lambda(x_2)\}$ can be shown to be P-Donsker similarly by using local uniform L_2-continuity of its members (see the first footnote in the proof of Lemma 1U), and so can the class $\mathcal{V}'' \triangleq \{\psi_{n,\kappa}(\cdot)[v_\lambda(\cdot,\cdot) - v_{\lambda_0}(\cdot,\cdot)] : (\kappa,\lambda) \in \mathcal{K}\times\Lambda_n\}$. By using standard arguments of U-processes and Theorem 2.14.5 of van der Vaart and Wellner (1996) that provides the bound for the $L_2(P)$ norm of empirical processes in terms of $L_1(P)$ norm of empirical processes, we can show that the sum in (50) is $O_P(n^{1/2-2b}||\bar{\psi}_n||_{2,P})$.

We turn to the second term in (49). We write the term as

$$\begin{aligned} & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_{\lambda,i} \le U_{\lambda,j}\} - 1\{U_i \le U_j\})\right]. \\ = \ & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_{\lambda,i} \le U_{\lambda,j}\} - 1\{U_{\lambda,i} \le U_j\})\right] \\ & +\sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_{\lambda,i} \le U_j\} - 1\{U_i \le U_j\})\right]. \end{aligned} \tag{51}$$

The first term above is equal to

$$\begin{aligned} & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_{\lambda,i} \le U_{\lambda,j}\} - 1\{U_{\lambda,i} \le U_j\})\right] \\ = \ & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)\mathbf{E}\left[1\{U_{\lambda,i} \le U_{\lambda,j}\} - 1\{U_{\lambda,i} \le U_j\}|W_j\right]\right] \\ = \ & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(U_{\lambda,j} - U_j)\right]. \end{aligned} \tag{52}$$

For the second term, note that

$$\begin{aligned} & \sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_{\lambda,i} \le U_j\} - 1\{U_i \le U_j\})\right] \\ = \ & -\sqrt{n}\mathbf{E}\left[\psi_{n,\kappa}(W_j)(1\{U_j \le U_{\lambda,i}\} - 1\{U_j \le U_i\})\right] \\ = \ & -\sqrt{n}\mathbf{E}\left[\mathbf{E}\left[\psi_{n,\kappa}(W_j)|U_j, X_i\right](1\{U_j \le U_{\lambda,i}\} - 1\{U_j \le U_i\})\right] \\ = \ & -\sqrt{n}\mathbf{E}\left[\mathbf{E}\left[\mathbf{E}\left[\psi_{n,\kappa}(W_j)|U_j\right](1\{U_j \le U_{\lambda,i}\} - 1\{U_j \le U_i\})|X_i\right]\right] \\ = \ & -\sqrt{n}\mathbf{E}\left[\int_{U_i}^{U_{\lambda,i}} \mathbf{E}\left[\psi_{n,\kappa}(W_j)|U_j = u\right] du\right]. \end{aligned} \tag{53}$$

By applying the mean-value expansion using Assumption 1U(iii) and the condition that $\lambda \in \Lambda_n$, the last term becomes

$$-\sqrt{n}\mathbf{E}\left[\mathbf{E}\left[\psi_{n,\kappa}(W_i)|U_i\right](U_{\lambda,i} - U_i)\right] + O_P(n^{1/2-2b}||\bar{\psi}_n||_{2,P}).$$

Combining this with (52), we obtain the wanted result.■

Lemma UB : As for the classes $\mathcal{J}_1$ and $\mathcal{J}_2$ defined by

$$\begin{aligned} \mathcal{J}_1 & \triangleq \{w(\cdot)b'p^K(\cdot) : (w,b) \in \mathcal{W} \times S_K\} \\ \mathcal{J}_2 & \triangleq \{q_K(\cdot,\cdot;\kappa,b)/\{\zeta_{0,K}|D^1 g_{\tilde{\kappa}}(\cdot)|\} : (\kappa,b) \in \mathcal{K} \times S_K\}, \end{aligned}$$

where $q_K(u_1,u_2;\kappa,b) \triangleq D^1 g_\kappa(u_1)b'p^K(u_1)\{u_1 - 1\{u_2 \le u_1\}\}$, it is satisfied that

$$\begin{aligned} \log N_{[]}(\varepsilon, \mathcal{J}_1, ||\cdot||_{2,P}) & \le C\{\varepsilon^{-b_3} + K\log(\varepsilon/K^{\sigma/2})\}, \\ \log N_{[]}(\varepsilon, \mathcal{J}_2, ||\cdot||_{2,P}) & \le C\{\varepsilon^{-b_1} + K\log(\varepsilon)\}. \end{aligned}$$

Proof of Lemma UB : First consider $\mathcal{J}_1$. We take $\{(w_j, \Delta_j)\}_{j=1}^{N_1}$ such that $\{[w_j - \Delta_j, w_j + \Delta_j]\}_{j=1}^{N_1}$ form ε-brackets that cover $\mathcal{W}$ and choose $\{b_k\}_{k=1}^{N_2}$, the centers of $2\varepsilon/K^{\sigma/2}$-balls that cover S_K. Take $\Delta_{jk} \triangleq \Delta_j \sqrt{b_k p^K p^{K\prime} b_k} + \tilde{w} \left\{tr\left(\{p^K p^{K\prime}\}^{M_\sigma/2}\right)\right\}^{1/M_\sigma}$. As in the proof of Theorem 1L in Song (2006), we note that

$$||\sqrt{b_k p^K p^{K\prime} b_k}||_{2,P} \leq C \text{ and } ||\left\{tr\left(\{p^K p^{K\prime}\}^{M_\sigma/2}\right)\right\}^{1/M_\sigma}||_{2,P} \leq CK^{\sigma/2}.$$

These bounds are obtained due to the rotation of p_K performed in the beginning of the proof of Lemma 1U. Then it is easy to check that the set $\{[w_j b_k' p^K - \Delta_{jk}, w_j b_k' p^K + \Delta_{jk}]\}_{j=1}^{N_1}$ forms $C\varepsilon$-brackets that cover $\mathcal{J}_1$. Hence

$$\begin{aligned} \log N_{[]}(\varepsilon, \mathcal{J}_1, ||\cdot||_{2,P}) &\leq \log N_{[]}(\varepsilon/C, \mathcal{W}, ||\cdot||_{2,P}) + \log N(\varepsilon/K^{\sigma/2}, S_K, ||\cdot||) \\ &\leq C\varepsilon^{-b_3} + CK\log(\varepsilon/K^{\sigma/2}). \end{aligned}$$

Let us turn to $\mathcal{J}_2$. Note that

$$\begin{aligned} \left|D^1 g_{\kappa_1}(u) - D^1 g_{\kappa_2}(u)\right| &= \left|\lim_{v\to 0}\left\{\mathbf{E}\left(\kappa_1(Y) - \kappa_2(Y)|U = u + v\right) - \mathbf{E}\left(\kappa_1(Y) - \kappa_2(Y)|U = u - v\right)\right\}/(2v)\right| \\ &\leq \lim_{v\to 0}\int |\kappa_1(y) - \kappa_2(y)|\frac{|f(y|u+v) - f(y|u-v)|}{2v}dy \leq C||\kappa_1 - \kappa_2||_{2,P} \end{aligned}$$

by Assumption 1U(iii). Hence similarly as before,

$$\log N_{[]}(\varepsilon, \mathcal{J}_2, ||\cdot||_{2,P}) \leq \log N_{[]}(\varepsilon/C, \mathcal{K}, ||\cdot||_{2,P}) + \log N(\varepsilon, S_K, ||\cdot||).$$

■

References

[1] Aït-Sahalia, Y., P. J. Bickel, and T. M. Stoker (2001): "Goodness-of-fit tests for regression using kernel methods," *Journal of Econometrics*, 105, 363-412.

[2] Andrews, D. W. K (1994): "Empirical process method in econometrics," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden, Amsterdam: North-Holland.

[3] Andrews, D. W. K (1995): "Nonparametric kernel estimation for semiparametric models," *Econometric Theory*, 11, 560-596.

[4] Andrews, D. W. K. (1997): "A conditional Kolmogorov test," *Econometrica* 65, 1097-1128.

[5] Begun, J. M., W. J. Hall, W-M, Huang, and J. A. Wellner (1983): "Information and asymptotic efficiency in parametric-nonparametric models," *Annals of Statistics*, 11, 432-452.

[6] Bickel, P. J. (1982): "On adaptive estimation," *Annals of Statistics*, 10, 647-671.

[7] Bickel, P. J. , A.J. Klaassen, Y. Rikov, and J. A. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, Springer Verlag, New York.

[8] Bierens, H. J. (1990): "A consistent conditional moment test of functional form" *Econometrica* 58, 1443-1458.

[9] Bierens, H. J. and W. Ploberger (1997): "Asymptotic theory of integrated conditional moment test" *Econometrica* 65, 1129-1151.

[10] Choi, S., W. J. Hall, and A. Schick (1996): "Asymptotically uniformly most powerful tests in parametric and semiparametric models," *Annals of Statistics*, 24, 841-861.

[11] Cox, D. R. and N. Reid (1987): "Parameter orthogonality and approximate conditional inference (with discussion)," *Journal of the Royal Statistical Society* Series B. 49, 1-39.

[12] Delecroix, M. W. Härdle, and M. Hristache (2003): "Efficient estimation in conditional single-index regression," *Journal of Multivariate Analysis* 86, 213-226.

[13] Duan, N. and K. Li (1989): "Regression analysis under link violation," *Annals of Statistics*, 17, 1009-1052.

[14] Escanciano, J. C. (2006): "Consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030-1051.

[15] Ermakov, M. S. (1995): "Asymptotic minimaxity of tests of the Kolmogorov and Omega-Square types," *Theory of Probability and its Applications*, 40, 41-51.

[16] Fan, Y. and Q. Li. (1996): "Consistent model specification tests: omitted variables and semi-parametric forms," *Econometrica*, 64, 865-890.

[17] Guerre E. and P. Lavergne (2002): "Optimal minimax rates for nonparametric specification testing in regression models," *Econometric Theory*, 18, 1139-1171.

[18] Hansen, L. P. (1982): "Large sample properties of generalized method of moment estimators," *Econometrica*, 50, 1029-1054.

[19] Härdle, W., P. Hall, and H. Ichimura (1993): "Optimal smoothing in single-index models," *Annals of Statistics*, 21, 157-178.

[20] Härdle, W. and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits" *Annals of Statistics* 21, 1926-1974.

[21] Hristache, M., A. Juditsky, and V. Spokoiny (2001): "Direct estimation of the index coefficient in a single-index model," *Annals of Statistics*, 29, 595-623.

[22] Horowitz, J. L. and V. G. Spokoiny (2001): "An adaptive rate-optimal test of a parametric model against a nonparametric alternative," *Econometrica* 69, 599-631.

[23] Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS of single index models," *Journal of Econometrics*, 58, 71-120.

[24] Ingster, Y. I. (1993): "Asymptotically minimax hypothesis testing for nonparametric alternatives, parts I, II, and III," *Mathematical Methods of Statistics*, 2, 85-114, 171-189, and 249-268.

[25] Ingster, Y. I. and I. A. Suslina (2003): *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, Springer, New York.

[26] Janssen, A. (2000): "Global power functions of goodness-of-fit tests," *Annals of Statistics*, 28, 239-253.

[27] Klein, R. W. and R. H. Spady (1993): "An efficient semiparametric estimation for binary response models," *Econometrica*, 61, 387-421.

[28] Leeb, H. and B. A. Pötcher (2005): "Model selection and inference: facts and fiction," *Econometric Theory*, 21, 21-59.

[29] Newey, W. K. (1985): "Generalized method of moment specification testing," *Journal of Econometrics*, 29, 229-256.

[30] Newey, W. K. and T. M. Stoker (1993): "Efficiency of weighted average derivative estimators and index models," *Econometrica*, 61, 1199-1223.

[31] Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79, 147-168.

[32] Nikitin, Y. (1995): *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press.

[33] Pollard, D. (1989): "A maximal inequality for sums of independent processes under a bracketing entropy condition". Unpublished manuscript.

[34] Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric estimation of index coefficients," *Econometrica* 57, 1403-1430.

[35] Song, K. (2006): "Uniform convergence of series estimators over function spaces", Working Paper.

[36] Stinchcombe, M., and H. White (1998): "Consistent specification testing with nuisance parameters present only under the alternative" *Econometric Theory,* 14, 295-325.

[37] Stute, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics,* 25, 613-641.

[38] Stute, W. and L. Zhu (2002): "Model checks for generalized linear models," *Scandinavian Journal of Statistics*, 29, 535-545.

[39] Stute, W. and L. Zhu (2005): "Nonparametric checks for single-index models," *Annals of Statistics,* 33, 1048-1083.

[40] Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models," *Journal of Econometrics*, 30, 415-443.

[41] Turki-Moalla, K. (1998): "Rates of convergence and law of the iterated logarithm for U-processes," *Journal of Theoretical Probability*, 11, 869-906.

[42] van der Vaart, A. (1996): "New Donsker classes," *Annals of Probability.* 24, 2128-2140.

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

[44] Xia, Y., W. K. Li, H. Tong, and D. Zhang (2004): "A goodness of fit test for single-index models", *Statistical Sinica*, 14, 1-39.