Testing Additive Separability of Error Term in Nonparametric Structural Models*

Liangjun Su[a], Yundong Tu[b], Aman Ullah[c]

[a] School of Economics, Singapore Management University

[b] Guanghua School of Management and Center for Statistical Science
Peking University

[c] Department of Economics, University of California, Riverside

September 12, 2012

Abstract

This paper considers testing additive error structure in nonparametric structural models, against the alternative hypothesis that the random error term enters the nonparametric model non-additively. We propose a test statistic under a set of identification conditions considered by Hoderlein, Su and White (2012), which require the existence of a control variable such that the regressor is independent of the error term given the control variable. The test statistic is motivated from the observation that, under the additive error structure, the partial derivative of the nonparametric structural function with respect to the error term is one under identification. The asymptotic distribution of the test is established and a bootstrap version is proposed to enhance its finite sample performance. Monte Carlo simulations show that the test has proper size and reasonable power in finite samples.

Key Words: Additive Separability; Hypotheses Testing; Nonparametric Structural Equation; Nonseparable Models

JEL Classification: C12; C13; C14

*The authors gratefully thank the Co-editors and two anonymous referees for their many constructive comments on the previous version of the paper. They are also thankful to Rosa Matzkin and Aditi Bhattacharya for some discussions on the subject matter of this paper. The third author is thankful to UCR, Academic Senate for its financial support. Correspondence to: Liangjun Su, School of Economics, SMU, 90 Stamford Road, Singapore 178903, E-mail: ljsu@smu.edu.sg. Yundong Tu, Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China 100871. E-mail: yundong.tu@gsm.pku.edu.cn. Aman Ullah, Department of Economics, University of California, Riverside, CA 92521, E-mail: aman.ullah@ucr.edu, Tel.: +1 951 827 1591.

1 Introduction

Economic models that incorporate stochastic features usually proceed by specifying the relationship between an observed dependent variable (or variable of interest), a set of observed independent variables (or explanatory variables), and some unobservable random term represented by error (or shock). This paper examines how to deal with this unobserved error in the econometric modeling process and whether it enters the econometric model as a separable additive component or as a nonseparable element.

When the error term is nonadditive, the conventional identification and estimation approaches for additive nonparametric models are not applicable anymore. Therefore, new approaches to identification and estimation are called upon for nonseparable nonparametric models. A great deal of efforts was devoted to progress in this direction in the past two decades or so. Some earlier work includes Roehrig (1988), Brown and Matzkin (1998), Matzkin (1991), Olley and Pakes (1996), Heckman and Vytlacil (1999, 2001), and Blundell and Powell (2003). Matzkin (2003) presents estimators for nonparametric nonadditive models and shows their asymptotic characteristics under a set of assumptions that may be implied by economic theory. Altonji and Matzkin (2005) adopt a conditional independence assumption to estimate the average derivative of a nonparametric function and the distribution of the unobservable random term, when the unobservable is nonadditive and the regressors are endogenous. Briesch, Chintagunta and Matzkin (2010) provide a method to estimate discrete choice models with unobserved heterogeneity that enters the subutility function nonadditively. Heckman, Matzkin and Nesheim (2010) establish nonparametric identification of structural functions and distributions in general nonparametric nonadditive hedonic models by relaxing the assumptions of additive marginal utility and additive marginal product function adopted in Ekeland, Heckman and Nesheim (2004). Altonji, Ichimura and Otsu (2012) present a simple method to estimate the marginal effects of observable variables on a limited dependent variable, when the dependent variable is a nonseparable function of observables and unobservables.

Albeit the literature is flooded with approaches that are capable of tackling both separable and nonseparable nonparametric models, there is no valid method available to distinguish which model is more appropriate for the problem confronted by the researchers. We believe that there are at least four reasons that amplify the urgent need for some convincing testing procedures to detect the way through which the unobservable random term enters the economic structure. They are: (1) The economic meaning of an unobservable random term varies from case to case; (2) The identification and statistical properties of the estimated underlying economic structure depend on whether additive separability holds; (3) The identification and estimation of other economic structures also relies on the separability properties; and (4) There is a lack of consistent testing procedures to detect additive separability of unobservables in the literature. These are described below.

Economic meaning of an unobservable. An additive unobservable takes on the traditional explanation as measurement error of the variable of interest, or a level shift of the dependent variable due to some random shocks to the economy, or some minor factors other than the included regressors that may affect the dependent variable. A nonadditive unobservable random term, on the other hand, may adopt explanations such as a *heterogeneity* parameter in a utility function, the productivity shock or utility value for some unobserved attributes, etc. See, for example, Heckman (1974), Heckman and Willis (1977), McFadden (1974), and Lancaster (1979), among others. A clarification of the additivity property of the unknown economic structure helps to identify the economic meaning of an unobservable, which facilitates further evaluation of sources of heterogeneity, improvement of productivity for firms and better economic policy proposals.

Identification and statistical properties of the estimators. Classical additive nonparametric models can be identified under standard conditional moment restrictions, and estimated, for example, by conventional nonparametric kernel or sieve methods. The consistency and asymptotic normality of these nonparametric estimators have been well understood. In contrast, methods to identify and estimate nonadditive nonparametric functions are relatively new in the literature and have not yet been fully explored. Matzkin (2003) presents an estimator of the nonseparable nonparametric random function, and shows that it is consistent and asymptotically normal under certain identification conditions. She argues that her identification conditions are not very strong since they may be implied by some economic theory and are rather straightforward to derive if certain parametric functional forms are tolerated. Yet, one concern regarding these identification conditions is that the underlying economic theory itself be subject to valid tests, not to mention its implications or the parametric functional forms that are implicitly needed to facilitate the formulation of identification conditions. Therefore, there is a potentially high cost of applying these conditions for identification purposes.

Estimation of other economic structure. Quite often it is also of interest to estimate other sensible economic structure. Examples are available in the policy evaluation literature. Heckman and Vytlacil (2005) point out that the entire recent literature on instrumental variable estimators with heterogeneous responses "relies critically on the assumption that the treatment choice equation has a representation in the additive separable form." Heckman and Vytlacil (2001) show that, even after some transformation, the defined marginal treatment effect (MTE) is still not identified through linear instrumental variable (LIV), and MTE defined in this way precludes getting treatment parameters via integration. Furthermore, Heckman and Vytlacil (2005) also notice that nonseparability will lead to failure of the index sufficiency. In other words, additive separable assumption simplifies the estimation of some economic structure. Yet, there is no convincing testing procedure to provide evidence that the economic structure under investigation is indeed additive.

Lack of specification tests for separability. Since Hausman's (1978) seminal work a large literature on testing for the correct specification of functional forms has developed; see Bierens (1982, 1990), Ruud (1984), Newey (1985), Tauchen (1985), White (1987), Robinson (1989), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White (1995), Zheng (1996), Andrews (1997), Bierens and Ploberger (1997), Li and Wang (1998), Stinchcombe and White (1998), Hsiao, Li and Racine (2007), Su and Ullah (2012), among others. Although much progress has been made towards econometric model specification, almost all the literature has been confined to functional forms that only accommodate additive random errors. Rare exceptions include Hoderlein, Su and White (2012, **HSW** hereafter) and Lu and White (2012, **LW** hereafter). The former paper proposes a nonparametric test for monotonicity in unobservables in nonparametric nonseparable structural models whereas the latter considers a nonparametric test for additive separability in structural models based on a test for conditional independence. As **LW** argue, many important identification results in the econometrics literature depend on the separability of structural equations, and when correctly imposed, separability helps achieve estimation efficiency in various scenarios. Thus it is desirable to consider tests for separability.

In this paper we propose a consistent testing procedure that is able to differentiate an additively separable model from a nonadditive one. Like **LW,** we consider testing the null hypothesis of additive separability in a nonparametric structural model (see eq. (2.1) below) under a conditional exogeneity condition (see Assumption I.3 below). Unlike **LW,** we follow **HSW** and also assume a monotonicity condition to identify the structural equation without imposing additive separability because our testing strategy requires the identification and estimation of the nonparametric structural function under both the null and the alternative. Note that the monotonicity condition is naturally guaranteed under the null but it may not be ensured under the alternative. **LW** do not need to impose such a condition under the alternative because they transform their test of additive separability to a test of conditional independence, which is implied by but in general does not imply the null. So they avoid the identification and estimation of the nonparametric structural model under the alternative. The cost is that their test is not consistent against all global alternatives because of the gap between the implied hypothesis and the original null hypothesis.[1] In contrast, our test is based on the estimate of the partial derivative of the structural function with respect to the unobservable which is identically one under the null and not otherwise. We shall study the asymptotic distributions of our test under the null hypothesis and a sequence of Pitman local alternatives and establish the consistency of our test.

The rest of the paper is structured as follows. Section 2 states our testing problem and presents the test statistic. Section 3 provides asymptotic properties of our proposed test. We perform a small set of Monte Carlo experiments in Section 4 to investigate the finite sample size and power behavior of our test. In Section 5, we conclude and remark on future research. All proofs are relegated to the appendix.

Notation: Throughout the paper we use upper case letters (e.g., X, Y, Z, ε) to denote random variables and their corresponding lower case letters (e.g., x, y, z, e) to indicate the realizations.

2 Testing Additive Separability

The model of interest can be formulated as

$$Y = m(X, \varepsilon) \tag{2.1}$$

where Y and X are observables, ε is an unobserved random shock, and $m(\cdot,\cdot)$ is an unknown but smooth function defined on $\mathcal{X}\times\mathcal{E}$, where $\mathcal{X}\subset\mathbb{R}^{d_X}$ and $\mathcal{E}\subset\mathbb{R}$. $m(\cdot,\cdot)$ is termed as "nonadditive random function" by Matzkin (2003). We are interested in testing whether the random error ε enters the model as an additive term.

[1]Interestingly, **LW** show that by imposing monotonicity in unobservables for the nonparametric structural function, they can establish the equivalence between the conditional independence and additive separability hypotheses. In this case, their test is also consistent.

2.1 Identification

The model specified in (2.1) is generally not identified without further restriction. For testing purpose, we only consider the situation in which $m(\cdot,\cdot)$ is identified. Matzkin (2003, 2007) studies the identification issue extensively. **HSW** revisit the identification issue and give a set of identification conditions that are analogous to Specification I in Matzkin (2003) but much easier to use. The identification conditions in **HSW** require the existence of a control variable Z such that X is independent of ε given Z, or in short, $X \perp \varepsilon \mid Z$. We shall use $\mathcal{Z}$ to denote the support of Z, and $G(\cdot|x,z)$ to denote the conditional cumulative distribution function (CDF) of Y given $(X,Z)=(x,z)$.

Following **HSW**, we make the following identification assumptions.

Assumption I.1 For all $x \in \mathcal{X}$, $m(x,\cdot)$ is strictly increasing.

Assumption I.2 There exists $\bar{x} \in \mathcal{X}$ such that $m(\bar{x},e)=e$ for all $e \in \mathcal{E}$.

Assumption I.3 $X \perp \varepsilon \mid Z$, where Z is not measurable$-\sigma(X)$.

Assumption I.4 For each $(x,z) \in \mathcal{X} \times \mathcal{Z}$, $G(\cdot|x,z)$ is invertible.

Remark 1. I.1-I.4 parallels Assumptions A.2, A.3, B.1, B.2, respectively, in **HSW**. I.1 and I.3 are also analogous to Assumptions I.2 and I.3 in Matzkin (2003) and I.2 corresponds to their Specification I discussed in their Section 3.1 where an assumption similar to I.4 is also implicitly made.

Remark 2. As **HSW** remark, given I.1 and the structural functional relationship in (2.1), for any $\bar{x} \in \mathcal{X}$ there exists a function, say $\bar{m}$, for which I.1 and I.2 hold. This implies that under I.1, any point in $\mathcal{X}$ can play the role of $\bar{x}$ in I.2. Given this $\bar{x}$, we can replace m with $\bar{m}$, such that $\bar{m}(x,\cdot)$ is strictly increasing for all $x \in \mathcal{X}$, and $\bar{m}(\bar{x},\varepsilon)=\varepsilon$ a.s. With this normalization in mind, we can drop the reference to $\bar{m}$ and simply work with m, as what is stated in I.2. In what follows, we simply choose a particular value $\bar{x}$, such as the vector of sample medians of X,[2] and adopt the normalization rule $m(\bar{x},e)=e$.

The following lemma summarizes some of the identification results in **HSW**.

Lemma 2.1 *Suppose (2.1) and Assumptions I.1-I.4 hold. Then*

$$m(x,e) = G^{-1}(G(e|\bar{x},z) \mid x,z) \qquad \forall (e,x,z) \in \mathcal{E} \times \mathcal{X} \times \mathcal{Z}, \text{ and}$$
$$\varepsilon = G^{-1}(G(Y|X,z) \mid \bar{x},z) \qquad \forall z \in \mathcal{Z}.$$

The above identification result lays down the foundation for our test of additive separability. It says that under I.1-I.4, the structural response function $m(\cdot,\cdot)$ and the unobserved error term ε can be identified. Note that we do not need the existence of the conditional probability density function (PDF), say, $g(\cdot|,x,z)$, of Y given $(X,Z)=(x,z)$ in Lemma 2.1. If $g(\cdot|,x,z)$ exists, the first result in the above lemma implies that

$$D_e m(x,e) \equiv \frac{\partial m(x,e)}{\partial e} = \frac{g(e \mid \bar{x},z)}{g(m(x,e) \mid x,z)} \tag{2.2}$$

[2] The sample median is random but converges to the population median at the parametric rate. Noting that our test is of nonparametric nature and has power against local alternatives converging to the null at the nonparametric rate, this implies that one can treat the sample median as the population median without affecting the asymptotic theory studied below.

where $D_e m(\cdot,\cdot)$ denotes the partial derivative of $m(\cdot,\cdot)$ with respect to its second argument. Note that the partial derivative $D_e m(x,e)$ is also identified provided g is well defined. Note also that z appears only on the right hand side of (2.2).

2.2 Hypotheses

Given the model specified in (2.1) we are interested in testing whether $m(\cdot,\cdot)$ is additively separable, that is, whether there exist some measurable functions $m_1(\cdot)$ and $m_2(\cdot)$ such that $m(X,\varepsilon) = m_1(X) + m_2(\varepsilon)$ almost surely (a.s.). Therefore the null hypotheses is

$$\mathbb{H}_0 : m(X,\varepsilon) = m_1(X) + m_2(\varepsilon) \text{ a.s.} \tag{2.3}$$

for some measurable functions $m_1(\cdot)$ and $m_2(\cdot)$, and the alternative hypothesis is the negation of $\mathbb{H}_0$:

$$\mathbb{H}_1 : P[m(X,\varepsilon) = m_1(X) + m_2(\varepsilon)] < 1 \tag{2.4}$$

for all measurable functions $m_1(\cdot)$ defined on $\mathcal{X}$ and $m_2(\cdot)$ on $\mathcal{E}$.

The simulation experiment in Matzkin (2003) shows that the nonparametric estimate of an additive model without imposing the additive restriction is significantly worse than that with the additive restriction correctly imposed. This highlights the importance of testing the additivity structure of the unknown relationship between the observables and unobservables.

Under I.1, $m_2(\cdot)$ is strictly increasing in (2.3). Given I.2 and $\mathbb{H}_0$ in (2.3), we have

$$m(\bar{x},\varepsilon) = m_1(\bar{x}) + m_2(\varepsilon) = \varepsilon \text{ a.s.},$$

implying that $m_2(\varepsilon) - \varepsilon$ is a constant with probability one. Therefore we observe that under $\mathbb{H}_0$ and I.1-I.2,

$$D_e m(X,\varepsilon) \equiv \frac{\partial m(X,\varepsilon)}{\partial e} = \left.\frac{\partial m(X,e)}{\partial e}\right|_{e=\varepsilon} = 1 \text{ a.s.} \tag{2.5}$$

This observation is very important because it motivates us to propose a test based on the derivative of $m(\cdot,\cdot)$ with respect to its second argument. In particular, we will consider a test for $\mathbb{H}_0$ based on the following weighted L_2-distance measure between $D_e m(x,e)$ and 1:

$$J = \int [D_e m(x,e) - 1]^2 a_0(x,e)\, dP(x,e) \tag{2.6}$$

where $P(\cdot)$ is the joint CDF of X and ε and $a_0(\cdot,\cdot)$ is a nonnegative weight function defined on $\mathcal{X}_0 \times \mathcal{E}_0$, where $\mathcal{X}_0$ and $\mathcal{E}_0$ are a compact subset of $\mathcal{X}$ and $\mathcal{E}$, respectively.[3]

[3]Here and below we restrict (x,z,e) to $\mathcal{X}_0 \times \mathcal{Z}_0 \times \mathcal{E}_0$ because we need to estimate $G(e|x,z)$ and its inverse $G^{-1}(\cdot \mid x,z)$ which can not be estimated sufficiently well if $G(e|x,z)$ is close to either 0 or 1, say, when (x,z,e) lies at the boundary of its support $\mathcal{X} \times \mathcal{Z} \times \mathcal{E}$.

2.3 Estimation and test statistic

Let $\{(Y_i, X_i, Z_i), \ i = 1, ..., n\}$ denote a random sample for (Y, X, Z) that has support $\mathcal{Y} \times \mathcal{X} \times \mathcal{Z}$. Let $U_i \equiv (X_i', Z_i')'$. Let $u \equiv (x', z')'$ be a $d \times 1$ vector, $d \equiv d_X + d_Z$, where x is $d_X \times 1$ and z is $d_Z \times 1$. Let $W_i \equiv (Y_i, U_i')'$ and $w \equiv (y, u')'$.

To propose a feasible test statistic, we need to estimate $G(\cdot|u)$, $G^{-1}(\cdot|u)$, $g(\cdot|u)$, and ε_i. Throughout, we rely on local constant estimates.[4] First, we estimate $G(y|u)$ by

$$\hat{G}_b(y|u) \equiv \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u) 1\{Y_i \leq y\} / \hat{g}_b(u)$$

where $\hat{g}_b(u) \equiv \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u)$, $K_b(\cdot) \equiv K(\cdot/b)/b$, $K(\cdot)$ a kernel function defined on $\mathbb{R}^d$, $b \equiv b(n)$ is a bandwidth parameter, and $1\{\cdot\}$ is the usual indicator function. Then we estimate $G^{-1}(\cdot|u)$ by inverting $\hat{G}_b(\cdot|u)$ to obtain

$$\hat{G}_b^{-1}(\cdot\ |u) = \inf\left\{y \in \mathbb{R} : \hat{G}_b(y|u) \geq \cdot\right\},$$

which is well defined if K is always nonnegative such that $\hat{G}_b(y|u)$ is always between zero and one and monotone in y. Nevertheless, to reduce the bias of these kernel estimates, we permit the use of a higher order kernel for K when d is large (e.g., $d \geq 4$). In this case, we may only consider the estimates $\hat{G}_b$ and $\hat{G}_b^{-1}$ on a subset of the observations for which $\hat{G}_b$ lies on a compact subset of $(0, 1)$ for large n, which is also required in our asymptotic analysis. Alternatively, as a referee kindly suggests, one can consider some sort of rearrangement technique to ensure the monotonicity of $\hat{G}_b(\cdot|u)$ in finite samples in the case of higher order kernel. Chernozhukov, Fernández-Val and Galichon (2010) address the longstanding problem of lack of monotonicity in the estimation of conditional and structural quantile functions by rearrangement. We conjecture that similar technique can be used to yield monotone estimate of the CDF but leave this for future study.

Similarly, we estimate the conditional PDF $g(y|u)$ of Y_i given $U_i = u$ by

$$\hat{g}_c(y|u) = \sum_{i=1}^{n} L_c(W_i - w) / \sum_{i=1}^{n} L_c(U_i - u)$$

where $L_c(\cdot) \equiv L(\cdot/c)/c$, $L(\cdot)$ a kernel function defined on $\mathbb{R}^d$ or $\mathbb{R}^{d+1}$, and $c \equiv c(n)$ is a bandwidth parameter.[5]

With $\hat{G}_b$ and $\hat{G}_b^{-1}$ on hand, Lemma 2.1 motivates us to estimate $m(x, e) = G^{-1}(G(e|\bar{x}, z) \mid x, z)$ by

$$\hat{m}_b(x, e) = \int \hat{G}_b^{-1}(\hat{G}_b(e|\bar{x}, z)|x, z)\, dH(z) \tag{2.7}$$

and $\varepsilon_i = G^{-1}(G(Y_i|X_i, z) \mid \bar{x}, z)$ by

$$\hat{\varepsilon}_i = \int \hat{G}_b^{-1}(\hat{G}_b(Y_i|X_i, z) \mid \bar{x}, z) dH(z), \tag{2.8}$$

[4] Alternatively one can follow **HSW** and apply the local polynomial method to obtain all necessary estimates. But we find that the local constant method is less computational expensive than the latter.

[5] We abuse the notation a little bit. The multivariate kernel function L can be defined either on $\mathbb{R}^d$ for U_i or $\mathbb{R}^{d+1}$ for W_i, which is self evident from its argument.

where $H(\cdot)$ is a CDF that has a PDF $h(\cdot)$ with compact support $\mathcal{Z}_0 \subset \mathcal{Z}$. Note that here we suppress the dependence of $\hat{m}_b$ and $\hat{\varepsilon}_i$ on H and that of $\hat{\varepsilon}_i$ on b. Like **HSW**, the use of H helps to eliminate the variability of estimators of $m(x,e)$ and ε_i based on an arbitrary choice of z. In view of the fact that the left hand side of (2.2) does not depend on z, we propose to estimate $D_e m(x,e)$ by[6]

$$\widehat{D_e m}(x,e) = \int \frac{\hat{g}_c(e|\bar{x},z)}{\hat{g}_c(\hat{m}_b(x,e) \mid x,z)} dH(z). \tag{2.9}$$

Based on (2.6), we might consider

$$\breve{J}_n = n^{-1} \sum_{i=1}^{n} \left[\int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)} dH(z) - 1 \right]^2 a_0(X_i, \hat{\varepsilon}_i), \tag{2.10}$$

which can be regarded as a sample analogue of J defined in (2.6). To simplify the analysis, in view of $\varepsilon_i = m^{-1}(X_i, Y_i)$ where $m^{-1}(x,\cdot)$ denotes the inverse of $m(x,\cdot)$ $\forall x \in \mathcal{X}$, we define $a(X_i,Y_i) = a_0\left(X_i, m^{-1}(X_i,Y_i)\right)$ and consider the following simpler test statistic

$$\hat{J}_n = n^{-1} \sum_{i=1}^{n} \left[\int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)} dH(z) - 1 \right]^2 a(X_i, Y_i). \tag{2.11}$$

Apparently, the support of a_0 and that of a are closely related to each other, and the nonnegativity of a is inherited from that of a_0. We will make assumptions on the support of a directly so that $\hat{J}_n$ is well defined. Let $\mathcal{Z}_0$ denote the compact support of $h(\cdot) \equiv \partial H(\cdot)/\partial z$ that is a proper subset of $\mathcal{Z}$. Let $\mathcal{X}_0 \times \mathcal{Y}_0$ denote the compact support of $a(\cdot,\cdot)$ where $\mathcal{X}_0$ and $\mathcal{Y}_0$ are a proper subset of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\mathcal{E}_0$ denote the support of $\varepsilon_i = \int G^{-1}(G(Y_i|X_i,z) \mid \bar{x},z)dH(z)$ when (X_i,Y_i) is constrained to lie in $\mathcal{X}_0 \times \mathcal{Y}_0$. $\hat{G}_b(y|x,z)$ will be bounded away below from 0 and above from 1 for all $(y,x,z) \in \mathcal{Y}_0 \times \mathcal{X}_0 \times \mathcal{Z}_0$ for sufficiently large sample size n by the consistency of $\hat{G}_b$. This will ensure $\hat{\varepsilon}_i = \int \hat{G}_b^{-1}(\hat{G}_b(Y_i|X_i,z) \mid \bar{x},z)dH(z)$ to be well defined for observations with nonzero value of $a(X_i,Y_i)$.

We study the asymptotic distribution of $\hat{J}_n$ in the next section.

3 Asymptotic Distribution

In this section we first present assumptions that are used in deriving the asymptotic distribution of our test statistic $\hat{J}_n$. Then we study its asymptotic distribution under the null hypothesis and a sequence of Pitman local alternatives. We also prove the consistency of the test and propose a bootstrap method to obtain the bootstrap p-value.

3.1 Assumptions

Let $\boldsymbol{j} \equiv (j_1,...,j_d)$ be a d-vector of non-negative integers and $|\boldsymbol{j}| \equiv \sum_{i=1}^{d} j_i$. To study asymptotic distribution of our test statistic, we use the following assumptions.

[6] When G and G^{-1} are estimated by the local polynomial regressions, the asymptotic distributions of $\hat{m}_b(x,e)$, $\hat{\varepsilon}_i$, and $\widehat{D_e m}(x,e)$ are quite complicated and studied in **HSW**.

Assumption A.1 Let $W_i \equiv (Y_i, X_i', Z_i')'$, $i = 1, 2, ..., n$, be IID random variables with W_i distributed identically to (Y, X', Z').

Let $g(w)$, $g(u)$, and $g(y|u)$ denote the PDF of W_i, that of U_i, and the conditional PDF of Y_i given $U_i = u$, respectively. Let $\mathcal{U} \equiv \mathcal{X} \times \mathcal{Z}$ and $\mathcal{U}_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0$. Let $\mathcal{Y}_0 \equiv [\underline{y}, \bar{y}]$ denote a proper subset of $\mathcal{Y}$.

Assumption A.2 (i) $g(u)$ is continuous in $u \in \mathcal{U}$, and $g(y|u)$ is continuously differentiable in $y \in \mathcal{Y}$ for all $u \in \mathcal{U}$.

(ii) There exist $C_1, C_2 \in (0, \infty)$ such that $C_1 \leq \inf_{u \in \mathcal{U}_0} g(u) \leq \sup_{u \in \mathcal{U}_0} g(u) \leq C_2$ and $C_1 \leq \inf_{(y,u) \in \mathcal{Y}_0 \times \mathcal{U}_0} g(y|u) \leq \sup_{(y,u) \in \mathcal{Y}_0 \times \mathcal{U}_0} g(y|u) \leq C_2$.

Assumption A.3 (i) There exist $\underline{\tau}, \bar{\tau} \in (0, 1)$ such that $\underline{\tau} \leq \inf_{u \in \mathcal{U}_0} G(\underline{y}|u) \leq \sup_{u \in \mathcal{U}_0} G(\bar{y}|u) \leq \bar{\tau}$ and $\underline{\tau} \leq \inf_{z \in \mathcal{Z}_0} G(\underline{y}|\bar{x}, z) \leq \sup_{z \in \mathcal{Z}_0} G(\bar{y}|\bar{x}, z) \leq \bar{\tau}$.

(ii) $G(\cdot|u)$ admits the PDF $g(\cdot|u)$ and is equicontinuous: $\forall \epsilon > 0, \exists \delta > 0 : |y - \tilde{y}| < \delta \Rightarrow \sup_{u \in \mathcal{U}_0} |G(y|u) - G(\tilde{y}|u)| < \epsilon$. For each $y \in \mathcal{Y}_0$, $G(y \mid \cdot)$ has all partial derivatives up to order r_1 where $r_1 \geq 2$ is an even integer.

(iii) Let $D^{\mathbf{j}} G(y|u) \equiv \partial^{|\mathbf{j}|} G(y|u) / \partial^{j_1} u_1 ... \partial^{j_d} u_d$ where $u = (u_1, ..., u_d)'$. For each $y \in \mathcal{Y}_0$, $D^{\boldsymbol{j}} G(y \mid \cdot)$ with $|\boldsymbol{j}| = r_1$ is uniformly bounded and Lipschitz continuous on $\mathcal{U}_0$: for all u, $\tilde{u} \in \mathcal{U}_0$, $|D^{\boldsymbol{j}} G(y \mid u) - D^{\boldsymbol{j}} G(y \mid \tilde{u})| \leq C_3 ||u - \tilde{u}||$ for some $C_3 \in (0, \infty)$ where $\|\cdot\|$ is the Euclidean norm.

(iv) For each $u \in \mathcal{U}_0$ and for all $y, \tilde{y} \in \mathcal{Y}_0$, $|D^{\boldsymbol{j}} G(y \mid u) - D^{\boldsymbol{j}} G(\tilde{y} \mid u)| \leq C_4 |y - \tilde{y}|$ for some $C_4 \in (0, \infty)$ where $|\boldsymbol{j}| = r_1$.

Assumption A.4 The joint PDF $g(w)$ of W_i has all r_2th partial derivatives that are uniformly continuous on $\mathcal{Y}_0 \times \mathcal{U}_0$ where $r_2 \geq 2$ is an even integer.

Assumption A.5 (i) The distribution function $H(z)$ admits a PDF $h(z)$ that is continuous on $\mathcal{Z}_0$.

(ii) The weight function $a(\cdot, \cdot)$ is a nonnegative function that is uniformly bounded on its compact support $\mathcal{X}_0 \times \mathcal{Y}_0$.

Assumption A.6 (i) For some even integer $r_1 \geq 2$, the kernel K is a product kernel of the bounded symmetric kernel $k : \mathbb{R} \to \mathbb{R}$ satisfying $\int_{\mathbb{R}} v^i k(v) dv = \delta_{i0}$ $(i = 0, 1, ..., r_1 - 1)$, $\int_{\mathbb{R}} v^{r_1} k(v) dv < \infty$, and $k(v) = O\left((1 + |v|^{r_1+1+\epsilon})^{-1}\right)$ for some $\epsilon > 0$, where δ_{ij} is Kronecker's delta.

(ii) For some even integer $r_2 \geq 2$, the kernel L is a product kernel of the bounded symmetric kernel $l : \mathbb{R} \to \mathbb{R}$ satisfying $\int_{\mathbb{R}} v^i l(v) dv = \delta_{i0}$ $(i = 0, 1, ..., r_2 - 1)$, $\int_{\mathbb{R}} v^{r_2} l(v) dv < \infty$, and $l(v) = O\left((1 + |v|^{r_2+1+\epsilon})^{-1}\right)$ for some $\epsilon > 0$.

Assumption A.7 As $n \to \infty$, $b \to 0$, $c \to 0$, and the following conditions are satisfied:

(i) $nc^{d_X+1} / \log n \to \infty$, $nc^{2r_2+d_X+\frac{1}{2}} \to 0$, $nb^{2r_1} c^{d_X+\frac{1}{2}} \to 0$,

(ii) $nb^{d_X} c / (\log n)^2 \to \infty$, $nb^{2(d_X+d_Z)} c^{-d_X-\frac{1}{2}} / (\log n)^2 \to \infty$,

(iii) $c^{\frac{1}{2}} (c/b)^{d_X} \left[1 + c^{2r_2} \log n\right] \to 0$, and $b^{2r_1} c^{-\frac{1}{2}} \log n \to 0$.

A.1-A.3 parallel Assumptions C.1-C.3 in **HSW**. As in **HSW**, the IID requirement in A.1 is standard in cross-section studies but can be relaxed to allow for weakly dependent time series observations. A.2-A.4 and A.6 are standard for nonparametric local constant estimation of conditional CDF and PDF when a higher order kernel may be called upon. Note that we permit the use of higher order kernel for either

K and L but neither is necessary if $d = d_X + d_Z$ is small, see the discussions below. A.5 specifies the weak conditions on the probability weight H and the weight function $a(\cdot,\cdot)$. In the simulations we simply choose H to be a scaled beta distribution that has a compact support $\mathcal{Z}_0$ and specify a as an indicator function with compact support $\mathcal{X}_0 \times \mathcal{Y}_0$. A.7 appropriately restricts the choices of bandwidth sequences and the orders of kernel functions.

Note that if we choose $b = c \propto n^{-1/\alpha}$ for some $\alpha > 0$, then A.7(iii) is automatically satisfied and A.7(i)-(ii) would require $nb^{d_X+1}/(\log n)^2 \to \infty$, $nb^{2r_2+d_X+\frac{1}{2}} \to 0$, $nb^{2r_1+d_X+\frac{1}{2}} \to 0$, and $nb^{d_X+2d_Z-\frac{1}{2}}/(\log n)^2 \to \infty$. The last set of conditions are met provided

$$d_X + 2d_Z - \frac{1}{2} < \alpha < d_X + \frac{1}{2} + 2\min(r_1, r_2). \tag{3.1}$$

Apparently, (3.1) requires $\min(r_1, r_2) > d_Z - \frac{1}{2}$. In the case where $d_Z = 1$ or 2, we can choose $r_1 = r_2 = 2$ and $\alpha \in (d_X + 2d_Z - \frac{1}{2},\ d_X + \frac{9}{2})$ such that (3.1) is satisfied. In this case, there is no need to use higher order kernels for either K or L.

More generally, we can consider choosing $b \propto n^{-1/\alpha}$ and $c \propto n^{-\kappa/\alpha}$. Then A.7 would require

$$\max\left\{(d_X+1)\kappa,\ d_X+\kappa,\ 2(d_X+d_Z)-(d_X+\frac{1}{2})\kappa\right\} < \alpha < \min\left\{d_X+\frac{1}{2}+2r_2,\ (d_X+\frac{1}{2})\kappa+2r_1\right\}$$

where $d_X/\left(d_X+\frac{1}{2}\right) < \kappa < 4r_1$. Due to the "curse of dimensionality" in nonparametric estimation, we expect that typical values of d_X and d_Z are 1, 2, or 3 such that $d_X + d_Z \leq 4$ for realistic applications, in which case we can verify that the above conditions can be satisfied for a variety of combinations for α, κ, r_1 and r_2. In particular, to ensure the conditional CDF estimate $\hat{G}_b(y|x,z)$ to lie between zero and 1 and to be monotone in y, it is always possible to restrict our attention to the use of a second order kernel for K (i.e., $r_1 = 2$) for properly chosen α, κ and r_2. In particular, if $d_Z \leq 2$, we recommend using the same second order kernel for K and L (implying that $r_1 = r_2 = 2$) and setting $b = c \propto n^{-1/\alpha}$. So one only needs to choose a single bandwidth.

3.2 Asymptotic null distribution

In this section, we study the asymptotic behavior of the test statistic in (2.11). To state the next result, we write $\tilde{w} \equiv (\tilde{y}, \tilde{x}', \tilde{z}')'$ and introduce the following notation:

$$\zeta_0(W_i, W_j) \equiv \int g(Y_i|X_i,z)^{-1}\left[g(\bar{x},z)^{-1}\bar{L}_{cj,(\varepsilon_i,\bar{x},z)} - g(X_i,z)^{-1}\bar{L}_{cj,(Y_i,X_i,z)}\right]dH(z) \tag{3.2}$$

and

$$\varphi(w,\tilde{w}) \equiv E\left[\zeta_0(W_i,w)\,\zeta_0(W_i,\tilde{w})\,a(X_i,Y_i)\right], \tag{3.3}$$

where $\bar{L}_{ci,w} = L_{ci,w} - E(L_{ci,w})$, and $L_{ci,w} = L_c(W_i - w)$.[7] We define the asymptotic bias and variance respectively by

$$\mathbb{B}_n \equiv n^{-1}c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\varphi(W_i,W_i) \text{ and } \sigma_n^2 = 2c^{2d_X+1}E[\varphi(W_1,W_2)^2].$$

[7] Even though X_i, Z_i, Y_i, and ε_i all enter the definition of ζ_0, we can still use $W_i = \left(Y_i, X_i', Z_i'\right)'$ to summarize these variables because $\varepsilon_i = m^{-1}(X_i, Y_i)$ is measurable under Assumption I.1 and the continuity of $m(\cdot,\cdot)$.

The following theorem establishes the asymptotic null distribution of the $\hat{J}_n$ test statistic.

Theorem 3.1 *Suppose Assumptions I.1-I.4 and A.1-A.7 hold. Then under* $\mathbb{H}_0$, *we have* $nc^{d_X+\frac{1}{2}}\hat{J}_n - \mathbb{B}_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$, *where* $\sigma_0^2 \equiv \lim_{n\to\infty} \sigma_n^2$.

The proof of the above theorem is quite involved. In fact, we prove it as a special case of Theorem 3.2 studied below. After a long and arduous effort, we can demonstrate that the key building block in obtaining the asymptotic bias and variance of the test statistic $\hat{J}_n$ is $\zeta_0(W_i, W_j)$. The first term, $g(Y_i|X_i, z)^{-1} g\left(\bar{x}, z\right)^{-1} \bar{L}_{cj,(\varepsilon_i,\bar{x},z)}$, in the definition of ζ_0 reflects the influence of the numerator estimator $\hat{g}_c(\varepsilon_i|\bar{x}, z)$ in the definition of $\hat{J}_n$ in (2.11), whereas the second term $g(Y_i|X_i, z)^{-1} g\left(X_i, z\right)^{-1} \bar{L}_{cj,(Y_i,X_i,z)}$ embodies the effect of the denominator estimator $\hat{g}_c(Y_i|X_i, z)$. Like the test statistic in **HSW**, these two terms contribute to the asymptotic bias of $\hat{J}_n$ symmetrically but to the asymptotic variance asymmetrically due to different roles played by $\bar{x}$ (the normalization point) and X_i (data). A careful analysis of $\mathbb{B}_n$ indicates that both terms contribute to the asymptotic bias of $\hat{J}_n$ to the order of $O\left(c^{-1/2}\right)$. On the other hand, a detailed study of σ_n^2 shows that they contribute asymmetrically to the asymptotic variance: the asymptotic variance of $\hat{J}_n$ is mainly determined by the numerator estimator, whereas the role played by the denominator estimator is asymptotically negligible. See **HSW** for further discussion of similar phenomena in a different context. They also explain why we need $c^{d_X+\frac{1}{2}}$ instead of the usual term $c^{(d_X+1)/2}$ as the normalization constant in the front of $\hat{J}_n$, which unavoidably reduces the size of the class of local alternatives that this test has power to detect.

To implement, we need consistent estimates of the asymptotic bias and variance. Let

$$\hat{\zeta}_0\left(W_i, W_k\right) \equiv \int \hat{g}_c(Y_i|X_i, z)^{-1} \left[\hat{g}_c\left(\bar{x}, z\right)^{-1} \hat{L}_{cj,(\hat{\varepsilon}_i,\bar{x},z)} - \hat{g}_c\left(X_i, z\right)^{-1} \hat{L}_{cj,(Y_i,X_i,z)}\right] dH\left(z\right)$$

where $\hat{L}_{cj,w} = L_{cj,w} - \frac{1}{n}\sum_{k=1}^{n} L_{ck,w}$, and $\hat{g}_c\left(x, z\right)$ is a kernel estimator of the PDF $g\left(x, z\right)$ by using kernel L and bandwidth c. We propose estimating the asymptotic bias $\mathbb{B}_n$ by

$$\hat{\mathbb{B}}_n = n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^{n}\sum_{j=1}^{n} \left[\hat{\zeta}_0\left(W_i, W_j\right)\right]^2 a\left(X_i, Y_i\right)$$

and the asymptotic variance σ_n^2 by

$$\hat{\sigma}_n^2 = \frac{2c^{2d_X+1}}{n^2} \sum_{i=1}^{n}\sum_{j=1}^{n} \left[\frac{1}{n}\sum_{l=1}^{n} \hat{\zeta}_0\left(W_l, W_i\right) \hat{\zeta}_0\left(W_l, W_j\right) a\left(X_l, Y_l\right)\right]^2.$$

It is tedious but straightforward to show $\hat{\mathbb{B}}_n - \mathbb{B}_n = o_P\left(1\right)$ and $\hat{\sigma}_n^2 - \sigma_n^2 = o_P\left(1\right)$. Then the following feasible test statistic

$$T_n \equiv \left(nc^{d_X+\frac{1}{2}}\hat{J}_n - \hat{\mathbb{B}}_n\right) / \sqrt{\hat{\sigma}_n^2} \tag{3.4}$$

is asymptotically distributed as $N\left(0, 1\right)$ and we reject the null for large value of T_n.

3.3 Local power property and consistency

To study the local power of the T_n test, consider the sequence of Pitman local alternatives:

$$\mathbb{H}_1(\gamma_n) : D_e m(x,e) = 1 + \gamma_n \delta_n(x,e), \tag{3.5}$$

where $\gamma_n \to 0$ as $n \to \infty$, and δ_n is a non-constant measurable function with $\mu_0 \equiv \lim_{n\to\infty} E[\delta_n (X_1, \varepsilon_1)^2 a(X_1, Y_1)] < \infty$.

Theorem 3.2 *Suppose Assumptions I.1-I.4 and A.1-A.7 hold. Then under* $\mathbb{H}_1(\gamma_n)$ *with* $\gamma_n = n^{-1/2} c^{-d_X/2-1/4}$, $T_n \xrightarrow{d} N(\mu_0/\sigma_0, 1)$. *That is, the asymptotic local power function of* T_n *is given by* $P(T_n > z|\mathbb{H}_1(\gamma_n)) = 1 - \Phi(z - \mu_0/\sigma_0)$, *where* Φ *is the standard normal CDF.*

Theorem 3.2 implies that the T_n test has non-trivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}c^{-d_X/2-1/4}$, provided $0 < \mu_0 < \infty$. As remarked above, this rate is different from the usual nonparametric rate $n^{-1/2}c^{-(d_X+1)/4}$ or $n^{-1/2}c^{-(d_X+d_Z+1)/4}$ when (d_X+1) or (d_X+d_Z+1) dimensional nonparametric objects need to be estimated.

The following theorem shows that the test is consistent.

Theorem 3.3 *Suppose Assumptions I.1-I.4 and A.1-A.7 hold. Suppose that* $\mu_A \equiv E\{[D_e m(X_i, \varepsilon_i) - 1]^2 a(X_i, Y_i)\} > 0$. *Then* $P(T_n > \lambda_n) \to 1$ *as* $n \to \infty$ *for any nonstochastic sequence* $\lambda_n = o(nc^{d_X+1/2})$.

3.4 A bootstrap version of the test

It is well known that nonparametric tests based on their asymptotic normal null distributions may perform poorly in finite samples. As an alternative, we can rely on bootstrap to obtain the bootstrap p-value.

To obtain the bootstrap replicates of $W_i = (Y_i, X_i', Z_i')'$, we need to impose various restrictions. First, we need to impose the identification conditions given in Assumptions I.1 and I.3. Simple resampling bootstrap does not impose these conditions and is thus not applicable. Fortunately, we can follow the local smooth bootstrap procedure of **HSW** (see also Su and White (2008)) to impose these identification conditions. Second, we need to impose the null of additive separability. In view of the discussion in Section 2.2, under $\mathbb{H}_0$ and Assumption I.2, we have

$$m(x,e) = \bar{m}_1(x) + e$$

for some measurable function $\bar{m}_1$ whose exact structure depend on the choice of the normalization point $\bar{x}$. This motivates us to estimate $\bar{m}_1(x)$ by

$$\hat{m}_{1,b}(x) = \int \hat{m}_b(x,e) dQ(e)$$

where $Q(\cdot)$ is a proper CDF on $\mathbb{R}$. Then $\hat{m}_{1,b}(x)$ is consistent for $\bar{m}_1(x) + \int e dQ(e)$ provided $\hat{m}_b(x,e)$ is consistent for $m(x,e)$. The last claim can be established as in **HSW** and the term $\int e dQ(e)$ is constant, which does not affect the asymptotic distribution of our bootstrap test statistic if we generate the bootstrap data Y_i^* through this relationship. See Step 3 below.

Let $\mathcal{W}_n \equiv \{W_i\}_{i=1}^n$. Following Su and White (2008) and **HSW**, we draw bootstrap resamples $\{X_i^*, Y_i^*, Z_i^*\}_{i=1}^n$ based on the following smoothed local bootstrap procedure:

1. For $i = 1, ..., n$, obtain a preliminary estimate of ε_i as $\hat{\varepsilon}_i = \int \hat{G}_b^{-1}(\hat{G}_b(Y_i|X_i, z) \mid \bar{x}, z)dH(z)$.

2. Draw a bootstrap sample $\{Z_i^*\}_{i=1}^n$ from the smoothed kernel density $\tilde{f}_Z(z) = n^{-1}\sum_{i=1}^n \mathbf{\Phi}_{\alpha_z}$ $(Z_i - z)$, where $\mathbf{\Phi}_\alpha(z) = \alpha^{-d_Z}\mathbf{\Phi}(z/\alpha)$ where $\mathbf{\Phi}(\cdot)$ is a product kernel formed from the standard normal PDF $\phi(\cdot)$, and $\alpha_z > 0$ is a bandwidth parameter.

3. For $i = 1, ..., n$, given Z_i^*, draw X_i^* and ε_i^* independently from the smoothed conditional density $\tilde{f}_{X|Z}(x|Z_i^*) = \sum_{j=1}^n \mathbf{\Phi}_{a_x}(X_j - x)\mathbf{\Phi}_{\alpha_z}(Z_j - Z_i^*) / \sum_{l=1}^n \mathbf{\Phi}_{\alpha_z}(Z_l - Z_i^*)$ and $\tilde{f}_{\varepsilon|Z}(e|Z_i^*) = \sum_{j=1}^n \mathbf{\Phi}_{\alpha_e}$ $(\hat{\varepsilon}_j - e)\Phi_{\alpha_z}(Z_j - Z_i^*) / \sum_{l=1}^n \mathbf{\Phi}_{\alpha_z}(Z_l - Z_i^*)$, respectively, where α_z, α_x, and α_e are given bandwidths.[8]

4. For $i = 1, ..., n$, generate the bootstrap analogue of Y_i as $Y_i^* = \hat{m}_{1,b}(X_i^*) + \varepsilon_i^*$.

5. Compute a bootstrap statistic T_n^* in the same way as T_n with $\{(Y_i^*, X_i^*, Z_i^*)\}_{i=1}^n$ replacing $\mathcal{W}_n$.

6. Repeat Steps 2-5 B times to obtain bootstrap test statistics $\left\{T_{nj}^*\right\}_{j=1}^B$. Calculate the bootstrap p-values $p^* \equiv B^{-1}\sum_{j=1}^B 1\left(T_{nj}^* \geq T_n\right)$ and reject the null hypothesis if p^* is smaller than the prescribed nominal level of significance.

4 Monte Carlo Simulations

In this section, we conduct a small set of Monte Carlo simulations to examine the finite sample performance of our test. We first consider the following two data generating processes (DGPs) for the level study:

DGP 1: $Y_i = X_i + \varepsilon_i$,

DGP 2: $Y_i = \Phi(X_i) - \frac{1}{2} + \varepsilon_i$,

where $i = 1, \ldots, n$, $\Phi(\cdot)$ is the standard normal CDF, $X_i = 0.25 + Z_i - 0.25Z_i^2 + v_{1i}$, $\varepsilon_i = 0.5Z_i + v_{2i}$ and Z_i, v_{1i} and v_{2i} are IID $N(0,1)$ and mutually independent. Clearly, the error terms in DGPs 1-2 are additively separable and we use the above two DGPs to evaluate the finite sample level behavior of our test. Note that

$$m(x,e) = \begin{cases} x + e & \text{in DGP 1,} \\ \Phi(x) - \frac{1}{2} + e & \text{in DGP 2.} \end{cases}$$

In both designs, $m(x,\cdot)$ is strictly monotone for each x and $m(\bar{x}, e) = e$ for $\bar{x} = 0$. The other two identification conditions used throughout the paper are easily verified.

To study the finite sample power behavior of our test, we consider the following four DGPs:

DGP 3: $Y_i = (0.5 + 0.1X_i^2)\varepsilon_i$,

DGP 4: $Y_i = \Phi((X_i + 1)\varepsilon_i/4)(X_i + 1)$,

DGP 5: $Y_i = X_i + \varepsilon_i - \frac{\delta X_i^2}{0.1+\exp(\varepsilon_i)}$,

DGP 6: $Y_i = \Phi(X_i) - \frac{1}{2} + \varepsilon_i - \frac{\delta(\sin X_i)^2}{0.1+\varepsilon_i^3}$,

[8] We abuse the notation $\mathbf{\Phi}$ a little bit here: $\mathbf{\Phi}_\alpha(z) = \alpha^{-d_Z}\mathbf{\Phi}(z/\alpha)$ and $\mathbf{\Phi}_\alpha(x) = \alpha^{-d_X}\mathbf{\Phi}(z/\alpha)$. So the argument of $\mathbf{\Phi}$ can be of dimension d_X or d_Z. The bandwidths here are all set according to the Silverman's rule of thumb in our simulations below.

DGP 7: $Y_i = X_i + \varepsilon_i + \frac{\delta X_i}{0.1+\exp(\varepsilon_i)}$,
DGP 8: $Y_i = \Phi(X_i) - \frac{1}{2} + \varepsilon_i + \frac{\delta \sin X_i}{0.1+\varepsilon_i^2}$,
where $i = 1, \ldots, n$, X_i, ε_i and the instrument Z_i are generated as in DGPs 1-2, and δ is a parameter that adapts the corresponding DGP for different simulation purposes.

DGPs 3 and 4 are used by **HSW** to test for the monotonicity in the unobservable (ε_i here). It is easy to verify that the identification conditions specified in Assumptions I.1-I.4 are all satisfied for DGPs 3-4. But these two DGPs do not satisfy the additive separability condition.

When $\delta = 0$, DGPs 5 and 7 (resp. DGPs 6 and 8) reduce to DGP 1 (resp. DGP 2). For other values of δ, the structural function $m(x, e)$ implied by DGPs 5-8 is not additively separable in error terms. In addition, DGPs 5 and 6 satisfy all the identification conditions specified in Assumptions I.1-I.4; DGPs 7 and 8 violate Assumption I.1 but satisfies the other identification conditions (e.g., $m(0, e) = e$ regardless of the value of δ in DGPs 7-8). It is worth mentioning that when monotonicity is violated as in DGPs 7 and 8, the structural function $m(\cdot, \cdot)$ is generally not identified so that Lemma 2.1 in this paper does not apply and our test is not applicable either.[9] The inclusion of these two DGPs aims to investigate whether our test still has some power in the case where the maintained key identifying assumption (monotonicity) is not satisfied.

To construct our standardized test statistic T_n in (3.4), we need to compute sequentially $\hat{J}_n$, $\hat{\mathbb{B}}_n$ and $\hat{\sigma}_n$. We first obtain local constant estimates $\hat{G}_b(y|u)$, $\hat{G}_b^{-1}(\tau|u)$, $\hat{g}_c(y|u)$ and $\hat{\varepsilon}_i = \int \hat{G}_b^{-1}(\hat{G}_b(Y_i|X_i, z) \mid \bar{x}, z) dH(z)$ by using standard normal kernel function and Silverman's rule of thumb for bandwidth choice, i.e., $b = c = \left(1.06 S_X n^{-1/5}, 1.06 S_Z n^{-1/5}\right)$ with S_X and S_Z being the sample standard deviation of $\{X_i\}$ and $\{Z_i\}$, respectively. We choose $H(z)$ to be a scaled beta(3,3) distribution on $\left[\zeta_\kappa, \zeta_{1-\kappa}\right]$, where ζ_κ denotes the κ-th sample quantile of $\{Z_i\}$ and $\kappa = 0.05$. $N = 30$ evenly-spaced points are chosen for numerical integration. We set $a(X_i, Y_i) = 1\left\{\zeta_{\lambda,X} \le X_i \le \zeta_{1-\lambda,X}\right\} \times 1\left\{\zeta_{\lambda,Y} \le Y_i \le \zeta_{1-\lambda,Y}\right\}$, where, e.g., $\zeta_{\lambda,X}$ is the λ-th sample quantile of $\{X_i\}$ and $\lambda = 0.0125$. Note that we only establish the asymptotic theory for the case where the trimming function $a(\cdot, \cdot)$ has a fixed support $\mathcal{X}_0 \times \mathcal{Y}_0$. But since the sample quantiles converge to their population analogue at the parametric rate, we conjecture that the asymptotic theories established above continue to be valid for our data-driven choice of the weighting function. For the computation of $\hat{\mathbb{B}}_n$ and $\hat{\sigma}_n$, we need to further compute $\hat{g}_c(x, z)$ with a standard normal kernel function and bandwidth c chosen as before. The same trimming function $a(X_i, Y_i)$ and weight function $H(z)$ are utilized everywhere.

To obtain the bootstrap p-values, we follow the procedure stated in Section 3.4 to compute the rejection probabilities. We consider two sample sizes ($n = 100$ and 200) with 250 replications. Due to the high computational burden, we only use $B = 100$ bootstrap resamples in each replication. Before conducting the bootstrap with $B = 100$, we study the sensitivity of the test to the bandwidth b as suggested by Giacomini, Politis and White (2007), using the warp-speed bootstrap procedure based on a single bootstrap resample. We find that the our test is not very sensitive to the choice of $b = \left(c_1 S_X n^{-1/5}, c_1 S_Z n^{-1/5}\right)$ as long as c_1 is between 1 and 2. We report the results for $c_1 = 1.06$. In

[9] Hoderlein and Mammen (2007) show that an average over the marginal effects can be identified without the monotonicity assumption.

Table 1: Empirical level for DGPs 1-2

DGP	n	1%	5%	10%
1	100	0.008	0.044	0.116
	200	0.012	0.048	0.112
2	100	0.004	0.040	0.088
	200	0.008	0.052	0.108

Table 2: Empirical power for DGPs 3-8

DGP	n	1%	5%	10%
3	100	0.908	0.992	0.996
	200	0.980	0.996	1
4	100	0.996	1	1
	200	1	1	1
5	100	0.896	0.912	0.984
	200	0.924	0.952	0.992
6	100	0.872	0.904	0.952
	200	0.916	0.936	0.988
7	100	0.440	0.468	0.492
	200	0.484	0.524	0.572
8	100	0.476	0.544	0.648
	200	0.504	0.584	0.692

addition, we consider $\delta = 1$ in DGPs 5-8.

Table 1 reports the empirical level of our bootstrapped test for DGPs 1-2 where the nominal levels are 1%, 5% and 10%. We see that the level of our test is fairly well behaved and it gets closer to the nominal level as the sample size increases. Table 2 presents the empirical power of our bootstrapped test at various nominal levels. Surprisingly our test has fantastic power to reject additive separability for DGPs 3-4. The power is also reasonably good and increases as the sample size doubles in DGPs 5-8. Comparing the results for DGPs 7-8 with DGPs 5-6, we observe that the power performance of our test is adversely affected by the violation of the monotonicity assumption. This is interesting as our test is designed to test for additivity by maintaining the monotonicity hypothesis - a key identifying assumption in the literature on nonparametric nonseparable models. In the general case, if one rejects the null hypothesis, one may argue that either the null hypothesis or the monotonicity hypothesis may be violated. Our limited simulation results here suggest that the violation of both may not enhance the power of our test. An obvious reason for this is that our test, by construction, is only designed to test for the violation of additivity, and it has no power against the violation of monotonicity.

5 Concluding Remarks

The prevalent additive error structure has been an important assumption in many economic and econometric models. This paper develops a simple consistent test to detect whether this critical assumption holds in the presence of economic data. The test is motivated from the simple observation that the partial derivative of the unknown structural function with respect to the unobserved error term is one under the null hypothesis of additive separability and certain identification conditions. We derive the asymptotic distributions of our test statistic under the null and a sequence of Pitman local alternatives and prove its consistency. We also propose a bootstrap version of the test. Monte Carlo simulations are conducted to examine the finite sample performance of the bootstrapped test. The test enjoys appropriate size and reasonable power in finite samples.

There are some interesting topics for further research. First, under the same set of identification conditions considered in this paper, one can develop other tests for additive separability. For example, one may consider a test based on the observation that the cross derivatives with respect to the regressor and the error term is zero under additivity. But this would need consistent estimate of cross derivatives and thus is expected to be less powerful. For another example, we can consider the estimation of the structural function under both the null and the alternative, and base a test on the weighted L_2 distance between these estimates. To this goal, one needs to develop an estimate of the structural function under the additive separability condition. Under Assumption I.3 and the null: $m(X, \varepsilon) = \bar{m}_1(X) + \varepsilon$, $E(Y|X, Z) = \bar{m}_1(X) + E(\varepsilon|Z)$. This motivates us to obtain a consistent estimate $\tilde{m}_1(x)$ of $\bar{m}_1(x)$ by using either the marginal integration or backfitting technique. Then we can compare this estimate with $\hat{m}_{1,b}(x)$ used in Section 3.4. The theoretical study of this test is left for future research.

Second, one may consider relaxing some of the identification conditions used to identify and estimate the nonparametric structural function under the alternative. For example, one may follow **LW** and relax the monotonicity assumption. The problem is that without monotonicity, one cannot identify $m(x, e)$ or its partial derivative with respect to e under the alternative without further assumptions. It is interesting to know whether it is possible to develop a consistent test in this case. Alternatively, one may consider relaxing the conditional exogeneity condition: $X \perp \varepsilon \mid Z$. Again, without this assumption, one cannot identify $m(x, e)$ or its partial derivative as in this paper. Some other assumptions have to be in place.

Appendix

A Proof of Some Technical Lemmas

In this appendix, we prove some technical lemmas that are used in the establishment of the main results in Section 3.

Recall that $\mathcal{U}_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0$, $U_i \equiv (X_i', Z_i')'$, $u \equiv (x', z')'$, $W_i \equiv (Y_i, U_i')'$ and $w \equiv (y, u')'$. Let $1_i \equiv 1\{X_i \in \mathcal{X}_0, Y_i \in \mathcal{Y}_0\}$. Define

$$\begin{aligned} V_{n,b}(y;u) &\equiv \frac{1}{n}\sum_{i=1}^n K_b(U_i - u)\left[1\{Y_i \leq y\} - G(y|U_i)\right] = \frac{1}{n}\sum_{i=1}^n K_{bi,u}\bar{\mathbf{1}}_i(y), \\ \mathbf{V}_{n,c}(y;u) &\equiv \frac{1}{n}\sum_{i=1}^n \{L_c(W_i - w) - E\left[L_c(W_i - w)\right]\} = \frac{1}{n}\sum_{i=1}^n \bar{L}_{ci,w}, \end{aligned}$$

where $K_{bi,u} \equiv K_b(U_i - u)$, $L_{ci,w} = L_c(W_i - w)$, $\bar{L}_{ci,w} = L_{ci,w} - E(L_{ci,w})$, and $\bar{\mathbf{1}}_i(y) = 1\{Y_i \leq y\} - G(y|U_i)$. Let

$$\nu_{1b} \equiv n^{-1/2} b^{-d_X/2}\sqrt{\log n},\ \nu_{2b} \equiv n^{-1/2} b^{-(d_X+d_Z)/2}\sqrt{\log n},\ \text{and } \nu_{3b} \equiv n^{-1/2} b^{-(d_X+d_Z+1)/2}\sqrt{\log n}.$$

ν_{1c}, ν_{2c}, and ν_{3c} are similarly defined.

Lemma A.1 *Suppose that Assumptions A.1-A.3, A.6(i) and A.7 hold. Let $\mathcal{T}_0 = [\underline{\tau}, \bar{\tau}]$ denote a closed interval of $(0,1)$. Then*

(a) $\hat{G}_b(y|u) - G(y|u) = g(u)^{-1} V_{n,b}(y;u) + O_P(\nu_{2b}^2 + b^{r_1})$ *uniformly in* $(y,u) \in \mathbb{R} \times \mathcal{U}_0$,

(b) $\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u) = O_P(\nu_{2b} + b^{r_1})$ *uniformly in* $(\tau,u) \in \mathcal{T}_0 \times \mathcal{U}_0$,

(c) $\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u) = -\frac{V_{n,b}(G^{-1}(\tau|u);u)\{1+o(1)\}}{g(G^{-1}(\tau|u)|u)g(u)} + O_P(\nu_{2b}^2 + b^{r_1})$ *uniformly in* $(\tau,u) \in \mathcal{T}_0 \times \mathcal{U}_0$.

Proof. For (a), we make the following bias-variance decomposition:

$$\hat{G}_b(y|u) - G(y|u) = \hat{g}_b(u)^{-1}\frac{1}{n}\sum_{i=1}^n K_b(U_i - u)\left[G(y|U_i) - G(y|u)\right] + \hat{g}_b(u)^{-1} V_{n,b}(y;u)$$

By Assumptions A.1-A.3 and A.6(i) and the standard arguments in kernel estimation (e.g., Masry (1996a, 1996b), Hansen (2008)), $\sup_{u\in\mathcal{U}_0}|\hat{g}_b(u) - g(u)| = O_P(\nu_{2b} + b^{r_1})$, $\sup_{u\in\mathcal{U}_0}|\frac{1}{n}\sum_{i=1}^n K_b(U_i - u)$ $[G(y|U_i) - G(y|u)]| = O_P(b^{r_1})$, and $\sup_{u\in\mathcal{U}_0}|V_{n,b}(y;u)| = O_P(\nu_{2b})$. It follows that uniformly in $u \in \mathcal{U}_0$,

$$\hat{G}_b(y|u) - G(y|u) = g(u)^{-1} V_{n,b}(y;u) + O_P\left(b^{r_1} + \nu_{2b}^2\right).$$

By the same argument as used in the proof of Theorem 4.1 of Boente and Fraiman (1991), we can show that the last result also holds uniformly in $y \in \mathbb{R}$ under Assumption A.3.

For (b), noting that $\hat{G}_b\left(\hat{G}_b^{-1}(\tau|u)|u\right) = \tau = G\left(G^{-1}(\tau|u)|u\right)$, we have

$$\left|G\left(\hat{G}_b^{-1}(\tau|u)|u\right) - G\left(G^{-1}(\tau|u)|u\right)\right| = \left|G\left(\hat{G}_b^{-1}(\tau|u)|u\right) - \hat{G}_b\left(\hat{G}_b^{-1}(\tau|u)|u\right)\right| \leq \sup_{y\in\mathbb{R}}\left|G(y|u) - \hat{G}_b(y|u)\right|.$$

So the pointwise consistency of $\hat{G}_b^{-1}(\tau|u)$ follows from that of $\hat{G}_b(y|u)$ and the continuity of $G(\cdot|u)$. By Assumption A.3(ii) and the first order Taylor expansion,

$$G\left(\hat{G}_b^{-1}(\tau|u)|u\right) - G\left(G^{-1}(\tau|u)|u\right) = \left[\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u)\right] g\left(\tilde{G}^{-1}(\tau|u)|u\right)$$

where $\tilde{G}^{-1}(\tau|u)$ lies between $\hat{G}_b^{-1}(\tau|u)$ and $G^{-1}(\tau|u)$. Therefore by (a) and Assumption A.2(ii)

$$\begin{aligned}\sup_{(\tau,u)\in\mathcal{T}_0\times\mathcal{U}_0}\left|\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u)\right| &\leq \frac{\sup_{(\tau,u)\in\mathcal{T}_0\times\mathcal{U}_0}\left|G\left(\hat{G}_b^{-1}(\tau|u)|u\right) - G\left(G^{-1}(\tau|u)|u\right)\right|}{\inf_{(\tau,u)\in\mathcal{T}_0\times\mathcal{U}_0} g\left(\tilde{G}^{-1}(\tau|u)|u\right)} \\ &\leq \frac{\sup_{u\in\mathcal{U}_0}\sup_{y\in\mathbb{R}}\left|G(y|u) - \hat{G}_b(y|u)\right|}{\inf_{(\tau,u)\in\mathcal{T}_0\times\mathcal{U}_0} g\left(\tilde{G}^{-1}(\tau|u)|u\right)} = O_P(\nu_{2b} + b^{r_1}).\end{aligned}$$

To obtain the uniform Bahadur representation for $\hat{G}_b^{-1}(\tau|u)$, we apply the Hadamard differentiability of the (conditional) quantile operator (see e.g., Doss and Gill (1992, Theorem 1)) to obtain

$$\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u) = \frac{\hat{G}_b(G^{-1}(\tau|u)|u) - \tau}{g\left(G^{-1}(\tau|u)|u\right)}\left\{1 + o(1)\right\}.$$

This together with (a) implies that $\hat{G}_b^{-1}(\tau|u) - G^{-1}(\tau|u) = -\frac{V_{n,b}(G^{-1}(\tau|u);u)\{1+o(1)\}}{g(G^{-1}(\tau|u)|u)g(u)} + O_P(\nu_{2b}^2 + b^{r_1})$. ■

If $G(y|x,z) \in \mathcal{T}_0 = [\underline{\tau}, \bar{\tau}] \subset (0,1)$ for $(y,x,z) \in \mathcal{Y}_0 \times \mathcal{X}_0 \times \mathcal{Z}_0$, by Lemma A.1($a$) $\hat{G}_b(y|x,z) \in \mathcal{T}_0^\epsilon$ with probability approaching 1 (w.p.a.1) as $n \to \infty$, where $\mathcal{T}_0^\epsilon \equiv [\underline{\tau} - \epsilon, \bar{\tau} + \epsilon] \subset (0,1)$ for some $\epsilon > 0$. Note that the result in Lemma A.1(c) also holds uniformly in $(\tau, u) \in \mathcal{T}_0^\epsilon \times \mathcal{U}_0$ w.p.a.1.

Lemma A.2 *Suppose that Assumptions A.1-A.4, A.6 and A.7 hold. Then*

(a) $\sup_{\tilde{y},y\in\mathcal{Y}_0,|\tilde{y}-y|\leq M(v_{2b}+b^{r_1})}\sup_{u\in\mathcal{U}_0}\sqrt{nc^{d_X+1/2}}\left\|V_{n,b}(\tilde{y};u) - V_{n,b}(y;u)\right\| = o_P(1)$;

(b) $\sup_{\tilde{y},y\in\mathcal{Y}_0,|\tilde{y}-y|\leq M(v_{2b}+b^{r_1})}\sup_{u\in\mathcal{U}_0}\sqrt{nc^{d_X+1/2}}\left\|\mathbf{V}_{n,c}(\tilde{y};u) - \mathbf{V}_{n,c}(y;u)\right\| = o_P(1)$.

Proof. The proof is analogous to that of Lemma A.3 in **HSW** and thus omitted. ■

Lemma A.3 *Suppose that Assumptions A.1-A.4, A.6 and A.7 hold. Then for any* $\delta_n = O(\nu_{2b} + b^{r_1})$, *we have*

(a) $\hat{G}_b(a+\delta_n|u) - \hat{G}_b(a|u) = g(a|u)\delta_n + o_P\left(n^{-1/2}c^{-d_X/2-1/4}\right)$ *uniformly in* $u \in \mathcal{U}_0$,

(b) $\hat{G}_b^{-1}(a+\delta_n|u) - \hat{G}_b^{-1}(a|u) = g(G^{-1}(a|u)|u)^{-1}\delta_n + o_P\left(n^{-1/2}c^{-d_X/2-1/4}\right)$ *uniformly in* $u \in \mathcal{U}_0$,

Proof. By Lemma A.1(a), $\hat{G}_b(a+\delta_n|u) - \hat{G}_b(a|u) = [G(a+\delta_n|u) - G(a|u)] + g(u)^{-1}[V_{n,b}(a+\delta_n;u) - V_{n,b}(a;u)] + O_P(\nu_{2b}^2 + b^{r_1})$. By Assumption A.4 and Taylor expansions, the first term on the right hand side of the last expression is $g(a|u)\delta_n + O(\delta_n^2)$. By Lemma A.2($a$), $V_{n,b}(a+\delta_n;u) - V_{n,b}(a;u) = o_P(n^{-1/2}c^{-d_X/4-1/4})$ uniformly in $u \in \mathcal{U}_0$. Thus (a) follows by Assumption A.7. The proof of (b) is analogous and thus omitted. ■

Lemma A.4 *Suppose Assumptions A.1-A.4, A.6 and A.7 hold. Then uniformly in* $(y,u) \in \mathcal{Y}_0 \times \mathcal{U}_0$,

(a) $\hat{g}_c(y|u) - g(y|u) = g(u)^{-1}\mathbf{V}_{n,c}(y;u) + O_P(c^{r_2} + \nu_{3c}^2)$,

(b) $\mathbf{V}_{n,c}(y;u) = O_P(\nu_{3c})$,

(c) $\hat{g}_c(y+\delta_n|u) - \hat{g}_c(y|u) = D_y g(y|u)\delta_n + o_P\left(n^{-1/2}c^{-d_X/2-1/4}\right)$ *for any* $\delta_n = O(\nu_{2b} + b^{r_1})$,

where $D_y g(y|u) \equiv \partial(g(y|u))/\partial y$.

Proof. Recall $W_i \equiv (Y_i, U_i')'$ and $w \equiv (y, u')'$. We make the following bias-variance decomposition:

$$\hat{g}_c(y|u) - g(y|u) = \hat{g}_c(u)^{-1} \frac{1}{n} \sum_{i=1}^{n} \{g(y,u) - E[L_c(W_i - w)]\} + \hat{g}_c(u)^{-1} \mathbf{V}_{n,c}(y;u)$$

By Assumptions A.1, A.4 and A.6(ii) and the standard arguments in kernel estimation, $\sup_{u \in \mathcal{U}_0} |\hat{g}_c(u) - g(u)| = O_P(\nu_{2c} + c^{r_2})$, $\sup_{w \in \mathcal{W}_0} |\frac{1}{n}\sum_{i=1}^n E[L_c(W_i - w)] - g(y,u)| = O_P(c^{r_2})$, and $\sup_{w \in \mathcal{W}_0} |\mathbf{V}_{n,c}(y;u)| = O_P(\nu_{3c})$. Thus ($a$) and ($b$) follow. Furthermore, $\hat{g}_c(y+\delta_n|u) - \hat{g}_c(y|u) = [g(y+\delta_n|u) - g(y|u)] + g(u)^{-1}[\mathbf{V}_{n,c}(y+\delta_n;u) - \mathbf{V}_{n,c}(y;u)] + O_P(c^{r_2} + \nu_{3c}^2)$. Then ($c$) follows from Taylor expansions and Lemma A.2(b). ■

Lemma A.5 *Suppose that Assumptions A.1-A.4, A.6 and A.7 hold. Then uniformly in* i,

(a) $(\hat{\varepsilon}_i - \varepsilon_i) 1_i = s_{\varepsilon n,i} 1_i \{1 + o(1)\} + o_P(n^{-1/2} c^{-d_X/2 - 1/4})$,

(b) $(\hat{\varepsilon}_i - \varepsilon_i) 1_i = O_P(v_{1b} + b^{r_1})$,

where $s_{\varepsilon n,i} = \int \frac{1}{g(G^{-1}(\tau_{iz}|\bar{x},z)|\bar{x},z)} [-\frac{V_{n,b}(G^{-1}(\tau_{iz}|\bar{x},z);\bar{x},z)}{g(\bar{x},z)} + \frac{V_{n,b}(Y_i;X_i,z)}{g(X_i,z)}] dH(z)$ *and* $\tau_{iz} \equiv G(Y_i|X_i,z)$.

Proof. Let $\hat{\tau}_{iz} \equiv \hat{G}_b(Y_i|X_i,z)$. Then $(\hat{\varepsilon}_i - \varepsilon_i) 1_i = \varepsilon_{1i} + \varepsilon_{2i}$, where

$$\begin{aligned} \varepsilon_{1i} &\equiv \left[\int \hat{G}_b^{-1}(\hat{\tau}_{iz}|\bar{x},z) dH(z) - \int G^{-1}(\hat{\tau}_{iz}|\bar{x},z) dH(z)\right] 1_i, \text{ and} \\ \varepsilon_{2i} &\equiv \left[\int G^{-1}(\hat{\tau}_{iz}|\bar{x},z) dH(z) - \int G^{-1}(\tau_{iz}|\bar{x},z) dH(z)\right] 1_i. \end{aligned}$$

By Lemmas A.1 and A.2,

$$\begin{aligned} \varepsilon_{1i} &= -\int \frac{V_{n,b}\left(G^{-1}(\hat{\tau}_{iz}|\bar{x},z);\bar{x},z\right)\{1+o(1)\}}{g\left(G^{-1}(\hat{\tau}_{iz}|\bar{x},z)|\bar{x},z\right) g(\bar{x},z)} dH(z) 1_i + O_P(\nu_{2b}^2 + b^{r_1}) \\ &= -\int \frac{V_{n,b}\left(G^{-1}(\tau_{iz}|\bar{x},z);\bar{x},z\right)\{1+o(1)\}}{g\left(G^{-1}(\tau_{iz}|\bar{x},z)|\bar{x},z\right) g(\bar{x},z)} dH(z) 1_i + o_P(n^{-1/2} c^{-d_X/2 - 1/4}), \end{aligned}$$

and

$$\begin{aligned} \varepsilon_{2i} &= \int g(G^{-1}(\tau_{iz}|\bar{x},z)|\bar{x},z)^{-1} (\hat{\tau}_{iz} - \tau_{iz})\, dH(z) 1_i + O_p\left(b^{2r_1} + v_{2b}^4\right) \\ &= \int g(G^{-1}(\tau_{iz}|\bar{x},z)|\bar{x},z)^{-1} g(X_i,z)^{-1} V_{n,b}(Y_i;X_i,z)\, dH(z) 1_i + o_P(n^{-1/2} c^{-d_X/2 - 1/4}). \end{aligned}$$

Combining these results yields (a). (b) follows from (a) and the standard arguments as used in showing $\sup_{u \in \mathcal{U}_0} |V_{n,b}(y;u)| = O_P(\nu_{2b})$. ■

Lemma A.6 *Suppose that Assumptions A.1-A.4, A.6 and A.7 hold. Then*

(a) $\alpha_1(x,e) \equiv \int \left[\frac{\hat{g}_c(e|\bar{x},z)}{\hat{g}_c(y|x,z)} - \frac{g(e|\bar{x},z)}{g(y|x,z)}\right] dH(z) = s_{1n}(x,e) + O_P(c^{r_2} + n^{-1} c^{-(d_X+1)} \log n)$ *uniformly in* $(e,x) \in \mathcal{E}_0 \times \mathcal{X}_0$,

(b) $\alpha_{2i} \equiv \int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z) - \hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)} dH(z) 1_i = s_{2i} 1_i + o_P(n^{-1/2} c^{-d_X/2 - 1/4})$ *uniformly in* i,

where $y = m(x,e)$, $s_{1n}(x,e) = \int \frac{1}{g(y|x,z)} \left[\frac{\mathbf{V}_{n,c}(e;\bar{x},z)}{g(\bar{x},z)} - D_e m(x,e) \frac{\mathbf{V}_{n,c}(y;x,z)}{g(x,z)}\right] dH(z)$, *and* $s_{2n,i} = \int \frac{D_y g(\varepsilon_i|\bar{x},z)}{g(Y_i|X_i,z)} dH(z) s_{\varepsilon n,i}$.

Proof. First, observe that $\alpha_1(x,e) = \alpha_{11}(x,e) + \alpha_{12}(x,e)$, where $\alpha_{11}(x,e) = \int \hat{g}_c(y|x,z)^{-1}[\hat{g}_c(e|\bar{x},z) - g(e|\bar{x},z)]dH(z)$, and $\alpha_{12}(x,e) = \int g(e|\bar{x},z)[\hat{g}_c(y|x,z)^{-1} - g(y|x,z)^{-1}]dH(z)$. By Lemma A.4($a$), we can show that

$$\begin{aligned}\int [\hat{g}_c(y|x,z) - g(y|x,z)]^2 \, dH(z) &= \int g(x,z)^{-2} \, \mathbf{V}_{n,c}(y;x,z)^2 \, dH(z) + O_P(c^{2r_2} + \nu_{3c}^4) \\ &= O_P(v_{1c}^2 c^{-1} + c^{2r_2}) \text{ uniformly in } (y,x) \in \mathcal{Y}_0 \times \mathcal{X}_0. \qquad \text{(A.1)}\end{aligned}$$

By Lemma A.4(a), equation (A.1) and the Cauchy-Schwarz inequality, we have that uniformly in $(e,x) \in \mathcal{E}_0 \times \mathcal{X}_0$

$$\begin{aligned}\alpha_{11}(x,e) &= \int g(y|x,z)^{-1} \left[\hat{g}_c(e|\bar{x},z) - g(e|\bar{x},z)\right] dH(z) \\ &\quad - \int \hat{g}_c(y|x,z)^{-1} g(y|x,z) \left[\hat{g}_c(y|x,z) - g(y|x,z)\right] \left[\hat{g}_c(e|\bar{x},z) - g(e|\bar{x},z)\right] dH(z) \\ &= \int g(y|x,z)^{-1} \left[\hat{g}_c(e|\bar{x},z) - g(e|\bar{x},z)\right] dH(z) + O_P(n^{-1}c^{-(d_X+1)}\log n + c^{2r_2}) \\ &= \int g(y|x,z)^{-1} g(\bar{x},z)^{-1} \, \mathbf{V}_{n,c}(e;\bar{x},z) \, dH(z) + O_P(c^{r_2} + \nu_{3c}^2),\end{aligned}$$

and

$$\begin{aligned}\alpha_{12}(x,e) &= -\int g(e|\bar{x},z)\hat{g}_c(y|x,z)^{-1} g(y|x,z)^{-1}[\hat{g}_c(y|x,z) - g(y|x,z)]dH(z) \\ &= -\int g(e|\bar{x},z) g(y|x,z)^{-2}[\hat{g}_c(y|x,z) - g(y|x,z)]dH(z) + O_P(n^{-1}c^{-(d_X+1)}\log n + c^{2r_2}) \\ &= -\int g(e|\bar{x},z) g(y|x,z)^{-2} g(x,z)^{-1} \, \mathbf{V}_{n,c}(y;x,z) \, dH(z) + O_P(n^{-1}c^{-(d_X+1)}\log n + c^{r_2}).\end{aligned}$$

Then by equation (2.2) we have that uniformly in $(e,x) \in \mathcal{E}_0 \times \mathcal{X}_0$

$$\begin{aligned}\alpha_1(x,e) &= \int \frac{1}{g(y|x,z)} \left[\frac{\mathbf{V}_{n,c}(e;\bar{x},z)}{g(\bar{x},z)} - D_e m(x,e) \frac{\mathbf{V}_{n,c}(y;x,z)}{g(x,z)}\right] dH(z) + O_P(c^{r_2} + n^{-1}c^{-(d_X+1)}\log n) \\ &= s_{1n}(x,e) + O_P(c^{r_2} + n^{-1}c^{-(d_X+1)}\log n).\end{aligned}$$

For (b), note that $\alpha_{2i} = \alpha_{21i} - \alpha_{22i}$, where $\alpha_{21i} = \int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z) - \hat{g}_c(\varepsilon_i|\bar{x},z)}{g(Y_i|X_i,z)} dH(z) \, 1_i$, and $\alpha_{22i} = \int \frac{[\hat{g}_c(Y_i|X_i,z) - g(Y_i|X_i,z)][\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z) - \hat{g}_c(\varepsilon_i|\bar{x},z)]}{\hat{g}_c(Y_i|X_i,z) g(Y_i|X_i,z)} dH(z) \, 1_i$. By Lemmas A.4($c$) and A.5($a$),

$$\alpha_{21i} = \int \frac{D_y g(\varepsilon_i|\bar{x},z)}{g(Y_i|X_i,z)} dH(z) \, (\hat{\varepsilon}_i - \varepsilon_i) \, 1_i + o_P(n^{-1/2}c^{-d_X/2-1/4}) = s_{2n,i}1_i + o_P(n^{-1/2}c^{-d_X/2-1/4}),$$

where $o_P(n^{-1/2}c^{-d_X/2-1/4})$ holds uniformly in i. By Assumption A.2, Lemmas A.4(a)-(b) and A.5(b), and equation (A.1), we have that uniformly in i

$$\begin{aligned}\alpha_{22i} &= \int \frac{[\hat{g}_c(Y_i|X_i,z) - g(Y_i|X_i,z)]\,[\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z) - \hat{g}_c(\varepsilon_i|\bar{x},z)]}{g(Y_i|X_i,z) g(Y_i|X_i,z)} dH(z) \, 1_i \, \{1 + o(1)\} \\ &= \int |\hat{g}_c(Y_i|X_i,z) - g(Y_i|X_i,z)| \, dH(z) O_P(v_{1b} + b^{r_1}) \\ &= \left[\int \{[\hat{g}_c(Y_i|X_i,z) - g(Y_i|X_i,z)]\}^2 \, dH(z)\right]^{1/2} 1_i O_P(v_{1b} + b^{r_1}) \\ &= O_P(n^{-1/2}c^{-(d_X+1)/2}\sqrt{\log n} + c^{r_2}) O_P(v_{1b} + b^{r_1}) = o_P(n^{-1/2}c^{-d_X/2-1/4}).\end{aligned}$$

It follows that $\alpha_{2i} = s_{2n,i}1_i + o_P(n^{-1/2}c^{-d_X/2-1/4})$ uniformly in i. ■

B Proof of the Main Results

Proofs of Theorem 3.1 and 3.2

We only prove Theorem 3.2 as the proof of Theorem 3.1 is a special case. To conserve space, let $a_i \equiv a(X_i, Y_i)$. We first make the following decomposition:

$$\begin{aligned} nc^{d_X+\frac{1}{2}}\hat{J}_n &= c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\left\{\int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z)-\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}\,dH(z)+\int\left[\frac{\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}-1\right]dH(z)\right\}^2 a_i \\ &= c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\left\{\int\left[\frac{\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}-1\right]dH(z)\right\}^2 a_i \\ &\quad + c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\left[\int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z)-\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}\,dH(z)\right]^2 a_i \\ &\quad + 2c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\int \frac{\hat{g}_c(\hat{\varepsilon}_i|\bar{x},z)-\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}\,dH(z)\int\left[\frac{\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}-1\right]dH(z)a_i \\ &\equiv \hat{J}_{n1}+\hat{J}_{n2}+2\hat{J}_{n3},\ \text{say.} \end{aligned}$$

Propositions B.1, B.2, and B.3 study $\hat{J}_{n1}, \hat{J}_{n2}$, and $\hat{J}_{n3}$, respectively. Combining the results in these propositions yields $nc^{d_X+\frac{1}{2}}\hat{J}_n = J_n + \mu_0 + o_P(1)$, where $J_n = c^{d_X+\frac{1}{2}}\sum_{i=1}^{n} s_{n,i}^2 a_i$,

$$s_{n,i} = \int g_{1iz}^{-1}\left[\frac{\mathbf{V}_{n,c}(\varepsilon_i;\bar{x},z)}{g(\bar{x},z)} - \frac{\mathbf{V}_{n,c}(Y_i;X_i,z)}{g(X_i,z)}\right]dH(z) = n^{-1}\sum_{j=1}^{n}\zeta_0(W_i,W_j), \tag{B.1}$$

$g_{1iz} = g(Y_i|X_i,z)$, and $\zeta_0(W_i,W_j) = \int g_{1iz}^{-1}\left[g(\bar{x},z)^{-1}\bar{L}_{cj,(\varepsilon_i,\bar{x},z)} - g(X_i,z)^{-1}\bar{L}_{cj,(Y_i,X_i,z)}\right]dH(z)$ is as defined in (3.2). The rest of the proof follows that of **HSW** closely.

First, using ζ_0, we can write J_n as a third order V-statistic:

$$J_n = c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\left[n^{-1}\sum_{j=1}^{n}\zeta_0(W_i,W_j)\right]^2 a_i = n^{-2}c^{d_X+\frac{1}{2}}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\zeta(W_{i_1},W_{i_2},W_{i_3}),$$

where $\zeta(W_{i_1},W_{i_2},W_{i_3}) \equiv \zeta_0(W_{i_1},W_{i_2})\zeta_0(W_{i_1},W_{i_3})a_{i_1}$. To study the asymptotic distribution of J_n, we need to use the U-statistic theory (e.g., Lee (1990)). Let $\varphi(w_{i_1},w_{i_2}) \equiv E[\zeta(W_1,w_{i_1},w_{i_2})]$, and $\bar{\zeta}(w_{i_1},w_{i_2},w_{i_3}) \equiv \zeta(w_{i_1},w_{i_2},w_{i_3}) - \varphi(w_{i_2},w_{i_3})$. Then we can decompose J_n as follows

$$\begin{aligned} J_n &= n^{-1}c^{d_X+\frac{1}{2}}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\varphi(W_{i_1},W_{i_2}) + n^{-2}c^{d_X+\frac{1}{2}}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3}) \\ &\equiv J_{1n}+J_{2n},\ \text{say.} \end{aligned}$$

Consider J_{2n} first. Write $E(J_{n2}^2) = n^{-4}c^{2d_X+1}\sum_{i_1,\ldots,i_6=1}^{n}E\left[\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})\right]$. Observing that $E\left[\bar{\zeta}(W_{i_1},w_{i_2},w_{i_3})\right] = E\left[\bar{\zeta}(w_{i_1},W_{i_2},w_{i_3})\right] = E[\bar{\zeta}(w_{i_1},w_{i_2},W_{i_3})] = 0$, $E[\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})] = 0$ if there are more than three distinct elements in $\{i_1,\ldots,i_6\}$. In view of this, we can show that

$$E(J_{2n}^2) = O(n^{-1}c^{-d_X-1} + n^{-2}c^{-2d_X-1} + n^{-3}c^{-2d_X-2}) = o(1).$$

Then $J_{2n} = o_P(1)$ by the Chebyshev inequality.

For J_{1n}, let $\varphi(W_i, W_j) = \int \zeta_0(\tilde{w}, W_i)\, \zeta_0(\tilde{w}, W_j)\, a\left(\tilde{x}, m^{-1}(\tilde{x}, \tilde{y})\right) dG(\tilde{w})$, where $G(\cdot)$ is the CDF of W_i.Then $J_{1n} = \mathbb{B}_n + \mathbb{V}_n$, where $\mathbb{B}_n = n^{-1} c^{d_X} \sum_{i=1}^n \varphi(W_i, W_i)$ and $\mathbb{V}_n = 2n^{-1} c^{d_X + \frac{1}{2}} \sum_{1 \le i < j \le n} \varphi(W_i, W_j)$ contribute to the asymptotic bias and variance of our test statistic, respectively. Observing that $\mathbb{V}_n$ is a second-order degenerate U-statistic, we can easily verify that all the conditions of Theorem 1 of Hall (1984) are satisfied and a central limit theorem applies to it: $\mathbb{V}_n \xrightarrow{d} N\left(0, \sigma^2\right)$, where $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$ and $\sigma_n^2 = 2c^{2d_X+1} E\left[\varphi(W_1, W_2)\right]^2$.[10] ■

Proposition B.1 $\hat{J}_{n1} = c^{d_X + \frac{1}{2}} \sum_{i=1}^n s_{n,i}^2 a_i + \mu_0 + o_P(1)$ *under* $\mathbb{H}_1(\gamma_n)$.

Proof. To begin with, we decompose $\hat{J}_{n1}$ as follows:

$$
\begin{aligned}
\hat{J}_{n1} &= c^{d_X+\frac{1}{2}} \sum_{i=1}^n \left\{ \int \left[\frac{\hat{g}_c(\varepsilon_i|\bar{x}, z)}{\hat{g}_c(Y_i|X_i, z)} - \frac{g(\varepsilon_i|\bar{x}, z)}{g(Y_i|X_i, z)} \right] dH(z) \right\}^2 a_i \\
&\quad + c^{d_X+\frac{1}{2}} \sum_{i=1}^n \left\{ \int \left[\frac{g(\varepsilon_i|\bar{x}, z)}{g(Y_i|X_i, z)} - 1 \right] dH(z) \right\}^2 a_i \\
&\quad + 2c^{d_X+\frac{1}{2}} \sum_{i=1}^n \int \left[\frac{\hat{g}_c(\varepsilon_i|\bar{x}, z)}{\hat{g}_c(Y_i|X_i, z)} - \frac{g(\varepsilon_i|\bar{x}, z)}{g(Y_i|X_i, z)} \right] dH(z) \int \left[\frac{g(\varepsilon_i|\bar{x}, z)}{g(Y_i|X_i, z)} - 1 \right] dH(z) a_i \\
&= J_{n11} + J_{n12} + 2J_{n13}.
\end{aligned}
$$

Using Lemma A.6 and the fact that $D_e m(x, e) = 1 + \gamma_n \delta_n(x, e)$ under $\mathbb{H}_1(\gamma_n)$ (see (3.5)), we can show that

$$
\begin{aligned}
J_{n11} &= c^{d_X+\frac{1}{2}} \sum_{i=1}^n s_{1n,i}^2 a_i + n c^{d_X+\frac{1}{2}} O_P((c^{r_2} + n^{-1} c^{-(d_X+1)} \log n)^2) \\
&= c^{d_X+\frac{1}{2}} \sum_{i=1}^n s_{n,i}^2 a_i + o_P(1)
\end{aligned}
$$

where $s_{1n,i} = s_{1n}(X_i, \varepsilon_i)$ and $s_{n,i}$ is defined in (B.1). By (2.2), (3.5), and the weak law of large numbers (WLLN), we have

$$
\begin{aligned}
J_{n12} &= c^{d_X+\frac{1}{2}} \sum_{i=1}^n \left[D_e m(X_i, \varepsilon_i) - 1 \right]^2 a_i = c^{d_X+\frac{1}{2}} \sum_{i=1}^n \gamma_n^2 \delta_n(X_i, \varepsilon_i)^2 a_i + o_P(1) \\
&= n^{-1} \sum_{i=1}^n \delta_n(X_i, \varepsilon_i)^2 a_i + o_P(1) \xrightarrow{P} \lim_{n\to\infty} E\left[\delta_n(X_i, \varepsilon_i)^2 a(X_i, Y_i) \right] \equiv \mu_0.
\end{aligned}
$$

For J_{n13}, by Lemma A.6 and (3.5), we have

$$
\begin{aligned}
J_{n13} &= c^{d_X+\frac{1}{2}} \sum_{i=1}^n \int \left[\frac{\hat{g}_c(\varepsilon_i|\bar{x}, z)}{\hat{g}_c(Y_i|X_i, z)} - \frac{g(\varepsilon_i|\bar{x}, z)}{g(Y_i|X_i, z)} \right] dH(z) \gamma_n \delta_n(X_i, \varepsilon_i) a_i \\
&= \gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i s_{1n,i} \delta_n(X_i, \varepsilon_i) + n \gamma_n c^{d_X+\frac{1}{2}} O_P(c^{r_2} + n^{-1} c^{-(d_X+1)} \log n) \\
&= \bar{J}_{n13} + o_P(1),
\end{aligned}
$$

[10] Write $\zeta_0(W_i, W_j) = \int g_{1iz}^{-1} g(\bar{x}, z)^{-1} \bar{L}_{cj,(\varepsilon_i, \bar{x}, z)} dH(z) - \int g_{1iz}^{-1} g(X_i, z)^{-1} \bar{L}_{cj,(Y_i, X_i, z)} dH(z) \equiv \zeta_{1ij} - \zeta_{2ij}$, say. A careful calculation suggests that both ζ_{1ij} and ζ_{2ij} contribute to the asymptotic bias of J_{1na} but only ζ_{1ij} contributes to the asymptotic variance of J_{1na}.

where $\bar{J}_{n13} \equiv \gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i s_{n,i} \delta_n(X_i, \varepsilon_i)$. Note that $\bar{J}_{n13} = \bar{J}_{n131} + \bar{J}_{n132}$, where $\bar{J}_{n13s} = n^{-1}\gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1}^n a_i \zeta_{sij} \delta_n(X_i, \varepsilon_i)$ for $s = 1, 2$, where ζ_{1ij} and ζ_{2ij} are defined in footnote B. We further write $\bar{J}_{n131} = n^{-1}\gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i \zeta_{1ii} \delta(X_i, \varepsilon_i) + n^{-1}\gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j\neq i}^n a_i \zeta_{1ij} \delta_n(X_i, \varepsilon_i)$. It is easy to show that the first term is $O_P(\gamma_n c^{d_X+\frac{1}{2}})$ and the second term is $O_P(c^{1/4})$ by moment calculations. It follows that $\bar{J}_{n131} = o_P(1)$. Similarly, $\bar{J}_{n132} = n^{-1}\gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i \zeta_{2ii} \delta_n(X_i, \varepsilon_i) + n^{-1}\gamma_n c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j\neq i}^n a_i \zeta_{2ij} \delta_n(X_i, \varepsilon_i) = O_P(\gamma_n c^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}} c^{-\frac{1}{4}} + c^{\frac{1}{2}d_X+\frac{1}{4}}) = o_P(1)$. It follows that $J_{n13} = o_P(1)$.

Combining the above results yields the desired result: $\hat{J}_{n1} = c^{d_X+\frac{1}{2}} \sum_{i=1}^n s_{n,i}^2 a_i + \mu_0 + o_P(1)$. ■

Proposition B.2 $\hat{J}_{n2} = o_P(1)$ *under* $\mathbb{H}_1(\gamma_n)$.

Proof. By Lemma A.6(b) and the Cauchy-Schwarz inequality, we have

$$\hat{J}_{n2} \leq 2c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i s_{2n,i}^2 + 2nc^{d_X+\frac{1}{2}} o_P((n^{-1/2}c^{-d_X/2-1/4})^2) = 2J_{n2} + o_P(1),$$

where $J_{n2} \equiv c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i \beta_i^2 s_{\varepsilon n,i}^2$ and $\beta_i \equiv \int \frac{D_y g(\varepsilon_i|\bar{x},z)}{g(Y_i|X_i,z)} dH(z)$. Let $g_{2iz} \equiv g(G^{-1}(\tau_{iz}|\bar{x},z)|\bar{x},z)$ where recall $\tau_{iz} = G(Y_i|X_i,z)$. Then

$$\begin{aligned} J_{n2} &= c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i \beta_i^2 \left[\frac{1}{n} \sum_{j=1}^n \left(-\eta_{1ij} + \eta_{2ij} \right) \right]^2 \\ &= n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_i \beta_i^2 \left(\eta_{1ij}\eta_{1ik} + \eta_{2ij}\eta_{2ik} - 2\eta_{1ij}\eta_{2ik} \right) \\ &= J_{n21} + J_{n22} + J_{n23}, \text{ say,} \end{aligned}$$

where $\eta_{1ij} \equiv \int g_{2iz}^{-1} g(\bar{x},z)^{-1} K_{bj,(\bar{x},z)} \bar{\mathbf{1}}_j \left(G^{-1}(\tau_{iz}|\bar{x},z) \right) dH(z)$, $\eta_{2ij} \equiv \int g_{2iz}^{-1} g(X_i,z)^{-1} K_{bj,(X_i,z)} \bar{\mathbf{1}}_j (Y_i)\, dH(z)$, and e.g., $J_{n211} \equiv n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_i \beta_i^2 \eta_{1ij}\eta_{1ik}$. For J_{n21}, we decompose it as follows:

$$\begin{aligned} J_{n21} &= n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1, j\neq i}^n \sum_{k=1, k\neq i,j}^n a_i \beta_i^2 \eta_{1ij}\eta_{1ik} + n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1, j\neq i}^n a_i \beta_i^2 \eta_{1ij}^2 \\ &\quad + 2n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n \sum_{j=1, j\neq i}^n a_i \beta_i^2 \eta_{1ii}\eta_{1ij} + n^{-2} c^{d_X+\frac{1}{2}} \sum_{i=1}^n a_i \beta_i^2 \eta_{1ii}^2 \\ &\equiv J_{n211} + J_{n212} + J_{n213} + J_{n214}, \text{ say.} \end{aligned}$$

In view of $E(J_{n211a}) = 0$, $E\left(J_{n211}^2\right) = O\left(c^{2d_X+1} b^{-2d_X}\right)$, $E|J_{n212}| = E(J_{n212}) = O(c^{d_X+\frac{1}{2}} b^{-d_X})$, $E|J_{n213}| = O(c^{d_X+\frac{1}{2}} b^{-d_X})$, and $E|J_{n214}| = O(n^{-1} c^{d_X+\frac{1}{2}} b^{-2d_X})$, we have $J_{n21} = O_P(c^{d_X+\frac{1}{2}} b^{-d_X} + n^{-1} c^{d_X+\frac{1}{2}} b^{-2d_X}) = o_P(1)$ by the Chebyshev and Markov inequalities. By the same token, we can show that $J_{n22} = o_P(1)$. Then $J_{n23} = o_P(1)$ by the Cauchy-Schwarz inequality. Consequently, we have shown that $J_{n2} = o_P(1)$. ■

Proposition B.3 $\hat{J}_{n3} = o_P(1)$ *under* $\mathbb{H}_1(\gamma_n)$.

Proof. Following the proof of Propositions B.1 and B.2, we can show that

$$\begin{aligned}\hat{J}_{n3} &= c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}s_{2n,i}\left[s_{1n,i}+\gamma_n\delta(X_i,\varepsilon_i)\right]a_i+o_P\left(1\right)\\ &= c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}s_{2n,i}s_{1n,i}a_i+\gamma_n c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}s_{2n,i}\delta(X_i,\varepsilon_i)a_i+o_P\left(1\right)\\ &\equiv J_{n31}+J_{n32}+o_P\left(1\right),\text{ say.}\end{aligned}$$

We prove the lemma by demonstrating that $J_{n31}=o_P\left(1\right)$ and $J_{n32}=o_P\left(1\right).$ Recall $\zeta_{1ij}\equiv\int g_{1iz}^{-1}g\left(\bar{x},z\right)^{-1}$ $\bar{L}_{cj,(\varepsilon_i,\bar{x},z)}dH\left(z\right)$ and $\zeta_{2ij}\equiv\int g_{1iz}^{-1}g\left(X_i,z\right)^{-1}\bar{L}_{cj,(Y_i,X_i,z)}dH\left(z\right).$ Let $\bar{\zeta}_{2ij}=D_e m\left(X_i,\varepsilon_i\right)\zeta_{2ij}$. Then

$$\begin{aligned}J_{n31} &= c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}a_i\frac{1}{n}\sum_{j=1}^{n}\beta_i\left(-\eta_{1ij}+\eta_{2ij}\right)\frac{1}{n}\sum_{k=1}^{n}\left(\zeta_{1ik}-\bar{\zeta}_{2ik}\right)\\ &= n^{-2}n^{-2}c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}a_i\beta_i\left(-\eta_{1ij}\zeta_{1ik}+\eta_{1ij}\bar{\zeta}_{2ik}+\eta_{2ij}\zeta_{1ik}-\eta_{2ij}\bar{\zeta}_{2ik}\right)\\ &\equiv -J_{n311}+J_{n312}+J_{n313}-J_{n314},\text{ say.}\end{aligned}$$

As in the analysis of J_{n211}, we can readily demonstrate that $J_{n31s}=o_P\left(1\right)$ by straightforward moment calculations and the Chebyshev/Markov inequalities. Thus $J_{n31}=o_P\left(1\right).$ Note that $J_{n32}=J_{n321}+J_{n322}$, where $J_{n32s}=n^{-1}\gamma_n c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\sum_{j=1}^{n}a_i\beta_i\eta_{sij}\delta_n(X_i,\varepsilon_i)$ for $s=1,2$. We further write $J_{n321}=n^{-1}\gamma_n c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}a_i\beta_i\eta_{1ii}\delta_n(X_i,\varepsilon_i)+n^{-1}\gamma_n c^{d_X+\frac{1}{2}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}a_i\beta_i\eta_{1ij}\delta_n(X_i,\varepsilon_i)$. It is easy to show that the first term is $O_P(\gamma_n c^{d_X+\frac{1}{2}})$ and the second term is $O_P((c/b)^{d_X/2}c^{1/4})$ by moment calculations. It follows that $J_{n321}=o_P\left(1\right).$ Similarly, $J_{n321}=o_P\left(1\right).$ Thus we have shown that $J_{n32}=o_P\left(1\right).$ ■

Proof of Theorem 3.3

The proof is simpler than that of Theorem 3.2. Under $\mathbb{H}_1$, we can readily apply Lemmas A.6, A.5, and the WLLN to obtain

$$\begin{aligned}\hat{J}_n &= n^{-1}\sum_{i=1}^{n}\left\{\int\left[\frac{\hat{g}_c(\varepsilon_i|\bar{x},z)}{\hat{g}_c(Y_i|X_i,z)}-1\right]dH(z)\right\}^2 a_i+o_P\left(1\right)\\ &= n^{-1}\sum_{i=1}^{n}\left\{\int\left[\frac{g(\varepsilon_i|\bar{x},z)}{g(Y_i|X_i,z)}-1\right]dH(z)\right\}^2 a_i+o_P\left(1\right)\\ &= n^{-1}\sum_{i=1}^{n}\left[D_e m\left(X_i,\varepsilon_i\right)-1\right]^2 a_i+o_P\left(1\right)\xrightarrow{P}E\left\{\left[D_e m\left(X_i,\varepsilon_i\right)-1\right]^2 a_i\right\}.\end{aligned}$$

The result follows by noticing that $\hat{\sigma}_n^2=O_P\left(1\right)$ and $\hat{B}_n=o_P\left(nc^{d_X+1/2}\right)$ under $\mathbb{H}_1$. ■

References

Altonji, J. G., Ichimura H., Otsu, T. (2012). Estimating derivatives in nonseparable models with limited dependent variables. *Econometrica* 80: 1701–1719.

Altonji, J. G., Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73:1053-1102.

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

Blundell, R., Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. In: M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications*, *Eight World Congress*, Vol. II, pp.312-357. Cambridge University Press, Cambridge.

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

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

Bierens, H. J., Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* 65: 1129-1151.

Boente, G., Fraiman, R. (1991). Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. *International Statistical Review* 59: 355-372.

Briesch, R., Chintagunta, P., Matzkin, R. L. (2010). Nonparametric discrete choice models with unobserved heterogeneity. *Journal of Business and Economic Statistics* 28: 291-307.

Brown, D. J., Matzkin, R. L. (1998). Estimation of nonparametric functions in simultaneous equations models, with an application to consumer demand. *Mimeo*, Northwestern University.

Chernozhukov, V., Fernández-Val, I., Galichon, A. (2010). Quantile and probability curves without crossing. 2010, *Econometrica* 78: 1093-1125.

Doss, H., Gill, R. D. (1992). An elementary approach to weak convergence for quantile processes with applications to censored survival data. *Journal of the American Statistical Association* 87: 869-877.

Ekeland, I., Heckman, J. J., Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of Political Economy* 112(S1): S60-109.

Giacomini, R., Politis, D. N., White, H. (2007). A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Discussion Paper*, Dept. of Economics, UCSD.

Hall, P. (1984). Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14: 1-16.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24: 726-748.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21: 1926-1947.

Hausman, J. A. (1978). Specification testing in econometrics. *Econometrica* 46: 1251-1271.

Heckman, J. J. (1974). Effects of child-care programs on women's work effort. *Journal of Political Economy* 82: S136-S163.

Heckman, J., Matzkin, R., Nesheim, L. (2010). Nonparametric estimation of nonadditive hedonic models. *Econometrica* 78: 1569-1591.

Heckman, J. J., Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Science* 96: 4730-4734.

Heckman, J. J., Vytlacil, E. J. (2001). Local instrumental variables. In: C. Hsiao, K. Morimune, and J. Powell (eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, pp. 1-146. Cambridge University Press, Cambridge.

Heckman, J. J., Vytlacil, E. J. (2005). Structural equations, treatment effects, and economic policy evaluation. *Econometrica* 73: 669-738.

Heckman, J. J., Willis, R. (1977). A beta-logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy* 85: 27-58.

Hoderlein, S., Mammen, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75: 1513-1518.

Hoderlein, S., Su, L., White, H. (2012). Specification testing for nonparametric structural models with monotonicity in unobservables. *Discussion Paper*, Dept. of Economics, UCSD.

Hong, Y., White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica* 63: 1133-1159.

Hsiao, C., Li, Q., Racine, J. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics* 140: 802–826.

Lancaster, T. (1979). Econometric methods for the analysis of unemployment. *Econometrica* 47: 939-956.

Lee, A. J. (1990). *U-statistics: Theory and Practice.* Taylor & Francis Group, New York.

Li, Q., Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87: 145–165.

Lu, X., White, H. (2012). Testing for separability in structural equations. *Discussion Paper*, Dept. of Economics, UCSD.

Masry, E. (1996a). Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and their Applications* 65: 81–101.

Masry, E. (1996b). Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17: 571–599.

Matzkin, R. L. (1991): A nonparametric maximum rank correlation estimator. In: W. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica* 71: 1339-1375.

Matzkin, R. L. (2007). Nonparametric identification. In J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, Vol. 6b, pp.5307-5368, Elsevier, New York.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (eds.), *Frontiers in Econometrics*, . pp.105-142, Academic Press, New York.

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

Olley, G. S., Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64: 1263-1297.

Robinson, P. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* 56: 511-534.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametric Models. *Econometrica* 56: 433-447.

Ruud, P. A. (1984). Tests of specification in econometrics. *Econometric Reviews* 3: 211-242.

Stinchcombe, M.,White, H. (1998). Consistent specification testing with nuisance parameters present only under alternative. *Econometric Theory* 14: 295-324.

Su, L., Ullah, A. (2012) A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory*, forthcoming.

Su, L., White, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24: 829-864.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30: 415443.

White, H. (1987). Specification testing in dynamic models. In T. F. Bewley (eds.), *Advances in Econometrics,* pp. 1-58. Fifth world congress, Vol. I. Cambridge University Press, Cambridge.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory* 8: 452-475.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory* 8: 435-451.

Zheng, X. (1996). Consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* 75: 263-289.