

Specification Testing for Transformation Models with an Application to Generalized Accelerated Failure-time Models

Arthur Lewbel,^a Xun Lu^b and Liangjun Su^c

^a Department of Economics, Boston College

^b Department of Economics, Hong Kong University of Science and Technology

^c School of Economics, Singapore Management University

May 1, 2013

Abstract

Consider a nonseparable model $Y = R(X, U)$ where Y and X are observed, while U is unobserved and conditionally independent of X . This paper provides the first nonparametric test of whether R takes the form of a transformation model, meaning that Y is monotonic in the sum of a function of X plus a function of U . Transformation models of this form are commonly assumed in economics, including, e.g., standard specifications of duration models and hedonic pricing models. Our test statistic is asymptotically normal under local alternatives and consistent against nonparametric alternatives. Monte Carlo experiments show that our test performs well in finite samples. We apply our results to test for specifications of generalized accelerated failure-time (GAFT) models of the duration of strikes and of marriages.

Keywords: additivity, control variable, endogenous variable, monotonicity, nonparametric nonseparable model, hazard model, specification test, transformation model, unobserved heterogeneity

JEL Classification: C12, C14

Acknowledgements: Halbert White inspired this project, brought us together to work on it, and provided substantial advice, discussion, and enthusiasm. We deeply mourn his passing. We also want to thank Songnian Chen for helpful comments and suggestions. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

1 Introduction

We consider a general nonseparable structural equation

$$Y = R(X, U), \tag{1.1}$$

where Y is a scalar observable outcome, X a $d_x \times 1$ vector of observable covariates of interest, U a $d_u \times 1$ vector of unobservable causes or errors, and R an unknown measurable function. Our goal is to test the

following hypothesis:

\mathbb{H}_{10} : There exist three measurable functions $G : \mathbb{R} \rightarrow \mathbb{R}$, $H_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$ and $H_2 : \mathbb{R}^{d_u} \rightarrow \mathbb{R}$ such that $Y = G[H_1(X) + H_2(U)]$ *a.s.*, and G is strictly monotonic.

\mathbb{H}_{1A} : \mathbb{H}_{10} is false.

Specifications that are monotonic functions of additive models have been called “transformation models” (e.g., Chiappori et al., 2011), or “transformed additively separable models” (e.g., Jacho-Chávez et al., 2010), or “generalized additive models with unknown link function” (e.g., Horowitz, 2001, and Horowitz and Mammen, 2004).

Broadly speaking, there are two kinds of transformation models that are common in the economics literature. The first type assumes that Y and X are observable, U is unobservable, and the link function $G(\cdot)$ may be known or unknown. Our paper belongs to this category. Ridder (1990), Horowitz (1996), Ekeland et al. (2004), Chiappori et al. (2011), and Ichimura and Lee (2011) discuss identification and estimation for transformation models of this category. Since U is unobservable in this class of models, only the functions G and H_1 are identified and estimated. The second kind of transformation model assumes both X and U are observable, and takes Y to be an object that can be estimated like a conditional mean or quantile function. Horowitz (2001), Horowitz and Mammen (2004, 2007, 2011), Horowitz and Lee (2005), and Jacho-Chávez et al. (2010) provide identification and estimation results for this second kind of transformation model, while Gozalo and Linton (2001) consider specification tests for such models. See also Horowitz (2013) for a recent survey on the latter class of models.

The transformation models under our null are commonly used (and hence assumed to hold) in a wide range of economic applications. For example, they are often used to study duration data (see, e.g., Heckman and Singer, 1984, Keifer, 1988, Mata and Portugal, 1994, Engle, 2000, and Abbring et al., 2008), including generalized accelerated failure-time (GAFT) models, which includes accelerated failure-time (AFT) models, proportional hazard (PH) models, and mixed proportional hazard (MPH) models as special cases. The MPH specification in particular is a widely used class of duration data specifications (for a review, see Van den Berg, 2001).

Despite its popularity, economic theory rarely justifies the MPH specification. For example, Van den Berg (2001, p. 3400) points out that “the MPH model specification is not derived from economic theory and it remains to be seen whether the MPH specification is actually able to capture important theoretical relations.” He also provides many specific economic examples where the MPH specification is violated. In their microeconomics textbook, Cameron and Trivedi (2005, p. 613) say that “the multiplicative heterogeneity assumption [in MPH models] is also rather special, but it is mathematically convenient...” Given the popularity (and the limitations) of GAFT models, especially MPH models, it is obvious that a formal specification test of these models would be useful for empirical research. While some specification tests for certain parametric forms of duration models exist (see, e.g., Fernandes and Grammig, 2005), to the best of our knowledge, ours is the first that specifically tests for the general specification of GAFT models.

Another major set of applications of transformation model specifications where U is unobservable are hedonic models (see, e.g., Ekeland et al., 2004, and Heckman et al., 2005). Here again, we believe that our

paper is the first to provide a general specification test for this class of transformation models. Chiappori and Komunjer (2011) discuss a hypothesis similar to ours, but they do not provide specific test statistics.

A conditional exogeneity assumption is imposed to test \mathbb{H}_{10} , i.e., we assume that U and X are conditionally independent, conditioning on an observable covariate vector Z . This is analogous to the conditional unconfoundedness assumption in the treatment effect literature, and to the assumptions required for use of control function type methods of dealing with endogeneity (see, e.g., Blundell and Powell 2003). Chiappori et al. (2011) provide a nonparametric estimator for the transformation model under similar assumptions.¹

We first show that (given some regularity conditions) the data are generated by a transformation model, so \mathbb{H}_{10} holds, if and only if the ratio of the derivatives with respect to Y and to X of the conditional CDF of Y given (X, Z) is a multiplicative function of X and Y .² We then use local polynomial methods to estimate these derivatives, and construct test statistics based on the L_2 distance between restricted and unrestricted estimators of this ratio of derivatives. We show that our test statistic is asymptotically normal under the null and under a sequence of Pitman local alternatives.

To facilitate application of our test, we propose and compare a few different methods of obtaining limit distributions. These are direct estimation of the limiting variance, two different bootstrap methods, and subsampling. We also evaluate our test both in a Monte Carlo setting, and in two different empirical applications. Both applications have data sets with similar numbers of observations and have the same dimension. In the first application, concerning duration of strikes by manufacturing workers, the GAFT model is not rejected, while in a second application, on the duration of first marriages of divorced couples, GAFT and hence also MPH are strongly rejected.

Our null \mathbb{H}_{10} is weaker than additive separability but stronger than monotonicity. Lu and White (2013) and Su et al. (2013) propose tests for additive separability under the same conditional exogeneity assumption we make, i.e., they test whether there exist two unknown measurable functions G_1 and G_2 such that

$$Y = G_1(X) + G_2(U) \text{ a.s.}$$

Testing \mathbb{H}_{10} is more general than testing for separability, since our null is equivalent to additive separability in the special case where G is known to be the identity function. Hence if we reject \mathbb{H}_{10} , then we also reject additive separability.

Hoderlein et al. (2011) (HSW) test for monotonicity under a conditional exogeneity assumption. Let $\tilde{U} \equiv H_2(U)$. HSW test whether there exists a function \tilde{R} such that

$$Y = \tilde{R}(X, \tilde{U})$$

where \tilde{R} is strictly monotonic in its second argument. Our null is stronger than monotonicity, so if the HSW test rejects monotonicity, then our null \mathbb{H}_{10} is also rejected. Our null \mathbb{H}_{10} combines monotonicity

¹Specifically, Chiappori et al. (2011) provide identification (up to some normalizations) and an estimator for the transformation model, assuming the data are generated by this model, while we provide a test for whether this assumption is valid. The model they consider is more general than ours in that we only permit control function type endogeneity, while they allow for more general nonparametric instrumental variables assumptions

²Horowitz (1996) considers the estimation of the semiparametric model under our null, where H_1 takes a parametric form (unlike our nonparametric case) and without covariates Z . His estimator also relies on the implication that the ratio of the derivatives is a multiplicative function of X and Y .

with the additional restriction that the observable X and unobservable \tilde{U} are additively separable under a transformation function G . Our test exploits this additivity restriction, and so should be generically stronger than HSW for testing \mathbb{H}_{10} . Also, the HSW test requires that Z not be empty, while our test of \mathbb{H}_{10} can be applied even if we have no conditioning covariates Z .

Note that in all these models, under the null Y equals a function of X and a scalar unobservable \tilde{U} , e.g., $\tilde{U} \equiv H_2(U)$ or $\tilde{U} \equiv G_2(U)$, but under the alternative U may be a random vector.

The rest of the paper is organized as follows. In Section 2, we propose and motivate our test. In Section 3, we show that our test statistics are asymptotically normal under the null, and we analyze their global and local power. In Section 4, we conduct some Monte Carlo simulations to evaluate the finite sample performance of our test statistics. In Section 5, we provide two empirical applications, testing for the specification of GAFT models in data on the durations of strikes and of first marriages. In Section 6, we discuss extensions to other closely related hypotheses. Section 7 concludes, and mathematical proofs are relegated to the Appendix.

2 A Specification Test for Transformation Models

In this section, we describe implications of \mathbb{H}_{10} that are used to motivate our test construction, and then propose a test statistic.

2.1 Motivation

To construct our test, we first impose a conditional exogeneity assumption. Let $X \perp U \mid Z$ denote that X and U are independent given Z .

Assumption A.1 Let Z be an observable random vector of dimension $d_z \in \mathbb{N}$, such that $X \perp U \mid Z$ and that X and U are not measurable with respect to the sigma-field generated by Z .

Assumption A.1 is equivalent to the unconfoundedness assumption in the treatment effect literature and is widely used to identify causal effects. For detailed discussions, see Altonji and Matzkin (2005), Hoderlein and Mammen (2007), Imbens and Newey (2009), and White and Lu (2011), among others. It is also closely related to the assumptions used to allow for endogeneity in the control function literature, where Z would equal the residuals from a regression of X on exogenous instruments. See, e.g., Blundell and Powell (2003), (2004).

Under \mathbb{H}_{10} , the condition $X \perp U \mid Z$ can be relaxed a bit to $X \perp H_2(U) \mid Z$ in Theorem 2.1(a) below.

Let $F(y \mid x, z) \equiv F_{Y \mid X, Z}(y \mid x, z)$ and $f(y \mid x, z) \equiv f_{Y \mid X, Z}(y \mid x, z)$ denote the conditional cumulative distribution function (CDF) and probability density function (PDF) of Y given $(X, Z) = (x, z)$, respectively. Let $\mathcal{V} \equiv \mathcal{X} \times \mathcal{Z}$ denote the support of $V \equiv (X', Z)'$ and \mathcal{Y} the support of Y . Let $\mathcal{W} \equiv \mathcal{Y} \times \mathcal{V}$. Let $r(y; x, z) \equiv \frac{D_x F(y \mid x, z)}{f(y \mid x, z)}$, so $r(y; x, z)$ is the ratio of two partial derivatives of $F(y \mid x, z)$, since $f(y \mid x, z) = \partial F(y \mid x, z) / \partial y$ and $D_x F(y \mid x, z) \equiv \partial F(y \mid x, z) / \partial x$.

The following theorem characterizes some useful properties of the transformation model under \mathbb{H}_{10} .

Theorem 2.1 *Suppose that $f(y \mid x, z) \neq 0$ for all $(y, x, z) \in \mathcal{W}$.*

(a) If \mathbb{H}_{10} and A.1 hold and the first order (partial) derivatives of G and H_1 exist, then there exist two measurable functions $s_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_x}$ and $s_2 : \mathbb{R} \rightarrow \mathbb{R}_+$ (or $s_2 : \mathbb{R} \rightarrow \mathbb{R}_-$) such that

$$r(Y; X, Z) = s_1(X) s_2(Y) \text{ a.s.}, \quad (2.1)$$

where $s_1(x) = -\partial S_1(x)/\partial x$ for some measurable function $S_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$, and $1/s_2(y) = \partial S_2(y)/\partial y$ for some measurable function $S_2 : \mathbb{R} \rightarrow \mathbb{R}$.

(b) If there exist two measurable functions $s_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_x}$ and $s_2 : \mathbb{R} \rightarrow \mathbb{R}_+$ (or $s_2 : \mathbb{R} \rightarrow \mathbb{R}_-$) such that (2.1) holds, $s_1(x) = -\partial S_1(x)/\partial x$ for some function $S_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$, and $1/s_2(y) = \partial S_2(y)/\partial y$ for some measurable function $S_2 : \mathbb{R} \rightarrow \mathbb{R}$, then \mathbb{H}_{10} holds in the sense that there exist two measurable functions $G : \mathbb{R} \rightarrow \mathbb{R}$ and $H_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$ such that

$$Y = G \left[H_1(X) + \tilde{U} \right] \text{ a.s.} \quad (2.2)$$

where G is strictly monotonic and differentiable, all first order partial derivatives of H_1 exist, and \tilde{U} is a scalar unobservable random variable satisfying $X \perp \tilde{U} \mid Z$.

Remark 2.1 Theorem 2.1(a) says that under \mathbb{H}_{10} and the conditional exogeneity condition in A.1, the ratio $r(y; x, z)$ is free of z and can be factored out as the product of a function s_1 of x and a function s_2 of y , the function s_1 can be written as the derivative of a scalar function, and the function s_2 does not alternate in sign on its support. Theorem 2.1(b) says the converse is also true: as long as the factorization in (2.1) holds with s_1 and s_2 satisfying appropriate conditions, the observables (Y, X, Z) will satisfy the version of transformation model (2.2) under the null. Note that even though U can be a vector in the true data generating process, \tilde{U} is a scalar unobservable here and it satisfies the conditional exogeneity in A.1.

Remark 2.2 Theorem 2.1 gives a characterization of \mathbb{H}_{10} , but it does not by itself provide a test for \mathbb{H}_{10} . The proof of Theorem 2.1(a) shows that s_1 and s_2 in the theorem depend on the unknown functions H_1 and G , respectively, so we cannot directly test equation (2.1). We instead propose a feasible and straightforward test statistic that is based on implications of the factorization in (2.1).

Let $\mathcal{Y}_0 \equiv [y, \bar{y}] \subset \mathcal{Y}$ for finite real numbers y and \bar{y} . Let $\mathbf{1}\{\cdot\}$ denote the indicator function that equals one when \cdot is true and zero otherwise, and let $E_Y(\cdot)$ and $E_{XZ}(\cdot)$ denote expectations with respect to Y and (X, Z) , respectively. Define

$$\begin{aligned} r(y; x, z) &\equiv \frac{D_x F(y \mid x, z)}{f(y \mid x, z)} \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ r_0 &\equiv E_Y E_{XZ} [r(Y; X, Z)], \\ r_1(x) &\equiv E[r(Y; x, Z)], \\ r_2(y) &\equiv E[r(y; X, Z)], \end{aligned} \quad (2.3)$$

where with a little abuse of notation we have redefined $r(y; x, z)$ to denote a trimmed instead of untrimmed ratio of partial derivatives of $F(y \mid x, z)$. Note that r, r_0, r_1 and r_2 are all $d_x \times 1$ vectors and it is easy to see that \mathbb{H}_{10} implies that

$$r(y; x, z) \circ r_0 = r_1(x) \circ r_2(y), \quad (2.4)$$

where \circ denotes the Hadamard product. It would be possible to base a test similar to ours on equation (2.4) directly. However, when $d_x > 1$, comparing equations (2.4) and (2.1), $s_2(y)$ in (2.1) is a scalar, which is not exploited in (2.4). To incorporate the implications of \mathbb{H}_{10} into our test as much as possible, we consider a simple average of $r_2(y)$. We define $\pi \equiv (\pi_1, \dots, \pi_{d_x})'$ as a $d_x \times 1$ weight vector such that $\sum_{l=1}^{d_x} \pi_l = 1$. Then \mathbb{H}_{10} implies that

$$r(y; x, z) \cdot (\pi' r_0) = r_1(x) \cdot (\pi' r_2(y)).$$

In practice, we can simply choose $\pi = (1/d_x, \dots, 1/d_x)$. Let

$$r_2^\pi(y) \equiv \pi' r_2(y) \text{ and } r_0^\pi \equiv \pi' r_0.$$

The following corollary summarizes a testable implication of (2.1) under \mathbb{H}_{10} and A.1.

Corollary 2.2 *Suppose that \mathbb{H}_{10} and A.1 hold. If $r_0^\pi \neq 0$, then*

$$r(Y; X, Z) r_0^\pi = r_1(X) r_2^\pi(Y) \text{ a.s.} \quad (2.5)$$

Remark 2.3 This corollary remains valid if we drop the indicator $\mathbf{1}\{y \in \mathcal{Y}_0\}$ in the definition of r in (2.3). Equivalently, one can take $\mathcal{Y}_0 = \mathcal{Y}$ in the definition of r and still obtain the above result provided that r is well defined. We incorporate the indicator function in our theorem to permit the trimming of the data in the tails that facilitates the establishment of the asymptotic properties of our test. Specifically our asymptotic theory below requires consistent estimation of $r(y; x, z)$ uniformly in $(y; x, z) \in \mathcal{Y}_0 \times \mathcal{V}$. If $f(y | x, z)$ is too close to zero for some values of $(y; x, z) \in \mathcal{Y} \times \mathcal{V}$, then we cannot estimate $r(y; x, z)$ uniformly in $(y; x, z) \in \mathcal{Y} \times \mathcal{V}$ at a sufficiently fast rate. We therefore restrict our attention to a subset \mathcal{Y}_0 of \mathcal{Y} such that $f(y | x, z)$ is bounded away from zero on $\mathcal{Y}_0 \times \mathcal{V}$.

Based on Corollary 2.2, consider the following null hypothesis

$$\mathbb{H}_0 : \Pr[r(Y; X, Z) r_0^\pi - r_1(X) r_2^\pi(Y) = 0] = 1. \quad (2.6)$$

The alternative hypothesis \mathbb{H}_A is the negation of \mathbb{H}_0 , i.e.,

$$\mathbb{H}_A : \Pr[r(Y; X, Z) r_0^\pi - r_1(X) r_2^\pi(Y) = 0] < 1. \quad (2.7)$$

According to the characterization result in Theorem 2.1, rejection of (2.6) can only be due either to the violation of \mathbb{H}_{10} , the original null hypothesis of interest, or to the violation of conditional exogeneity in A.1. Maintaining the conditional exogeneity assumption, we may therefore use the null hypothesis \mathbb{H}_0 to test the original null of interest, \mathbb{H}_{10} . Alternatively, if we maintain the transformation model specification in \mathbb{H}_{10} , our test can be used to test the conditional exogeneity assumption A.1.

To test the null hypothesis \mathbb{H}_0 in (2.6), we follow the lead of Härdle and Mammen (1993) and consider the weighted L_2 distance between rr_0^π and $r_1r_2^\pi$:

$$\Gamma \equiv E \left[\|r(Y; X, Z) r_0^\pi - r_1(X) r_2^\pi(Y)\|^2 \cdot a(Y; X, Z) \right], \quad (2.8)$$

where $\|\cdot\|$ denotes the Euclidean norm, and $a(y; x, z)$ is a nonnegative weight function that has compact support $\mathcal{Y}_0 \times \mathcal{V}_0$, where $\mathcal{V}_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0 \subset \mathcal{V}$. Then $\Gamma = 0$ under \mathbb{H}_0 and generally deviates from zero under \mathbb{H}_A . In the next subsection we consider the sample version of Γ based on local polynomial estimates of r , r_0^π , r_1 , and r_2^π .

2.2 Estimation and test statistic

The derivations in the previous section allow the covariates Z to be continuous or discrete. To describe our estimators and associated test statistics, we first consider the (more difficult) case where Z is continuous. Remark 2.4 below then discusses the case where some or all of the elements of Z are discrete.

We employ local polynomial regression to estimate various unknown population objects. Let $v \equiv (x', z')' = (v_1, \dots, v_d)'$ 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 $\mathbf{j} \equiv (j_1, \dots, j_d)$ be a d -vector of non-negative integers. Following Masry (1996), adopt the notation

$$v^{\mathbf{j}} \equiv \prod_{i=1}^d v_i^{j_i}, \quad \mathbf{j}! \equiv \prod_{i=1}^d j_i!, \quad |\mathbf{j}| \equiv \sum_{i=1}^d j_i, \quad \sum_{0 \leq |\mathbf{j}| \leq p} \equiv \sum_{k=0}^p \sum_{j_1=0}^k \cdots \sum_{j_d=0}^k \mathbf{1}_{j_1 + \dots + j_d = k}.$$

From $v^{\mathbf{j}} \equiv \prod_{i=1}^d v_i^{j_i}$, the j_i 's represent powers applied to the elements of v when constructing polynomials.

Consider the p -th order local polynomial estimators $D_x \hat{F}_b(y|x, z)$ of $D_x F(y|x, z)$. The subscript $b = b_n$ is a bandwidth parameter. Let $V_i \equiv (X_i', Z_i')'$ so $V_i - v = ((X_i - x)', (Z_i - z)')'$. Given observations $\{(Y_i, V_i), i = 1, \dots, n\}$, we estimate $D_x F(y|v)$ by solving the weighted least squares problem

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{i=1}^n \left[\mathbf{1}\{Y_i \leq y\} - \sum_{0 \leq |\mathbf{j}| \leq p} \beta'_{\mathbf{j}} ((V_i - v)/b)^{\mathbf{j}} \right]^2 K_b(V_i - v). \quad (2.9)$$

Here $\boldsymbol{\beta}$ stacks the $\beta_{\mathbf{j}}$'s ($0 \leq |\mathbf{j}| \leq p$) in lexicographic order (with $\beta_{\mathbf{0}}$, indexed by $\mathbf{0} \equiv (0, \dots, 0)$, in the first position, the element with index $(0, 0, \dots, 1)$ next, etc.) and $K_b(\cdot) \equiv K(\cdot/b)/b^d$, where $K(\cdot)$ is a symmetric PDF on \mathbb{R}^d . Let $\hat{\boldsymbol{\beta}}(y|v)$ denote the solution to the above minimization problem.

Let $N_l \equiv (l + d - 1)! / (l!(d - 1)!)$ be the number of distinct d -tuples \mathbf{j} having $|\mathbf{j}| = l$. In the above estimation problem, this denotes the number of distinct l th order partial derivatives of $F(y|v)$ with respect to v . Let $N \equiv \sum_{l=0}^p N_l$. Let $\mu(\cdot)$ be a stacking function such that $\mu((V_i - v)/b)$ denotes an $N \times 1$ vector that stacks $((V_i - v)/b)^{\mathbf{j}}$, $0 \leq |\mathbf{j}| \leq p$, in lexicographic order (e.g., $\mu(v) = (1, v)'$ when $p = 1$). Let $\mu_b(v) \equiv \mu(v/b)$. Then

$$\hat{\boldsymbol{\beta}}(y|v) = [\mathbf{S}_b(v)]^{-1} n^{-1} \sum_{i=1}^n K_b(V_i - v) \mu_b(V_i - v) \mathbf{1}\{Y_i \leq y\}, \quad (2.10)$$

where $\mathbf{S}_b(v) \equiv n^{-1} \sum_{i=1}^n K_b(V_i - v) \mu_b(V_i - v) \mu_b(V_i - v)'$. The p -th order local polynomial estimator $D_x \hat{F}_b(y|x, z)$ of $D_x F(y|x, z)$ is given by

$$D_x \hat{F}_b(y|x, z) = e_1 \hat{\boldsymbol{\beta}}(y|x, z) / b \quad (2.11)$$

where $e_1 \equiv [0_{d_x \times 1}, I_{d_x}, 0_{d_x \times (N - d_x - 1)}]$ selects the estimator of the coefficient of $(X_i - x)/b$ in the above regression.

To estimate $f(y|v)$, the conditional PDF of Y_i given $V_i = v$, we again employ local polynomial regression. Like Fan et al. (1996), we estimate $f(y|v)$ as $\hat{f}_c(y|v)$, the minimizing constant in the weighted least squares problem

$$\min_{\gamma} n^{-1} \sum_{i=1}^n \left[L_c(Y_i - y) - \sum_{0 \leq |\mathbf{j}| \leq p} \gamma'_{\mathbf{j}} ((V_i - v)/c)^{\mathbf{j}} \right]^2 K_c(V_i - v),$$

where $\boldsymbol{\gamma}$ stacks the γ_j 's ($0 \leq |j| \leq p$) in lexicographic order and $L_c(\cdot) \equiv L(\cdot/c)/c$, with $L(\cdot)$ a symmetric kernel function defined on \mathbb{R} and $c \equiv c_n$ a bandwidth parameter. Here, we use the same bandwidth sequence for Y_i and V_i , although different choices of bandwidths are also possible. To reduce the bias of the estimator \hat{f}_c , we permit use of a higher-order kernel for L . It is straightforward to verify that

$$\hat{f}_c(y|v) = e_2' [\mathbf{S}_c(v)]^{-1} n^{-1} \sum_{i=1}^n K_c(V_i - v) \mu_c(V_i - v) L_c(Y_i - y), \quad (2.12)$$

where $e_2 \equiv (1, 0, \dots, 0)'$ is an $N \times 1$ vector.

Define

$$\begin{aligned} \hat{r}(y; x, z) &\equiv \frac{D_x \hat{F}_b(y|x, z)}{\hat{f}_c(y|x, z)} \mathbf{1}\{y \in \mathcal{Y}_0\}, & \hat{r}_0 &\equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \hat{r}(Y_i; X_j, Z_j), \\ \hat{r}_1(x) &\equiv \frac{1}{n} \sum_{i=1}^n \hat{r}(Y_i; x, Z_i), & \text{and } \hat{r}_2(y) &\equiv \frac{1}{n} \sum_{i=1}^n \hat{r}(y; X_i, Z_i). \end{aligned}$$

Let $\hat{r}_0^\pi \equiv \pi' \hat{r}_0$ and $\hat{r}_2^\pi(y) \equiv \pi' \hat{r}_2(y)$. Our proposed test statistic is

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n \|\hat{r}(Y_i; X_i, Z_i) \hat{r}_0^\pi - \hat{r}_1(X_i) \hat{r}_2^\pi(Y_i)\|^2 a(Y_i; X_i, Z_i), \quad (2.13)$$

which is a sample analogue of Γ in (2.8). We next study the asymptotic properties of $\hat{\Gamma}$ under \mathbb{H}_0 , \mathbb{H}_A , and a sequence of Pitman local alternatives.

Remark 2.4 The above estimators and associated tests are easily extended to allow some or all elements of Z to be discrete. To estimate $r(y; x, z)$ in this case, we can simply stratify the sample by each distinct discrete outcome. Specifically, suppose $Z = (Z_c, Z_d)$, where Z_c is continuous and Z_d discrete. Then estimate $r(y; x, z) = r(y; x, z_c, z_d)$ as above (replacing Z with Z_c everywhere), just using the data having $Z_{di} = z_d$, and repeat for each value z_d in the support of Z_d . The functions r_0 , r_1 and r_2 can be estimated exactly the same way, by averaging out (X_i, Y_i, Z_i) , (Y_i, Z_i) , and (X_i, Z_i) , respectively, and then our test statistic $\hat{\Gamma}$ is still given by (8.3). More sophisticated estimators (e.g., smoothing across the discrete Z_d cells as proposed in Li and Racine, 2003) could also be used to estimate r these functions. We omit the details for brevity.

3 Asymptotic Properties of the Test Statistic

3.1 Basic assumptions

To study asymptotic properties of $\hat{\Gamma}$, make the following assumptions.

Assumption C.1 Let $W_i \equiv (Y_i, X_i', Z_i)'$, $i = 1, 2, \dots, n$, be IID random variables on (Ω, \mathcal{F}, P) , with (Y_i, X_i, Z_i) distributed identically to (Y, X, Z) .

Assumption C.2 (i) The PDF $f(v)$ of V_i is continuous in $v \in \mathcal{V}$, and $f(y|v)$ is continuous in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$.

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

Assumption C.3 (i) $F(\cdot|v)$ is equicontinuous on \mathcal{Y}_0 : $\forall \epsilon > 0, \exists \delta > 0 : |y - \tilde{y}| < \delta \Rightarrow \sup_{y \in \mathcal{Y}_0} |F(y|v) - F(\tilde{y}|v)| < \epsilon$. For each $y \in \mathcal{Y}_0$, $F(y|\cdot)$ is Lipschitz continuous on \mathcal{V} and has all partial derivatives up to order $p+1$, $p \in \mathbb{N}$.

(ii) Let $D^{\mathbf{j}}F(y|v) \equiv \partial^{|\mathbf{j}|}F(y|v)/\partial^{j_1}v_1 \dots \partial^{j_d}v_d$. For each $y \in \mathcal{Y}_0$, $D^{\mathbf{j}}F(y|\cdot)$ with $|\mathbf{j}| = p+1$ is uniformly bounded and Lipschitz continuous on \mathcal{V} : for all $v, \tilde{v} \in \mathcal{V}$, $|D^{\mathbf{j}}F(y|v) - D^{\mathbf{j}}F(y|\tilde{v})| \leq C_3 \|v - \tilde{v}\|$ for some $C_3 \in (0, \infty)$ where $\|\cdot\|$ is the Euclidean norm.

(iii) For each $v \in \mathcal{V}$ and for all $y, \tilde{y} \in \mathcal{Y}_0$, $|D^{\mathbf{j}}F(y|v) - D^{\mathbf{j}}F(\tilde{y}|v)| \leq C_4 |y - \tilde{y}|$ for some $C_4 \in (0, \infty)$ where $|\mathbf{j}| = p+1$.

Assumption C.4 Let $r \geq 2$. The r th derivative $f^{(r)}(y|v)$ of $f(y|v)$ with respect to y and all the $(p+1)$ th partial derivatives of $f(y|v)$ with respect to v are uniformly continuous on $\mathcal{Y}_0 \times \mathcal{V}$.

Assumption C.5 (i) The kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(ii) $v \rightarrow \|v\|^{2p+1} K(v)$ is integrable on \mathbb{R}^d with respect to the Lebesgue measure.

(iii) Let $\mathbf{K}_{\mathbf{j}}(u) \equiv v^{\mathbf{j}}K(v)$ for all \mathbf{j} with $0 \leq |\mathbf{j}| \leq 2p+1$. For some finite constants $\sigma_K, \bar{\sigma}_1$, and $\bar{\sigma}_2$, either $K(\cdot)$ is compactly supported such that $K(v) = 0$ for $\|u\| > \sigma_K$, and $|\mathbf{K}_{\mathbf{j}}(v) - \mathbf{K}_{\mathbf{j}}(\tilde{v})| \leq \bar{\sigma}_2 \|v - \tilde{v}\|$ for any $v, \tilde{v} \in \mathbb{R}^d$ and for all \mathbf{j} with $0 \leq |\mathbf{j}| \leq 2p+1$; or $K(\cdot)$ is differentiable, $\|\partial \mathbf{K}_{\mathbf{j}}(v)/\partial v\| \leq \bar{\sigma}_1$, and for some $\iota_0 > 1$, $|\partial \mathbf{K}_{\mathbf{j}}(v)/\partial v| \leq \bar{\sigma}_1 \|v\|^{-\iota_0}$ for all $\|v\| > \sigma_K$ and for all \mathbf{j} with $0 \leq |\mathbf{j}| \leq 2p+1$.

Assumption C.6 The univariate kernel function L satisfies $\int L(y)^2 dy < \infty$ and is a symmetric r th order kernel, i.e., $\int L(y) dy = 1$, $\int y^s L(y) dy = 0$ for all $s = 1, \dots, r-1$, and $\int y^r L(y) dy < \infty$. The r th derivative of L exists and is continuous.

Assumption C.7 (i) $p > d/2$.

(ii) As $n \rightarrow \infty$, $(c^{p+1} + c^r)/b^{d/2} \rightarrow 0$, $(b^p + c^{p+1} + c^r)b^{d/2+2}/c^{d+1} \rightarrow 0$, $b^{d+4}/c^{d+1} \rightarrow 0$, $nb^{2(p+1)+d} \rightarrow 0$, and $nb^{d+2}(c^{2(p+1)} + c^{2r}) \rightarrow 0$.

(iii) As $n \rightarrow \infty$, $\min\{nb^{2d}, nb^{3d/2+1}/\ln n, nb^{d+2}, nb^{d_x+2}/\ln n, nb^{d+1}c^{(d+1)/2}/\ln n, nb^{d/2}c^{d+1}/\ln n, nb^{-(d/2+2)}c^{2(d+1)}/\ln n, nb^{-1}c^{3(d+1)/2}/\ln n, nb^{-(d+4)}c^{3(d+1)}\} \rightarrow \infty$.

We assume IID observations in Assumption C.1, which is standard in cross-section studies. Assumptions C.2-C.4 impose smoothness conditions on the conditional CDF $F(y|v)$ and PDF $f(y|v)$ that are used to ensure uniform consistency of our local polynomial estimators, based on results of Masry (1996) and Hansen (2008). Assumptions C.5 and C.6 impose conditions on the kernels K and L , which are standard in the literature for local polynomial regression or conditional density estimation. Assumption C.7 restricts the choice of bandwidth sequences b and c , the order p of local polynomial regressions, and the order r of the kernel L . This assumption allows c to differ from b , but in the case where $b = c$ Assumption C.7 simplifies to the following assumption.

Assumption C.7* (i) $p > d/2$ and $r > d/2$.

(ii) As $n \rightarrow \infty$, $nb^{2(p+1)+d} \rightarrow 0$ and $nb^{2r+d+2} \rightarrow 0$.

(iii) As $n \rightarrow \infty$, $\min\{nb^{2d}, nb^{3(d+1)/2}/\ln n, nb^{d+2}, nb^{d_x+2}/\ln n\} \rightarrow \infty$.

Note that we allow $d_z = 0$, otherwise the condition $nb^{d_x+2}/\ln n \rightarrow \infty$ as $n \rightarrow \infty$ becomes redundant.

3.2 Asymptotic null distribution

In this section, we study the asymptotic behavior of the test statistic in (2.13). To state the next result, let $w_i \equiv (y_i, v_i)'$ and introduce the following notation:

$$\begin{aligned}\zeta_{1k}(y; v) &\equiv b^{-1} f(y|v)^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mu_b(V_k - v) K_b(V_k - v) \bar{\mathbf{I}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ \zeta_{2k}(y; v) &\equiv f(y|v)^{-2} D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mu_c(V_k - v) K_c(V_k - v) \bar{\mathbf{L}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ \zeta_k(y; v) &\equiv \zeta_{1k}(y; v) - \zeta_{2k}(y; v), \quad \zeta(W_i, W_j, W_k) \equiv \zeta_j(Y_i; V_i)' \zeta_k(Y_i; V_i) a_i, \\ \varphi(w_i, w_j) &\equiv E[\zeta(W_1, w_i, w_j)],\end{aligned}$$

where $\bar{\mathbf{S}}_b(v) \equiv E[\mathbf{S}_b(v)]$, $\bar{\mathbf{I}}_y(W_k) \equiv \mathbf{1}\{Y_k \leq y\} - F(y|V_k)$, $\bar{\mathbf{L}}_y(W_k) \equiv L_c(Y_k - y) - \alpha(y|V_k)$, $\alpha(y|v) \equiv E[L_c(Y_k - y)|V_k = v]$ and $a_i \equiv a(Y_i; X_i, Z_i)$. Define an asymptotic bias term

$$\begin{aligned}\mathbb{B}_n &\equiv n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \varphi(W_i, W_i) + n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^n \left\| \sum_{j=1}^n \sum_{k=1}^n \zeta_k(Y_j; X_i, Z_j) r_2^\pi(Y_i) \right\|^2 a_i \\ &\quad - 2n^{-3} b^{\frac{d}{2}+2} r_0^\pi \sum_{i=1}^n \sum_{l=1}^n \zeta_l(Y_i; V_i)' \sum_{j=1}^n \sum_{k=1}^n \zeta_k(Y_j; X_i, Z_j) r_2^\pi(Y_i) a_i \\ &\equiv \mathbb{B}_{1n} + \mathbb{B}_{2n} - 2\mathbb{B}_{3n}, \text{ say.}\end{aligned}$$

We establish the asymptotic null distribution of the $\hat{\Gamma}$ test statistic as follows:

Theorem 3.1 *Suppose Assumptions A.1 and C.1-C.7 hold. Then*

$$nb^{\frac{d}{2}+2} \hat{\Gamma} - \mathbb{B}_n \xrightarrow{d} N(0, \sigma_0^2),$$

where $\sigma_0^2 \equiv \lim_{n \rightarrow \infty} \sigma_n^2$ and $\sigma_n^2 = 2b^{d+4} (r_0^\pi)^4 E[\varphi(W_1, W_2)^2]$.

Remark 3.1. The asymptotic bias \mathbb{B}_n of $nb^{\frac{d}{2}+2} \hat{\Gamma}$ contains three terms \mathbb{B}_{1n} , \mathbb{B}_{2n} , and $-2\mathbb{B}_{3n}$. The first two terms reflect the contributions of $\hat{r}(Y_i; V_i) \hat{r}_0$ and $\hat{r}_1(X_i) \hat{r}_2^\pi(Y_i)$ respectively, and the last term reflects the interaction between these latter two terms. We show that $\mathbb{B}_{1n} = O_P(b^{\frac{d}{2}+2}(b^{-d-2} + c^{-d-1}))$ in Lemma 8.4, $\mathbb{B}_{2n} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$ in Lemma 8.5(b), and $\mathbb{B}_{3n} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$ in Lemma 8.6(b). Clearly, \mathbb{B}_{1n} never vanishes asymptotically whereas \mathbb{B}_{2n} and \mathbb{B}_{3n} are asymptotically negligible under certain conditions, say when $b = c$ and $d_z > d_x$. The asymptotic variance σ_n^2 of $nb^{\frac{d}{2}+2} \hat{\Gamma}$ only reflects the contribution of $\hat{r}(Y_i; V_i) \hat{r}_0$, due to the faster convergence rate of $\hat{r}_1(X_i) \hat{r}_2^\pi(Y_i)$ to $r_1(X_i) r_2^\pi(Y_i)$ than that of $\hat{r}(Y_i; V_i) \hat{r}_0$ to $r(Y_i; V_i) r_0$.

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

$$\begin{aligned}\hat{\zeta}_{1k}(y; v) &\equiv b^{-1} \hat{f}_c(y|v)^{-1} e_1 \mathbf{S}_b(v)^{-1} \mu_b(V_k - v) K_b(V_k - v) \hat{\mathbf{I}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ \hat{\zeta}_{2k}(y; v) &\equiv \hat{f}_c(y|v)^{-2} D_x \hat{F}_b(y|v) e_2' \mathbf{S}_c(v)^{-1} \mu_c(V_k - v) K_c(V_k - v) \hat{\mathbf{L}}_y(W_k) \mathbf{1}\{y \in \mathcal{Y}_0\}, \\ \hat{\zeta}_k(y; v) &\equiv \hat{\zeta}_{1k}(y; v) - \hat{\zeta}_{2k}(y; v), \quad \hat{\varphi}(W_j, W_k) \equiv n^{-1} \sum_{i=1}^n \hat{\zeta}_j(Y_i; V_i)' \hat{\zeta}_k(Y_i; V_i) a_i,\end{aligned}$$

where $\hat{\mathbf{I}}_y(W_k) \equiv \mathbf{1}\{Y_k \leq y\} - \hat{F}_b(y|V_k)$, $\hat{\mathbf{L}}_y(W_k) \equiv L_c(Y_k - y) - n^{-1} \sum_{i=1}^n \hat{f}_c(y|V_k)$, and $\hat{F}_b(y|V_k)$ is the p th order local polynomial estimator of $F(y|V_k)$ by using the kernel K and bandwidth b . We propose

estimating the asymptotic bias \mathbb{B}_n and variance σ_n^2 respectively by

$$\begin{aligned}\hat{\mathbb{B}}_n &\equiv n^{-1}b^{\frac{d}{2}+2}(\hat{r}_0^\pi)^2 \sum_{i=1}^n \hat{\varphi}(W_i, W_i) + n^{-4}b^{\frac{d}{2}+2} \sum_{i=1}^n \left\| \sum_{j=1}^n \sum_{k=1}^n \hat{\zeta}_k(Y_j; X_i, Z_j) \hat{r}_2^\pi(Y_i) \right\|^2 a_i \\ &\quad - 2n^{-3}b^{\frac{d}{2}+2} \hat{r}_0^\pi \sum_{i=1}^n \sum_{l=1}^n \hat{\zeta}_l(Y_i; V_i)' \sum_{j=1}^n \sum_{k=1}^n \hat{\zeta}_k(Y_j; X_i, Z_j) \hat{r}_2^\pi(Y_i) a_i \\ \hat{\sigma}_n^2 &\equiv 2n^{-2}b^{d+4}(\hat{r}_0^\pi)^4 \sum_{i=1}^n \sum_{j=1}^n \hat{\varphi}(W_i, W_j)^2.\end{aligned}$$

It is straightforward to show $\hat{\mathbb{B}}_n - \mathbb{B}_n = o_P(1)$ and $\hat{\sigma}_n^2 - \sigma_n^2 = o_P(1)$. We can now compare

$$T_n \equiv \left(nb^{\frac{d}{2}+2} \hat{\Gamma} - \hat{\mathbb{B}}_n \right) / \sqrt{\hat{\sigma}_n^2} \quad (3.1)$$

to the critical value z_α defined as the upper α percentile from the $N(0, 1)$ distribution (since the test is one-sided) and reject the null when $T_n > z_\alpha$.

3.3 Consistency and asymptotic local power

The following theorem shows that the test T_n is consistent for the class of global alternatives

$$\mathbb{H}_A : \mu_A \equiv E \left\{ [r(Y; X, Z) r_0^\pi - r_1(X) r_2^\pi(Z)]^2 a(Y; X, Z) \right\} > 0.$$

Theorem 3.2 *Suppose Assumptions C.1-C.7 hold. Then under \mathbb{H}_A , $P(T_n > e_n) \rightarrow 1$ for any nonstochastic sequence $e_n = o(nb^{\frac{d}{2}+2})$.*

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

$$\mathbb{H}_A(\gamma_n) : r(y; x, z) r_0^\pi - r_1(x) r_2^\pi(y) = \gamma_n \delta_n(y; x, z), \quad (3.2)$$

where $\gamma_n \rightarrow 0$ as $n \rightarrow \infty$, and δ_n is a non-constant measurable function with $\mu_0 \equiv \lim_{n \rightarrow \infty} E[\delta_n(Y_1; X_1, Z_1)^2 a(Y_1; X_1, Z_1)] < \infty$.

Theorem 3.3 *Suppose Assumptions C.1-C.7 hold. Then under $\mathbb{H}_A(\gamma_n)$ with $\gamma_n = n^{-1/2}b^{-d/4-1}$, $T_n \xrightarrow{d} N(\mu_0/\sigma_0, 1)$.*

Theorem 3.3 implies that the T_n test has non-trivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}b^{-d/4-1}$, provided $0 < \mu_0 < \infty$. The asymptotic local power function of the test is given by $1 - \Phi(z_\alpha - \mu_0/\sigma_0)$, where Φ is the standard normal CDF.

3.4 Simulating the null distribution

As an alternative to estimating the bias and variance of the test's asymptotically normal distribution, in this subsection and Appendix II, we discuss simulation methods to obtain p -values. These methods may perform better than the normal critical-value-based tests in finite samples, or they may be more convenient to implement. Below we describe a simple subsampling procedure. In Appendix II, we propose two possible

bootstrap procedures and discuss potential advantages and disadvantages associated with each bootstrap method in the context of our tests.

Let $m = m_n$ be a sequence of positive integers such that $m \rightarrow \infty$ and $m/n \rightarrow 0$ as $n \rightarrow \infty$. Let B be a large integer. The subsampling procedure goes as follows:³

1. Randomly draw B subsamples $\left\{ \left(X_i^{*(k)}, Y_i^{*(k)}, Z_i^{*(k)} \right), i = 1, \dots, m \right\}_{k=1}^B$ of size m from the original sample $\{(X_i, Y_i, Z_i)\}_{i=1}^n$.
2. For $k = 1, \dots, B$, compute T_n using the subsample $\left\{ \left(X_i^{*(k)}, Y_i^{*(k)}, Z_i^{*(k)} \right) \right\}_{i=1}^m$ and denote this as $\hat{T}_{n,m}^{*(k)}$.
3. Calculate the subsampling p -value as

$$p = B^{-1} \sum_{k=1}^B \mathbf{1} \left\{ T_n < \hat{T}_{n,m}^{*(k)} \right\}.$$

The asymptotic validity of the above subsampling method can be established as in Politis et al. (1999). Under the null hypothesis both T_n and $\hat{T}_{n,m}^{*(k)}$ are asymptotically distributed as $N(0, 1)$ and thus the test based on this subsampling based p -value has the correct asymptotic size, and under the fixed alternative T_n diverges to infinity at a speed faster than $\hat{T}_{n,m}^{*(k)}$, giving the test its power.

4 Monte Carlo Simulations

In this section, we use simulations to examine the finite sample performance of our test. We consider eight data generating processes (DGPs), the first four of which are as follows:

DGP 1: $Y = X + U$;

DGP 2: $Y = X + U + X\sqrt{1 + U^2}$;

DGP 3: $Y = \Phi(X + U)$, where Φ is the standard normal CDF;

DGP 4: $Y = \Phi(X + U + X\sqrt{1 + U^2})$;

where $X \sim \text{Uniform}(-1, 1)$, $U \sim \text{Uniform}(-1, 1)$, and X and U are independent.

DGPs 5-8 are identical to DGPs 1-4, respectively, except that X and U are no longer independent: $X = 0.5Z + 0.5\varepsilon_1$ and $U = 0.5Z + 0.5\varepsilon_2$, where $\varepsilon_1 \sim \text{Uniform}(-1, 1)$, $\varepsilon_2 \sim \text{Uniform}(-1, 1)$, Z follows a standard normal distribution truncated by -2 and 2 in the tails, and $\varepsilon_1, \varepsilon_2$, and Z are mutually independent. By construction, $X \perp U \mid Z$, DGPs 1, 3, 5 and 7 satisfy the null, and DGPs 2, 4, 6, and 8 obey the alternative.

We use second order (quadratic) local polynomial estimators, i.e., $p = 2$, with a Gaussian PDF for the kernel function. For the bandwidth sequence b and c , we use the rule $\kappa \cdot \text{std}(V) \cdot n^{-\frac{1}{2(p+1)+1}}$ and $\kappa \cdot \text{std}(Y) \cdot n^{-\frac{1}{2(p+1)+1}}$ associated with V and Y , respectively, where κ is a constant and $\text{std}(V)$ and $\text{std}(Y)$ are sample standard deviations of V and Y , respectively. In general, the optimal κ depends on the underlying specific DGPs. For simplicity we let $\kappa = 1$ for DGPs 1-4 and $\kappa = 2$ for DGPs 5-8. For DGPs 1-4, we specify the

³Alternatively, one can refer this as to the m -out-of- n bootstrap procedure.

weight function $a = 1$, corresponding to no trimming, whereas for DGPs 5-8, a trims out 2.5% data on each tail of each dimension of (Y, X, Z) , so

$$a(Y, X, Z) = \mathbf{1}[y_{0.025} \leq Y \leq y_{0.975}] \cdot \mathbf{1}[x_{0.025} \leq X \leq x_{0.975}] \cdot \mathbf{1}[z_{0.025} \leq Z \leq z_{0.975}],$$

where $y_{0.025}$ and $y_{0.975}$ are the 0.025 and 0.975 quantiles of Y respectively, and similarly for $x_{0.025}$, $x_{0.975}$, $z_{0.025}$, and $z_{0.975}$.

We first consider the subsampling test with the sample sizes $n = 200$ and 300 . We try three different subsample sizes $m = \lfloor n^{0.8} \rfloor$, $\lfloor n^{0.85} \rfloor$, and $\lfloor n^{0.9} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of \cdot . The number of subsamples is $B = 200$ and the number of replications is 200 .

Table 1 presents the rejection frequencies for DGPs 1, 3, 5, and 7 under the null. We consider three conventional nominal levels: 0.01, 0.05 and 0.10. When the sample size is 200, the subsampling tests are undersized. However, when the sample size increases to 300, the performance improves and the rejection frequencies are closer to their nominal levels. This suggests that a moderate to large sample is required for the test to have good level behavior. This is not surprising, as the estimation of derivatives is much harder and has a slower convergence rate than the estimation of the conditional expectation itself. The level behavior is similar for different subsample sizes m .

Table 1: Empirical rejection frequency: levels

DGP	n	Subsample size								
		$\lfloor n^{0.80} \rfloor$			$\lfloor n^{0.85} \rfloor$			$\lfloor n^{0.90} \rfloor$		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
1	200	0	0.015	0.035	0	0.010	0.050	0	0.010	0.035
	300	0.010	0.035	0.130	0.005	0.030	0.125	0	0.020	0.085
3	200	0	0.010	0.035	0	0.010	0.040	0	0.005	0.020
	300	0.010	0.035	0.085	0	0.030	0.080	0	0.025	0.065
5	200	0	0.015	0.045	0	0	0.020	0	0.005	0.015
	300	0	0.025	0.100	0	0.015	0.060	0	0	0.035
7	200	0	0.015	0.050	0	0.015	0.030	0	0.005	0.020
	300	0	0.035	0.100	0.005	0.020	0.065	0	0.005	0.035

Table 2: Empirical rejection frequency: powers

DGP	n	Subsample size								
		$\lfloor n^{0.80} \rfloor$			$\lfloor n^{0.85} \rfloor$			$\lfloor n^{0.90} \rfloor$		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
2	200	0.540	0.930	0.985	0.470	0.875	0.965	0.235	0.760	0.905
	300	0.940	1.000	1.000	0.895	0.995	1.000	0.725	0.980	0.995
4	200	0.610	0.960	0.990	0.350	0.890	0.975	0.110	0.580	0.930
	300	0.790	0.990	1.000	0.505	0.975	1.000	0.145	0.745	0.970
6	200	0.340	0.810	0.950	0.180	0.665	0.875	0.075	0.465	0.770
	300	0.790	0.980	1.000	0.685	0.950	0.995	0.365	0.825	0.965
8	200	0.075	0.250	0.435	0.030	0.160	0.310	0	0.060	0.225
	300	0.135	0.460	0.705	0.100	0.395	0.655	0.025	0.210	0.440

Table 2 reports the rejection frequencies for DGPs 2, 4, 6 and 8 under the alternative. The power of the tests behaves well. When $n = 200$, even at the 0.01 nominal level, the test has substantial power; for example, the rejection frequency is 0.540 for DGP 1 when sample size $n = 200$ and subsample size $m = n^{0.80}$. The power increases rapidly as the sample size increases. For example, the rejection frequency becomes 0.940 for DGP 1 when $n = 300$ and $m = n^{0.80}$. The power of the tests increases when the subsample size m decreases. This is likely because, under the alternative, the test statistics diverge, so the difference between the original test statistics and subsampled test statistics is large when the difference between the original sample size and subsample size is large.⁴

5 Empirical Applications

In this section, we consider testing whether duration data obey the class of nonlinear generalized accelerated failure-time (GAFT) models. We then apply our test empirically on two different data sets. The first application is duration of strikes among manufacturing workers in the US, and the second is duration of first marriages among divorced couples.

For these applications, Y is the duration of a certain state (a nonnegative random variable) such as duration of a strike. Our test is directly applicable to nonlinear GAFT models, since such models can be written in the form $Y = G[H_1(X) + U]$, where X is a vector of covariates, and U an unobservable random variable (see, eq. (2.5) in Ridder, 1990). The GAFT models include accelerated failure-time (AFT) models, proportional hazards (PH) models, and mixed proportional hazard (MPH) models.

⁴We also consider the weighted bootstrap discussed in Appendix II. We find that for some DGPs, especially when the dimension of (X, Z) is large, this bootstrap over-rejects the null. However, we find that the weighted bootstrap procedure has better power than subsampling uniformly for all DGPs. Since the weighted bootstrap test can be significantly over-sized, we do not recommend its use. The weighted bootstrap simulation results are available upon request.

MPH models are a particularly popular class of GAFT models (for a detailed review, see Van den Berg (2001)). Below we provide a direct link between our null hypothesis and MPH models. Let $h(Y, X, \xi)$ denote the hazard function for Y . An MPH model of survival time Y is one where

$$h(Y, X, \xi) = \lambda(Y) \cdot \theta(X) \cdot \xi \tag{5.1}$$

holds for some baseline hazard function $\lambda(Y)$ and some nonnegative function of covariates $\theta(X)$. The MPH model is widely applied in empirical research. For example, when $\xi = 1$, this is the standard proportion hazard (PH) model developed by Cox (1972). A particularly popular parametric specification of the MPH model due to Lancaster (1979) assumes that $\lambda(Y) = \alpha Y^{\alpha-1}$, $\theta(X) = \exp(X'\beta)$ and ξ is a gamma distributed random variable. The following Proposition provides a general characterization of MPH models.

Proposition 5.1 *Suppose that the hazard function of the survival time Y is $h(Y, X, \xi)$, where $Y \in \mathbb{R}_+$, $X \in \mathbb{R}^{d_x}$, $\xi \in \mathbb{R}_+$ and $\xi \neq 0$ with probability 1. Let $\lambda : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\theta : \mathbb{R}^{d_x} \rightarrow \mathbb{R}_+$ be two measurable functions such that $\lambda(Y) = 0$ with probability 0 and $\theta(X) = 0$ with probability 0. Then $h(Y, X, \xi)$ is a MPH model:*

$$h(Y, X, \xi) = \lambda(Y) \cdot \theta(X) \cdot \xi,$$

if and only if

$$Y = G[H_1(X) + U],$$

where $G : \mathbb{R} \rightarrow \mathbb{R}_+$ is a strictly increasing function that is differentiable a.e. on its support, $H_1 : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$, and $U = \ln \left[\frac{-\ln(1-\varepsilon)}{\xi} \right]$, where ε is a uniform random variable on $[0, 1]$ and $\varepsilon \perp (X, \xi)$.

Proposition 5.1 shows that the MPH model has two important implications: (i) it equals a transformation model of the type given by our null, and (ii) U allows a distribution determined by $\ln(-\ln(1-\varepsilon)/\xi)$. In principle, both restrictions might be testable, though we focus on implication (i), corresponding to our null hypothesis.⁵ If our null is rejected, then the specification of MPH models is rejected, so our test can be used as a falsification test for MPH models.

5.1 Duration of strikes

In this subsection, we test the specification of GAFT models using data on the duration of strikes. Here Y is the duration of strikes in U.S. manufacturing firms, defined as the number of days since the start of a strike. Our X is a scalar variable indicator of the business cycle position of the economy, measured by the deviation of output from its trend. Positive values of X mean that the economy is above its growth trend. We assume that A.1 holds with $X \perp U$, i.e., Z is empty.

Our dataset was used in Kennan (1985) and is employed in several econometrics textbooks including as Cameron and Trivedi (2005) and Greene (2011). The sample size is 566. Table 3 presents data summary statistics.

⁵More broadly, this proposition shows that nonparametrically the only difference between GAFT and MPH models is some regularity conditions, since if one is given a GAFT model which by Ridder (1990) satisfies $Y = G[H_1(X) + U]$, then given the regularity assumed in Proposition 5.1, one can construct an equivalent MPH model by letting $\xi = [-\ln(1-\varepsilon)]e^{-U}$ where ε is uniform.

Table 3: Summary statistics for the strike data (sample size $n = 566$)

Variable	Name	Mean	Median	Standard Deviation	Min	Max
Y	Duration of strikes (days)	43.62	28.00	44.67	1	235
X	Business cycle position	0.006	0.008	0.050	-0.140	0.086

We apply our subsampling based test. Results based on 1000 subsamples are reported in Table 4. The p -values are high for all subsample sizes under investigation. This suggests that our test supports the specification of GAFT models.

Table 4: Test results for the strike data (sample size $n = 566$)

Subsample size	$\lfloor n^{0.80} \rfloor = 159$	$\lfloor n^{0.85} \rfloor = 219$	$\lfloor n^{0.90} \rfloor = 300$
p -values	0.651	0.578	0.572

5.2 Duration of marriage

In this subsection, we apply our test to study the duration of the first marriage of divorced couples. Let Y be the duration of the first marriage of a couple and X be the age difference of the couple. The dataset is the U.S. survey data taken from Lillard and Panis (2003). We choose a relatively homogeneous subpopulation where the divorced couples are white and have more than 10 years of education. The sample size is 542. Data summary statistics are provided in Table 5.

Table 5: Summary statistics for the marriage data (sample size $n = 542$)

Variable	Name	Mean	Median	Standard Deviation	Min	Max
Y	Duration of marriage	10.37	8.04	7.89	0.10	50.38
X	Age difference	2.13	2.35	4.21	-31.25	33.00

We implement our test with this dataset and find that the GAFT model (and hence also the MPH model) is soundly rejected. The p -values are all smaller than 0.01 for all subsample sizes. These test results based on 1000 subsamples are shown in Table 6.

Table 6: Test results for the marriage data (sample size $n = 542$)

Subsample size	$\lfloor n^{0.80} \rfloor = 154$	$\lfloor n^{0.85} \rfloor = 211$	$\lfloor n^{0.90} \rfloor = 289$
p -values	0.006	0.005	0.005

This rejection of the null hypothesis could be due either to inadequacy of the GAFT specification, or the assumption that $X \perp U$ in our homogeneous subpopulation could be violated. This subpopulation controls for education and race, but it is possible that GAFT would not be rejected if we observed and conditioned on other covariates Z such as the religious affiliation, the number of children, and the income of the couple, among others.

Given our results, one could either propose a more general duration model than the GAFT class, such as Chesher's (2002) nonseparable semiparametric model $h(Y, X, \xi) = \psi(Y, \theta(X)) / \xi$, where ψ and θ are unknown functions. Alternatively, one could seek out a data set with more covariates to condition on (and more observations to deal with the curse of dimensionality that would arise with more covariates), and test if GAFT holds in this larger data set.

6 Extensions

Our methodology can be extended to test other related hypotheses for specifications in nonseparable models. For example, suppose that X is multi-dimensional such that $X \equiv (X_1, X_2)$. Then our results can be used to test the hypotheses:

\mathbb{H}_{20} : There exist two measurable functions R_2 and H_3 such that

$$Y = R_2 [H_3 (X_1, X_2), U] \text{ a.s.}$$

\mathbb{H}_{2A} : \mathbb{H}_{20} is false;

and

\mathbb{H}_{30} : There exist three measurable functions R_3 , H_4 and H_5 such that

$$Y = R_3 [H_4 (X_1) + H_5 (X_2), U] \text{ a.s.}$$

\mathbb{H}_{3A} : \mathbb{H}_{30} is false.

Given the key conditional exogeneity assumption A.1, a testable implication of \mathbb{H}_{20} is

$$\frac{\partial F_{Y|X_1, X_2, Z}(y | x_1, x_2, z) / \partial x_1}{\partial F_{Y|X_1, X_2, Z}(y | x_1, x_2, z) / \partial x_2} = r_3(x_1, x_2), \quad (6.1)$$

where $F_{Y|X_1, X_2, Z}(y | x_1, x_2, z)$ is the conditional CDF of Y given (X_1, X_2, Z) and r_3 some unknown measurable function. Similarly, \mathbb{H}_{30} implies that

$$\frac{\partial F_{Y|X_1, X_2, Z}(y | x_1, x_2, z) / \partial x_1}{\partial F_{Y|X_1, X_2, Z}(y | x_1, x_2, z) / \partial x_2} = r_4(x_1) \cdot r_5(x_2) \quad (6.2)$$

for some unknown measurable functions r_4 and r_5 .

Our test can also be extended to test for semiparametric specifications. For example, one may be interested in testing

\mathbb{H}_{40} : There exist $\beta \in \mathbb{R}^{d_x}$ and two measurable functions R_4 and H_2 such that

$$Y = R_4 [X' \beta + H_2 (U)] \text{ a.s.}$$

\mathbb{H}_{4A} : \mathbb{H}_{40} is false;

Then \mathbb{H}_{40} implies that

$$r(y; x, z) \equiv \frac{D_x F(y | x, z)}{f(y | x, z)} = r_6(y) \quad (6.3)$$

for some unknown measurable function r_6 .

To test equations (6.1), (6.2), and (6.3), one can readily construct test statistics similar to ours, using marginal integration as proposed in testing \mathbb{H}_0 .

7 Concluding Remarks

In this paper, we proposed a specification test for a transformation model containing a vector of covariates and a vector of unobservable errors. This test is related to tests for separability and monotonicity in

nonseparable structural equations. We derive the testable implication of the transformation model that the ratio of the derivatives of a conditional CDF takes a product form. Our test statistics are based on the L_2 distance between restricted and unrestricted estimators of this ratio of derivatives. We show that the test statistics are asymptotically normal and consistent against the alternative of this testable implication. We provide limit normal distribution theory as well as bootstrap and subsampling methods for obtaining p -values under the null. Our simulations suggest that the test statistics perform well in moderate size samples. We apply our statistics to test the specification of generalized accelerated failure-time models for data on the durations of strikes among manufacturers in the US and of first marriages of divorced couples. Both data sets are similar size and have the same dimension. We fail to reject GAFT for the strikes data, while strongly rejecting it in the marriage data.

8 Appendix I: Mathematical Proofs

Proof of Theorem 2.1

We first prove (a). Let $\tilde{U} = H_2(U)$. Then

$$\begin{aligned} F(y | x, z) &= \Pr[Y \leq y | X = x, Z = z] \\ &= \Pr\left[G\left[H_1(X) + \tilde{U}\right] \leq y | X = x, Z = z\right] \\ &= \Pr\left[\tilde{U} \leq G^{-1}(y) - H_1(x) | Z = z\right] \\ &= F_{\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right], \end{aligned}$$

where $F_{\tilde{U}|Z}(\cdot, z)$ denotes the conditional CDF of \tilde{U} given $Z = z$. Let $F_{1,\tilde{U}|Z}$ be the derivative of $F_{\tilde{U}|Z}$ with respect to its first argument. Then,

$$\frac{\partial F(y | x, z) / \partial x}{\partial F(y | x, z) / \partial y} = \frac{F_{1,\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right] \cdot [-\partial H_1(x) / \partial x]}{F_{1,\tilde{U}|Z}\left[G^{-1}(y) - H_1(x), z\right] \cdot [\partial G^{-1}(y) / \partial y]} = \frac{-\partial H_1(x) / \partial x}{\partial G^{-1}(y) / \partial y}.$$

So the functions s_1 and s_2 exist and are given by

$$s_1(x) = -C \partial H_1(x) / \partial x \quad \text{and} \quad s_2(y) = \frac{1}{C \partial G^{-1}(y) / \partial y},$$

where $C \neq 0$ is an arbitrary constant. Clearly, $s_2 : \mathbb{R} \rightarrow \mathbb{R}_+$ if $C > 0$ and $s_2 : \mathbb{R} \rightarrow \mathbb{R}_-$ if $C < 0$. The measurable functions S_1 and S_2 are given by CH_1 and CG^{-1} , respectively.

We now prove (b). Without loss of generality, assume that $s_2 : \mathbb{R} \rightarrow \mathbb{R}_+$. We can always find two scalar functions S_1 and S_2 such that $\partial S_1(x) / \partial x = -s_1(x)$ and $\partial S_2(y) / \partial y = 1/s_2(y)$, where $S_2(\cdot)$ is strictly increasing. Combining this with the definition of $r(y; x, z)$ gives

$$\frac{D_x F(y | x, z)}{D_y F(y | x, z)} = s_1(x) s_2(y) = \frac{-\partial S_1(x) / \partial x}{\partial S_2(y) / \partial y} \quad \text{for all } (x, y, z) \in \mathcal{W}. \quad (8.1)$$

Let $\tilde{U} \equiv S_2(Y) - S_1(X)$ and $\tilde{u} \equiv S_2(y) - S_1(x)$. By the monotonicity of S_2 , we have $Y = S_2^{-1}[S_1(X) + \tilde{U}]$ and $y = S_2^{-1}[S_1(x) + \tilde{u}]$. It follows that

$$\begin{aligned} F_{\tilde{U}|X,Z}(\tilde{u}, x, z) &\equiv P\left(\tilde{U} \leq \tilde{u} | X = x, Z = z\right) = P\left(S_2(Y) - S_1(X) \leq \tilde{u} | X = x, Z = z\right) \\ &= P\left(Y \leq S_2^{-1}\left(S_1(x) + \tilde{u}\right) | X = x, Z = z\right) \\ &= P\left(Y \leq y | X = x, Z = z\right) = F(y | x, z). \end{aligned}$$

Then

$$\begin{aligned} \frac{D_x F(y|x, z)}{D_y F(y|x, z)} &= \frac{D_x F_{\tilde{U}|X, Z}(\tilde{u}, x, z)}{D_y F_{\tilde{U}|X, Z}(\tilde{u}, x, z)} = \frac{\partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial \tilde{u} \cdot s_1(x) + \partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial x}{\partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial \tilde{u} \cdot (1/s_2(y))} \\ &= s_1(x) s_2(y) + \frac{\partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial x \cdot (1/s_2(y))}{\partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial \tilde{u}} \text{ for all } (x, y, z) \in \mathcal{W}. \end{aligned} \quad (8.2)$$

Comparing (8.1) with (8.2) yields $\partial F_{\tilde{U}|X, Z}(\tilde{u}, x, z) / \partial x = 0$ for all $(\tilde{u}, x, z) \in \mathcal{U} \times \mathcal{V}$ where \mathcal{U} denotes the support of \tilde{U} . Therefore, $\tilde{U} \perp X|Z$. So far, we have shown that

$$Y = S_2^{-1}[S_1(X) + \tilde{U}]$$

where S_2^{-1} is strictly monotonic and $X \perp \tilde{U}|Z$. The conclusion in part (b) follows by setting $G = S_2^{-1}$ and $H_1 = S_1$. ■

Proof of Corollary 2.2

Under \mathbb{H}_{10} and A.1, (2.1) in Theorem 2.1(a) holds, implying that

$$\begin{aligned} r_0^\pi &\equiv \pi' E_Y E_{XZ} [r(Y; X, Z)] = \pi' E [s_1(X)] E [s_2(Y)], \\ r_1(x) &\equiv E [r(Y; x, Z)] = s_1(x) E [s_2(Y)], \\ r_2^\pi(y) &\equiv \pi' E [r(y; X, Z)] = s_2(y) \pi' E [s_1(X)]. \end{aligned}$$

It follows that

$$\begin{aligned} r(X, Y, Z) r_0^\pi - r_1(X) r_2^\pi(Y) &= [s_1(X) s_2(Y)] \{\pi' E [s_1(X)] E [s_2(Y)]\} \\ &\quad - \{s_1(X) E [s_2(Y)]\} \{s_2(Y) \pi' E [s_1(X)]\} \\ &= 0. \quad \blacksquare \end{aligned}$$

To prove Theorem 3.1, we first establish some technical lemmas. Recall that $V_i \equiv (X'_i, Z'_i)'$, $v \equiv (x', z')'$, $K_b(v) \equiv b^{-d} K(v/b)$, and $\mu_b(v) \equiv \mu(v/b)$. Let $W_i \equiv (Y_i, V_i)'$ and $w \equiv (y, v)'$. Define

$$\begin{aligned} \mathbf{B}_b(y; v) &\equiv \frac{1}{n} \sum_{i=1}^n K_b(V_i - v) \mu_b(V_i - v) \Delta_{i,y}(v), \\ \mathbf{V}_b(y; v) &\equiv \frac{1}{n} \sum_{i=1}^n K_b(V_i - v) \mu_b(V_i - v) \bar{\mathbf{1}}_y(W_i), \end{aligned}$$

where $\Delta_{i,y}(v) \equiv F(y|V_i) - F(y|v) - \sum_{1 \leq |j| \leq p} \frac{1}{j!} D^j F(y|v) (V_i - v)^j$, and $\bar{\mathbf{1}}_y(W_i) \equiv \mathbf{1}\{Y_i \leq y\} - F(y|V_i)$. Let $\bar{\mathbf{S}}_b(v) \equiv E[\mathbf{S}_b(v)]$ and $\bar{\mathbf{B}}_b(y; v) \equiv E[\mathbf{B}_b(y; v)]$, where $\mathbf{S}_b(v)$ is defined after (2.10). The next lemma establishes uniform consistency of $D_x \hat{F}_b(y|v)$.

Lemma 8.1 *Suppose that Assumptions C.1-C.3, C.5, and C.7 hold. Then uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,*

$$(a) \quad D_x \hat{F}_b(y|v) - D_x F(y|v) = b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} [\mathbf{V}_b(y; v) + \bar{\mathbf{B}}_b(y; v)] + O_P(\nu_{1b}^2 b^{-1} + \nu_{1b} b^p),$$

$$(b) \quad D_x \hat{F}_b(y|v) - D_x F(y|v) = O_P(\nu_{1b} b^{-1} + b^p),$$

where $\nu_{1b} \equiv n^{-1/2} b^{-d/2} \sqrt{\ln n}$.

Proof. By Lemma 10.1 in HSW (2011), $\hat{\beta}(y|v) - \beta(y|v) = \bar{\mathbf{S}}_b(v)^{-1} [\mathbf{V}_b(y;v) + \bar{\mathbf{B}}_b(y;v)] + O_P(\nu_{1b}^2 + \nu_{1b}b^{p+1}) = O_P(\nu_{1b} + b^{p+1})$ uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$. The results follow from the fact that $D_x \hat{F}_b(y|v) - D_x F(y|v) = e_1[\hat{\beta}(y|x, z) - \beta(y|v)]/b$. ■

Define

$$\begin{aligned} \mathbf{V}_c^{(L)}(y;v) &\equiv \frac{1}{n} \sum_{i=1}^n K_c(V_i - v) \mu_c(V_i - v) \bar{\mathbf{L}}_y(W_i), \\ \mathbf{B}_c^{(L)}(y;v) &\equiv \frac{1}{n} \sum_{i=1}^n K_c(V_i - v) \mu_c(V_i - v) \left[\alpha(y|V_i) - f(y|v) - \sum_{1 \leq |\mathbf{j}| \leq p} \frac{1}{\mathbf{j}!} \alpha^{(\mathbf{j})}(y|v) (V_i - v)^{\mathbf{j}} \right], \end{aligned}$$

where $\bar{\mathbf{L}}_y(W_i) \equiv L_c(Y_i - y) - \alpha(y|V_i)$ and $\alpha(y|v) \equiv E[L_c(Y_i - y)|V_i = v]$. Let $\bar{\mathbf{B}}_c^{(L)}(y;v) \equiv E[\mathbf{B}_c^{(L)}(y;v)]$. The next lemma establishes uniform consistency of $\hat{f}_c(y|v)$.

Lemma 8.2 *Suppose that Assumptions C.1-C.7 hold. Then uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,*

- (a) $\hat{f}_c(y|v) - f(y|v) = e_2' \bar{\mathbf{S}}_c(v)^{-1} [\mathbf{V}_c^{(L)}(y;v) + \bar{\mathbf{B}}_c^{(L)}(y;v)] + O_P(\nu_{2c}^2 + \nu_{2c}c^{p+1})$,
- (b) $\hat{f}_c(y|v) - f(y|v) = O_P(\nu_{2c} + c^{p+1} + c^r)$,

where $\nu_{2c} \equiv n^{-1/2}c^{-(d+1)/2}\sqrt{\ln n}$.

Proof. The results follow from Lemma 10.5 in HSW (2011). ■

Lemma 8.3 *Suppose that Assumptions C.1-C.7 hold. Then*

- (a) $\hat{r}(y;v) - r(y;v) = b^{-1}e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v) f(y|v)^{-1} - D_x F(y|v) e_2' \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y;v) f(y|v)^{-2} + O_P(\nu_{bc})$ uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,
 - (b) $\hat{r}_0 - r_0 = O_P(\nu_{bc} + n^{-1/2}b^{-1})$,
 - (c) $\sup_{y \in \mathcal{Y}_0} |\hat{r}_2(y) - r_2(y)| = O_P(\nu_{bc} + n^{-1/2}b^{-1}\sqrt{\ln n})$,
- where $\nu_{bc} \equiv \nu_{1b}^2 b^{-1} + b^p + \nu_{2c}^2 + c^{p+1} + c^r + \nu_{1b}\nu_{2c}b^{-1} = o(n^{-1/2}b^{-(\frac{d}{2}+1)})$.

Proof. (a) Let $\hat{q}(y;v) \equiv \hat{r}(y;v) - r(y;v)$. Noting that $\hat{f}_c(y|v)^{-1} = f(y|v)^{-1} - [\hat{f}_c(y|v) - f(y|v)]/f(y|v)^2 + R_1(y;v)$ where $R_1(y;v) \equiv [\hat{f}_c(y|v) - f(y|v)]^2/[f(y|v)^2 \hat{f}_c(y|v)]$, we have that for any $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned} \hat{q}(y;v) &= \frac{D_x \hat{F}_b(y|v)}{\hat{f}_c(y|v)} - \frac{D_x F(y|v)}{f(y|v)} \\ &= \frac{D_x \hat{F}_b(y|v) - D_x F(y|v)}{f(y|v)} + \left[\frac{1}{\hat{f}_c(y|v)} - \frac{1}{f(y|v)} \right] D_x F(y|v) + R_2(y;v) \\ &= \frac{D_x \hat{F}_b(y|v) - D_x F(y|v)}{f(y|v)} - \frac{\hat{f}_c(y|v) - f(y|v)}{f(y|v)^2} D_x F(y|v) + R_1(y;v) D_x F(y|v) + R_2(y;v) \\ &\equiv \hat{q}_1(y;v) - \hat{q}_2(y;v) + R_1(y;v) D_x F(y|v) + R_2(y;v), \text{ say,} \end{aligned}$$

where $R_2(y;v) \equiv [\hat{f}_c(y|v)^{-1} - f(y|v)^{-1}] [D_x \hat{F}_b(y|v) - D_x F(y|v)]$. Using Lemmas 8.1 and 8.2 we can bound the last two terms in the last expression uniformly by $O_P(\eta_{2c}(\eta_{1b} + \eta_{2c}))$, where $\eta_{1b} \equiv \nu_{1b}b^{-1} + b^p$ and $\eta_{2c} \equiv \nu_{2c} + c^{p+1} + c^r$. In addition, uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned} \hat{q}_1(y;v) &= [D_x \hat{F}_b(y|v) - D_x F(y|v)]/f(y|v) \\ &= b^{-1}e_1 \bar{\mathbf{S}}_b(v)^{-1} [\mathbf{V}_b(y;v) + \bar{\mathbf{B}}_b(y;v)]/f(y|v) + O_P(\nu_{1b}^2 b^{-1} + \nu_{1b}b^p) \\ &= b^{-1}e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y;v)/f(y|v) + O_P(\nu_{1b}^2 b^{-1} + b^p), \end{aligned}$$

and

$$\begin{aligned}
\hat{q}_2(y; v) &= D_x F(y|v) [\hat{f}_c(y|v) - f(y|v)] / f(y|v)^2 \\
&= D_x F(y|v) e'_2 \bar{\mathbf{S}}_c(v)^{-1} [\mathbf{V}_c^{(L)}(y; v) + \bar{\mathbf{B}}_c^{(L)}(y; v)] / f(y|v)^2 + O_P(\nu_{2c}^2 + \nu_{2c} c^{p+1}) \\
&= D_x F(y|v) e'_2 \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y; v) / f(y|v)^2 + O_P(\nu_{2c}^2 + c^{p+1} + c^r).
\end{aligned}$$

It follows that uniformly in $(y, v) \in \mathcal{Y}_0 \times \mathcal{V}$,

$$\begin{aligned}
\hat{q}(y; v) &= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y; v) f(y|v)^{-1} - D_x F(y|v) e'_2 \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y; v) f(y|v)^{-2} \\
&\quad + O_P(\nu_{1b}^2 b^{-1} + b^p + \nu_{2c}^2 + c^{p+1} + c^r + \eta_{2c}(\eta_{1b} + \eta_{2c})) \\
&= b^{-1} e_1 \bar{\mathbf{S}}_b(v)^{-1} \mathbf{V}_b(y; v) f(y|v)^{-1} - D_x F(y|v) e'_2 \bar{\mathbf{S}}_c(v)^{-1} \mathbf{V}_c^{(L)}(y; v) f(y|v)^{-2} + O_P(\nu_{bc}).
\end{aligned}$$

(b) Write $\hat{r}_0 - r_0 = \hat{r}_{01} + \hat{r}_{02}$, where

$$\hat{r}_{01} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\hat{r}(Y_i; X_j, Z_j) - r(Y_i; X_j, Z_j)], \text{ and } \hat{r}_{02} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [r(Y_i; X_j, Z_j) - r_0].$$

It is easy to show that $\hat{r}_{02} = O_P(n^{-1/2})$ by the Chebyshev inequality. For \hat{r}_{01} , we have by (a) that $\hat{r}_{01} = R_{1n} - R_{2n} + O_P(\nu_{bc})$, where

$$\begin{aligned}
R_{1n} &\equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mathbf{V}_b(Y_i; V_j) f_{ij}^{-1} \mathbf{1}_i, \\
R_{2n} &\equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_{xij} e'_2 \bar{\mathbf{S}}_c(V_j)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_j) f_{ij}^{-2} \mathbf{1}_i,
\end{aligned}$$

$\mathbf{1}_i \equiv \mathbf{1}\{Y_i \in \mathcal{Y}_0\}$, $f_{ij} \equiv f(Y_i|V_j)$, and $D_{xij} \equiv D_x F(Y_i|V_j)$. For R_{1n} , we have

$$\begin{aligned}
R_{1n} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_k - V_j) K_b(V_k - V_j) \bar{\mathbf{I}}_{Y_i}(W_k) \mathbf{1}_i \\
&= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \sum_{k=1, k \neq j, i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_k - V_j) K_b(V_k - V_j) \bar{\mathbf{I}}_{Y_i}(W_k) \mathbf{1}_i \\
&\quad + \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1, j \neq i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_i - V_j) K_b(V_i - V_j) \bar{\mathbf{I}}_{Y_i}(W_i) \mathbf{1}_i \\
&\quad + \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1, j \neq i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(0) K_b(0) \bar{\mathbf{I}}_{Y_i}(W_j) \mathbf{1}_i \\
&\quad + \frac{1}{n^3} \sum_{i=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mu_b(0) K_b(0) \bar{\mathbf{I}}_{Y_i}(W_i) \mathbf{1}_i \\
&\equiv R_{1n,1} + R_{1n,2} + R_{1n,3} + R_{1n,4}.
\end{aligned}$$

It is easy to show that $R_{1n,4} = O_P(n^{-2}b^{-d-1})$, $R_{1n,3} = O_P(n^{-3/2}b^{-d-1})$, and $R_{1n,2} = O_P(n^{-1}b^{-1})$. Noting that $R_{1n,1}$ is a third-order U -statistic with $E(R_{1n,1}) = 0$, it is straightforward to show that $E(R_{1n,1}^2) = O(n^{-1}b^{-2} + n^{-2}b^{-d-2})$. Thus $R_{1n,1} = O_P(n^{-1/2}b^{-1})$ and $R_{1n} = O_P(n^{-1/2}b^{-1})$ as $n^{-1}b^{-d} = o(1)$. By the same token, we can show that $R_{2n} = O_P(n^{-1/2})$. It follows that $\hat{r}_0 - r_0 = O_P(\nu_{bc} + n^{-1/2}b^{-1})$.

(c) Write $\hat{r}_2(y) - r_2(y) = \hat{r}_{21}(y) + \hat{r}_{22}(y)$, where

$$\hat{r}_{21}(y) = \frac{1}{n} \sum_{i=1}^n [\hat{r}(y; V_i) - r(y; V_i)] \quad \text{and} \quad \hat{r}_{22}(y) = \frac{1}{n} \sum_{i=1}^n [r(y; V_i) - r_2(y)].$$

By standard chaining arguments and the exponential inequality, we can show that $\sup_{y \in \mathcal{Y}_0} \|\hat{r}_{22}(y)\| = O(n^{-1/2} \sqrt{\ln n})$. By (a), $\hat{r}_{21}(y) = \bar{r}_{21}(y) + O_P(\nu_{bc})$ uniformly in $y \in \mathcal{Y}_0$, where $\bar{r}_{21}(y) \equiv \bar{r}_{21,1}(y) - \bar{r}_{21,2}(y)$, $\bar{r}_{21,1}(y) \equiv \frac{1}{n} \sum_{i=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(y; V_i) f(y|V_i)^{-1} \mathbf{1}\{y \in \mathcal{Y}_0\}$, and $\bar{r}_{21,2}(y) \equiv \frac{1}{n} \sum_{i=1}^n f(y|V_i)^{-2} D_x F(y|V_i) \times e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(y; V_i) \mathbf{1}\{y \in \mathcal{Y}_0\}$. Now write $\bar{r}_{21,1}(y)$ as the summation of a first order U -statistic and a second order U -statistic: $\bar{r}_{21,1}(y) = \bar{r}_{21,11}(y) + \bar{r}_{21,12}(y)$, where

$$\begin{aligned} \bar{r}_{21,11}(y) &\equiv \frac{1}{n^2} \sum_{i=1}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mu_b(0) K_b(0) \bar{\mathbf{I}}_y(W_i) f(y|V_i)^{-1} \mathbf{1}\{y \in \mathcal{Y}_0\}, \quad \text{and} \\ \bar{r}_{21,12}(y) &\equiv \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n b^{-1} e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mu_b(V_j - V_i) K_b(V_j - V_i) \bar{\mathbf{I}}_y(W_j) f(y|V_i)^{-1} \mathbf{1}\{y \in \mathcal{Y}_0\}. \end{aligned}$$

By the exponential inequality, we can show that $\sup_{y \in \mathcal{Y}_0} \|\bar{r}_{21,11}(y)\| = O(n^{-3/2} b^{-d-1} \sqrt{\ln n})$. For $\bar{r}_{21,12}(y)$, one can follow the proof of (A.10) in Gozalo and Linton (2001) and show that $\sup_{y \in \mathcal{Y}_0} \|\bar{r}_{21,12}(y)\| = O(n^{-1/2} b^{-1} \sqrt{\ln n})$.⁶ Hence $\sup_{y \in \mathcal{Y}_0} \|\bar{r}_{21,1}(y)\| = O(n^{-1/2} b^{-1} \sqrt{\ln n})$. Similarly, $\sup_{y \in \mathcal{Y}_0} \|\bar{r}_{21,2}(y)\| = O(n^{-1/2} \sqrt{\ln n})$. Thus $\sup_{y \in \mathcal{Y}_0} \|\hat{r}_2(y) - r_2(y)\| = O_P(\nu_{bc} + n^{-1/2} b^{-1} \sqrt{\ln n})$. ■

Proof of Theorem 3.1

Let $a_i \equiv a(Y_i; X_i, Z_i)$, $r_i \equiv r(Y_i; X_i, Z_i)$, $r_{1i} \equiv r_1(X_i)$, $r_{2i} \equiv r_2(Y_i)$, $r_{2i}^\pi \equiv r_2^\pi(Y_i)$, $\hat{r}_i \equiv \hat{r}(Y_i; X_i, Z_i)$, $\hat{r}_{1i} \equiv \hat{r}_1(X_i)$, $\hat{r}_{2i} \equiv \hat{r}_2(Y_i)$, and $\hat{r}_{2i}^\pi \equiv \hat{r}_2^\pi(Y_i)$. Then

$$\begin{aligned} nb^{\frac{d}{2}+2} \hat{\Gamma} &= b^{\frac{d}{2}+2} \sum_{i=1}^n \|\hat{r}_i \hat{r}_0^\pi - \hat{r}_{1i} \hat{r}_{2i}^\pi\|^2 a_i \\ &= b^{\frac{d}{2}+2} \sum_{i=1}^n \|[(\hat{r}_i - r_i) + r_i][(\hat{r}_0^\pi - r_0^\pi) + r_0^\pi] - [(\hat{r}_{1i} - r_{1i}) + r_{1i}][(\hat{r}_{2i}^\pi - r_{2i}^\pi) + r_{2i}^\pi]\|^2 a_i \\ &= b^{\frac{d}{2}+2} \sum_{i=1}^n \|[r_i r_0^\pi - r_{1i} r_{2i}^\pi] + [(\hat{r}_i - r_i) r_0^\pi + r_i (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi - r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi)] \\ &\quad + [(\hat{r}_i - r_i) (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)]\|^2 a_i \\ &= \Gamma_{1n} + \Gamma_{2n} + \Gamma_{3n} + 2\Gamma_{4n} + 2\Gamma_{5n} + 2\Gamma_{6n}, \end{aligned} \tag{8.3}$$

⁶If we ignore the boundary points, we can write $\bar{\mathbf{S}}_b(v) = f(v) \mathbb{S} + b \mathbb{V}(v) + o(b)$ uniformly in v in the interior of \mathcal{V} , where \mathbb{S} and \mathbb{V} are defined as M and V in Li, Lu and Ullah (2003, p.617). Following the proof of Lemma A.3 in their paper, one can show that $\bar{r}_{21,12}(y) = O(n^{-1/2})$ elementwise by using the degeneracy of the second order U -statistic defined analogously to $\bar{r}_{21,12}(y)$ but with $\bar{\mathbf{S}}_b(V_i)^{-1}$ replaced by its leading term $f(V_i)^{-1} \mathbb{S}^{-1}$. But their argument breaks down when v takes values on the boundary of \mathcal{V} .

where

$$\begin{aligned}
\Gamma_{1n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_i r_0^\pi - r_{1i} r_{2i}^\pi\|^2 a_i, \\
\Gamma_{2n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_i - r_i) r_0^\pi + r_i (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi - r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi)\|^2 a_i, \\
\Gamma_{3n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_i - r_i) (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)\|^2 a_i, \\
\Gamma_{4n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n [r_i r_0^\pi - r_{1i} r_{2i}^\pi]' [(\hat{r}_i - r_i) r_0^\pi + r_i (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi - r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi)] a_i, \\
\Gamma_{5n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n [r_i r_0^\pi - r_{1i} r_{2i}^\pi]' [(\hat{r}_i - r_i) (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)] a_i, \\
\Gamma_{6n} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n [(\hat{r}_i - r_i) r_0^\pi + r_i (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi - r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi)]' \\
&\quad \times [(\hat{r}_i - r_i) (\hat{r}_0^\pi - r_0^\pi) - (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)] a_i.
\end{aligned}$$

Under \mathbb{H}_0 , $\Gamma_{jn} = 0$ for $j = 1, 4, 5$. It suffices to prove the theorem by showing that (i) $\Gamma_{2n} - \mathbb{B}_n \xrightarrow{d} N(0, \sigma_0^2)$, (ii) $\Gamma_{3n} = o_P(1)$, and (iii) $\Gamma_{6n} = o_P(1)$.

To show (i), we write $\Gamma_{2n} = \sum_{j=1}^{10} \Gamma_{2n,j}$ where

$$\begin{aligned}
\Gamma_{2n,1} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_i - r_i) r_0^\pi\|^2 a_i, & \Gamma_{2n,6} &\equiv -2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' (\hat{r}_{1i} - r_{1i}) r_0^\pi r_{2i}^\pi a_i, \\
\Gamma_{2n,2} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_i (\hat{r}_0^\pi - r_0^\pi)\|^2 a_i, & \Gamma_{2n,7} &\equiv -2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' r_{1i} r_0^\pi (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i, \\
\Gamma_{2n,3} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_{1i} - r_{1i}) r_{2i}^\pi\|^2 a_i, & \Gamma_{2n,8} &\equiv -2b^{\frac{d}{2}+2} \sum_{i=1}^n r_i' (\hat{r}_{1i} - r_{1i}) (\hat{r}_0^\pi - r_0^\pi) r_{2i}^\pi a_i, \\
\Gamma_{2n,4} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi)\|^2 a_i, & \Gamma_{2n,9} &\equiv -2b^{\frac{d}{2}+2} \sum_{i=1}^n r_i' r_{1i} (\hat{r}_0^\pi - r_0^\pi) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i, \\
\Gamma_{2n,5} &\equiv 2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' r_i r_0^\pi (\hat{r}_0^\pi - r_0^\pi) a_i, & \Gamma_{2n,10} &\equiv 2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_{1i} - r_{1i})' r_{1i} r_{2i}^\pi (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i.
\end{aligned}$$

By Lemmas 8.4, 8.5(b) and 8.6(b) below, $\Gamma_{2n,1} + \Gamma_{2n,3} + \Gamma_{2n,6} - \mathbb{B}_n \xrightarrow{d} N(0, \sigma_0^2)$, where $\mathbb{B}_n \equiv \mathbb{B}_{1n} + \mathbb{B}_{2n} - 2\mathbb{B}_{3n}$. By Lemmas 8.5(a) and (c) and Lemmas 8.6(a) and (c)-(f), $\Gamma_{2n,s} = o_P(1)$ for $s = 2, 4, 5, 7, \dots, 10$. It follows that $\Gamma_{2n} - \mathbb{B}_n \xrightarrow{d} N(0, \sigma_0^2)$.

Next, we show (ii). By the Cauchy-Schwarz inequality, $\Gamma_{3n} \leq 2\Gamma_{3n,1} + 2\Gamma_{3n,2}$, where $\Gamma_{3n,1} \equiv b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi)^2 \times \sum_{i=1}^n \|\hat{r}_i - r_i\|^2 a_i$ and $\Gamma_{3n,2} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)\|^2 a_i$. By Lemmas 8.3(b) and 8.4,

$$\Gamma_{3n,1} = (\hat{r}_0^\pi - r_0^\pi)^2 (r_0^\pi)^{-2} \Gamma_{2n,1} = O_P(\nu_{bc}^2 + n^{-1}b^{-2}) O_P\left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1})\right) = o_P(1).$$

Following the proof of Lemma 8.5(b), we can show that $\bar{\Gamma}_{3n,2} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|\hat{r}_{1i} - r_{1i}\|^2 a_i = O_P(b^{\frac{d}{2}+2} (b^{-d_x-2} + c^{-d_x}))$. It follows that

$$\Gamma_{3n,2} \leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)|^2 \bar{\Gamma}_{3n,2} = O_P(\nu_{bc}^2 + n^{-1}b^{-2} \ln n) O_P\left(b^{\frac{d}{2}+2} (b^{-d_x-2} + c^{-d_x})\right) = o_P(1),$$

and hence $\Gamma_{3n} = o_P(1)$.

To show (iii), we first decompose Γ_{6n} as follows:

$$\begin{aligned}
\Gamma_{6n} &= b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi) r_0^\pi \sum_{i=1}^n \|\hat{r}_i - r_i\|^2 a_i - b^{\frac{d}{2}+2} r_0^\pi \sum_{i=1}^n (\hat{r}_i - r_i)' (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i \\
&\quad + b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi)^2 \sum_{i=1}^n r_i' (\hat{r}_i - r_i) a_i - b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi) \sum_{i=1}^n r_i' (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i \\
&\quad - b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi) \sum_{i=1}^n (\hat{r}_{1i} - r_{1i})' (\hat{r}_i - r_i) r_{2i}^\pi a_i + b^{\frac{d}{2}+2} \sum_{i=1}^n \|\hat{r}_{1i} - r_{1i}\|^2 (\hat{r}_{2i}^\pi - r_{2i}^\pi) r_{2i}^\pi a_i \\
&\quad - b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi) \sum_{i=1}^n r_{1i}' (\hat{r}_i - r_i) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i + b^{\frac{d}{2}+2} \sum_{i=1}^n r_{1i}' (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi)^2 r_{2i}^\pi a_i \\
&\equiv \Gamma_{6n,1} - \Gamma_{6n,2} + \Gamma_{6n,3} - \Gamma_{6n,4} - \Gamma_{6n,5} + \Gamma_{6n,6} - \Gamma_{6n,7} + \Gamma_{6n,8}.
\end{aligned}$$

By Lemmas 8.3(b) and 8.4,

$$\Gamma_{6n,1} = (\hat{r}_0^\pi - r_0^\pi) (r_0^\pi)^{-1} \Gamma_{2n,1} = O_P \left(\nu_{bc} + n^{-1/2} b^{-1} \right) O_P \left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) \right) = o_P(1).$$

Using $\bar{\Gamma}_{3n,2}$ and $\Gamma_{2n,1}$ defined above, by Lemmas 8.3(c) and 8.4 and the Cauchy-Schwarz inequality,

$$\begin{aligned}
|\Gamma_{6n,2}| &\leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)| b^{\frac{d}{2}+2} \sum_{i=1}^n |(\hat{r}_i - r_i)' (\hat{r}_{1i} - r_{1i}) r_0^\pi| a_i \\
&\leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)| (\Gamma_{2n,1} \bar{\Gamma}_{3n,2})^{1/2} \\
&= O_P \left(\nu_{bc} + n^{-1/2} b^{-1} \sqrt{\ln n} \right) \left\{ O_P \left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) \right) O_P \left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) \right) \right\}^{1/2} \\
&= o_P(1).
\end{aligned}$$

By Lemmas 8.3(b) and 8.6(a), $\Gamma_{6n,3} = (\hat{r}_0^\pi - r_0^\pi) \Gamma_{2n,5} / (2r_0^\pi) = o_P(1) o_P(1) = o_P(1)$. Following the proof of Lemma 8.6(f), we can show that $\bar{\Gamma}_{6n,4} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n r_i' (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i = o_P(1)$. This, in conjunction with Lemma 8.3(b), implies that $\Gamma_{6n,4} = (\hat{r}_0^\pi - r_0^\pi) \bar{\Gamma}_{6n,4} = o_P(1)$. By Lemma 8.3 and 8.6(b), $\Gamma_{6n,5} = -(\hat{r}_0^\pi - r_0^\pi) \Gamma_{2n,6} / (2r_0^\pi) = o_P(1)$. Analogously to the proof of Lemma 8.4, we can show that $\bar{\Gamma}_{6n,6} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|\hat{r}_{1i} - r_{1i}\|^2 |r_{2i}^\pi| a_i = O_P(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}))$. Combining this with Lemma 8.3(c) yields

$$|\Gamma_{6n,6}| \leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)| \bar{\Gamma}_{6n,6} = O_P \left(\nu_{bc} + n^{-1/2} b^{-1} \sqrt{\ln n} \right) O_P \left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) \right) = o_P(1).$$

By Lemmas 8.3(b) and 8.6(c), $\Gamma_{6n,7} = -(\hat{r}_0^\pi - r_0^\pi) \Gamma_{2n,7} / (2r_0^\pi) = o_P(1)$. Lastly, by the Cauchy-Schwarz inequality, Lemmas 8.3(c), 8.5(b), and 8.5(c), we have

$$\begin{aligned}
|\Gamma_{6n,8}| &\leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)| b^{\frac{d}{2}+2} \sum_{i=1}^n |r_{1i}' (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi) r_{2i}^\pi| a_i \\
&\leq \sup_{y \in \mathcal{Y}_0} |\hat{r}_2^\pi(y) - r_2^\pi(y)| \{ \Gamma_{2n,3} \Gamma_{2n,4} \}^{1/2} \\
&= O_P \left(\nu_{bc} + n^{-1/2} b^{-1} \sqrt{\ln n} \right) \left\{ O_P \left(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) \right) o_P(1) \right\}^{1/2} \\
&= o_P(1).
\end{aligned}$$

Consequently we have proved $\Gamma_{6n} = o_P(1)$. ■

Lemma 8.4 *Suppose Assumptions C.1-C.7 hold. Then $\Gamma_{2n,1} - \mathbb{B}_{1n} \xrightarrow{d} N(0, \sigma_0^2)$ where $\mathbb{B}_{1n} = n^{-1}b^{\frac{d}{2}+2}(r_0^\pi)^2 \times \sum_{i=1}^n \varphi(W_i, W_i) = O_P(b^{\frac{d}{2}+2}(b^{-d-2} + c^{-d-1}))$.*

Proof. Recall $\mathbf{1}_i \equiv \mathbf{1}\{Y_i \in \mathcal{Y}_0\}$. Let $f_i \equiv f(Y_i|V_i)$ and $D_{xi} \equiv D_x F(Y_i|V_i)$. Then

$$\hat{r}_i - r_i = b^{-1}e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(Y_i; V_i) f_i^{-1} \mathbf{1}_i - D_{xi} e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2} \mathbf{1}_i + O_P(\nu_{bc}).$$

It follows that

$$\begin{aligned} \Gamma_{2n,1} &= b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_i - r_i) r_0^\pi\|^2 a_i \\ &= b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \left\| b^{-1}e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(Y_i; V_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2} \right\|^2 a_i \\ &\quad + nb^{\frac{d}{2}+2} O_P(\nu_{bc}^2 + \nu_{bc}(\nu_{1b}b^{-1} + b^p + \nu_{2c} + c^{p+1} + c^r)) \\ &= \bar{\Gamma}_{2n,1} + o_P(1), \end{aligned}$$

where $\bar{\Gamma}_{2n,1} \equiv b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \left\| b^{-1}e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(Y_i; V_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2} \right\|^2 a_i$,⁷ and we use the fact that $\mathbf{1}_i a_i = a_i$ as $a(y; v)$ has compact support $\mathcal{Y}_0 \times \mathcal{V}_0$. Let $\zeta_k(w) \equiv \zeta_k(y; v)$ be as defined in Section 3.2. Then

$$\begin{aligned} \bar{\Gamma}_{2n,1} &= b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \left\| n^{-1} \sum_{k=1}^n \zeta_k(W_i) \right\|^2 a_i \\ &= n^{-2} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \zeta(W_{i_1}, W_{i_2}, W_{i_3}), \end{aligned}$$

where $\zeta(W_{i_1}, W_{i_2}, W_{i_3}) \equiv \zeta_{i_2}(W_{i_1})' \zeta_{i_3}(W_{i_1}) a_{i_1}$. Let $\varphi(w_{i_1}, w_{i_2}) \equiv E[\zeta(W_{i_1}, w_{i_1}, w_{i_2})]$, and $\bar{\zeta}(w_{i_1}, w_{i_2}, w_{i_3}) \equiv \zeta(w_{i_1}, w_{i_2}, w_{i_3}) - \varphi(w_{i_2}, w_{i_3})$. We can decompose $\bar{\Gamma}_{2n,1}$ as $\bar{\Gamma}_{2n,1} = \bar{\Gamma}_{2n,11} + \bar{\Gamma}_{2n,12}$, where

$$\bar{\Gamma}_{2n,11} = n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i_1=1}^n \sum_{i_2=1}^n \varphi(W_{i_1}, W_{i_2}) \quad \text{and} \quad \bar{\Gamma}_{2n,12} = n^{-2} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}).$$

Consider $\bar{\Gamma}_{2n,12}$ first. Write $E(\bar{\Gamma}_{2n,12}^2) = n^{-4} b^{d+4} (r_0^\pi)^4 \sum_{i_1, \dots, i_6}^n E[\bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}) \bar{\zeta}(W_{i_4}, W_{i_5}, W_{i_6})]$. Noting that $E[\bar{\zeta}(W_{i_1}, w_{i_2}, w_{i_3})] = E[\bar{\zeta}(w_{i_1}, W_{i_2}, w_{i_3})] = E[\bar{\zeta}(w_{i_1}, w_{i_2}, W_{i_3})] = 0$, $E[\bar{\zeta}(W_{i_1}, W_{i_2}, W_{i_3}) \bar{\zeta}(W_{i_4}, W_{i_5}, W_{i_6})] = 0$ if there are more than three distinct elements in $\{i_1, \dots, i_6\}$. With this, it is easy to show that $E(\bar{\Gamma}_{2n,12}^2) = O(n^{-1} b^{d+4} (b^{-4-3d} + c^{-3(d+1)})) = o(1)$. Hence $\bar{\Gamma}_{2n,12} = o_P(1)$ by the Chebyshev inequality.

For $\bar{\Gamma}_{2n,11}$, we have

$$\begin{aligned} \bar{\Gamma}_{2n,11} &= n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \varphi(W_i, W_i) + 2n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{1 \leq i < j \leq n} \varphi(W_i, W_j) \\ &\equiv \mathbb{B}_{1n} + \mathbb{V}_{1n}, \quad \text{say,} \end{aligned} \tag{8.4}$$

⁷Write $\bar{\Gamma}_{2n,1} = b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{i=1}^n \|\xi_{1ni} - \xi_{2ni}\|^2 a_i$ where $\xi_{1ni} \equiv b^{-1}e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(Y_i; V_i) f_i^{-1}$ and $\xi_{2ni} \equiv D_{xi} e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_i) f_i^{-2}$. By straightforward moment calculations, we can show that ξ_{1ni} contributes to both the asymptotic bias and variance of the test statistic whereas ξ_{2ni} only contributes to the asymptotic bias.

where $\varphi(W_i, W_j) = \int \zeta(w, W_i, W_j) dF(w) = \int \zeta_i(w)' \zeta_j(w) a(w) dF(w)$, and \mathbb{B}_{1n} and \mathbb{V}_{1n} contribute to the asymptotic bias and variance of $\bar{\Gamma}_{2n,11}$, respectively. Note that as \mathbb{V}_{1n} 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: $\mathbb{V}_{1n} \xrightarrow{d} N(0, \sigma_0^2)$, where $\sigma_0^2 = \lim_{n \rightarrow \infty} \sigma_n^2$ and $\sigma_n^2 = 2b^{d+4} (r_0^\pi)^4 E[\varphi(W_1, W_2)]^2$. Thus $\Gamma_{2n,11} - \mathbb{B}_{1n} \xrightarrow{d} N(0, \sigma_0^2)$.

Lastly, noting that $E|\mathbb{B}_{1n}| = b^{\frac{d}{2}+2} O(b^{-d-2} + c^{-d-1})$, we have $\mathbb{B}_{1n} = O_P(b^{\frac{d}{2}+2}(b^{-d-2} + c^{-d-1}))$ by the Markov inequality. ■

Lemma 8.5 *Suppose Assumptions C.1-C.7 hold. Then*

- (a) $\Gamma_{2n,2} = b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_i(\hat{r}_0^\pi - r_0^\pi)\|^2 a_i = o_P(1)$,
- (b) $\Gamma_{2n,3} = b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_{1i} - r_{1i}) r_{2i}^\pi\|^2 a_i = \mathbb{B}_{2n} + o_P(1)$,
- (c) $\Gamma_{2n,4} = b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_{1i}(\hat{r}_{2i}^\pi - r_{2i}^\pi)\|^2 a_i = o_P(1)$

where $\mathbb{B}_{2n} \equiv n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^n \left\| \sum_{j=1}^n \sum_{k=1}^n \zeta_k(Y_j; X_i, Z_j) r_{2i}^\pi \right\|^2 a_i = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}))$. If $d_z > 0$, then (b) also holds when we replace \mathbb{B}_{2n} by $\bar{\mathbb{B}}_{2n} \equiv n^{-4} b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \left\| \sum_{k=1}^n \zeta_j(Y_k; X_i, Z_k) r_{2i}^\pi \right\|^2 a_i$.

Proof. (a) Note that $\Gamma_{2n,2} = n b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi)^2 \bar{\Gamma}_{2n,2}$, where $\bar{\Gamma}_{2n,2} \equiv n^{-1} \sum_{i=1}^n \|r_i\|^2 a_i$. By Assumptions C.2(ii) and C.3(i), the compact support of a , and the Markov inequality, $\bar{\Gamma}_{2n,2} = O_P(1)$. Using this and Lemma 8.3(a) we have $\Gamma_{2n,2} = n b^{\frac{d}{2}+2} O_P(n^{-1} b^{-2}) O_P(1) = O_P(b^{\frac{d}{2}}) = o_P(1)$.

(b) Noting that $\hat{r}_1(x) - r_1(x) = \frac{1}{n} \sum_{i=1}^n [\hat{r}(Y_i; x, Z_i) - r(Y_i; x, Z_i)] + \frac{1}{n} \sum_{i=1}^n [r(Y_i; x, Z_i) - r_1(x)]$, we have $b^{\frac{d}{2}+2} \sum_{i=1}^n \|[\hat{r}_1(X_i) - r_1(X_i)] r_{2i}^\pi\|^2 a_i = R_{3n} + R_{4n} + 2R_{5n}$, where

$$\begin{aligned} R_{3n} &\equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)] r_{2i}^\pi \right\|^2 a_i, \\ R_{4n} &\equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n [r(Y_j; X_i, Z_j) - r(X_i)] r_{2i}^\pi \right\|^2 a_i, \text{ and} \\ R_{5n} &\equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)]' [r(Y_j; X_i, Z_j) - r(X_i)] (r_{2i}^\pi)^2 a_i. \end{aligned}$$

By Lemma 8.3(a) we can readily show that $R_{3n} = \mathbb{B}_{2n} + o_P(1)$. We further decompose \mathbb{B}_{2n} as $\mathbb{B}_{2n} = \mathbb{B}_{2n,1} + \mathbb{B}_{2n,2}$, where

$$\begin{aligned} \mathbb{B}_{2n,1} &\equiv \frac{b^{\frac{d}{2}+2}}{n^4} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \zeta_{i_3}(Y_{i_2}; X_{i_1}, Z_{i_2})' \zeta_{i_3}(Y_{i_4}; X_{i_1}, Z_{i_4}) (r_{2i_1}^\pi)^2 a_{i_1}, \text{ and} \\ \mathbb{B}_{2n,2} &\equiv \frac{b^{\frac{d}{2}+2}}{n^4} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \sum_{i_5=1, i_5 \neq i_3}^n \zeta_{i_3}(Y_{i_2}; X_{i_1}, Z_{i_2})' \zeta_{i_5}(Y_{i_4}; X_{i_1}, Z_{i_4}) (r_{2i_1}^\pi)^2 a_{i_1}. \end{aligned}$$

By direct moment calculations and the Chebyshev inequality, we can show that $\mathbb{B}_{2n,2} = O_P(b^{\frac{d_z}{2}} + b^{\frac{d}{2}+2} c^{-\frac{d_x}{2}})$ which is $o_P(1)$ under Assumption A.7 if $d_z > 0$, and that

$$\mathbb{B}_{2n,1} = \bar{\mathbb{B}}_{2n} = \frac{1}{n^4} b^{\frac{d}{2}+2} \sum_{i_1=1}^n \sum_{i_3=1}^n \left\| \sum_{i_2=1}^n \zeta_{i_3}(Y_{i_2}; X_{i_1}, Z_{i_2}) r_{2i_1}^\pi \right\|^2 a_{i_1} = O_P(b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x})).$$

It follows that $R_{3n} = \bar{\mathbb{B}}_{2n} + o_P(1)$ if $d_z > 0$. By the Markov inequality, $R_{4n} = O_P(b^{\frac{d}{2}+2}) = o_P(1)$. By the Cauchy-Schwarz inequality, $R_{5n} \leq \{R_{3n}R_{4n}\}^{1/2} = O_P(\{[b^{\frac{d}{2}+2}(b^{-d_x-2} + c^{-d_x}) + 1]b^{\frac{d}{2}+2}\}^{1/2}) = o_P(1)$. This completes the proof of part (b).

(c) Noting that $\hat{r}_2(y) - r_2(y) = \frac{1}{n} \sum_{i=1}^n [\hat{r}(y; X_i, Z_i) - r(y; X_i, Z_i)] + \frac{1}{n} \sum_{i=1}^n [r(y; X_i, Z_i) - r_2(y)]$, by the Cauchy-Schwarz inequality we have $\Gamma_{2n,4} \leq 2R_{6n} + 2R_{7n}$, where

$$\begin{aligned} R_{6n} &\equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n r_{1i} \pi' [\hat{r}(Y_i; X_j, Z_j) - r(Y_i; X_j, Z_j)] \right\|^2 a_i \text{ and} \\ R_{7n} &\equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n r_{1i} \pi' [r(Y_i; X_j, Z_j) - r_2(Y_i)] \right\|^2 a_i. \end{aligned}$$

By the Markov inequality $R_{7n} = O_P(b^{\frac{d}{2}+2}) = o_P(1)$. For R_{6n} we can first apply Lemma 8.3 to show that $R_{6n} = \bar{R}_{6n} + o_P(1)$, where

$$\bar{R}_{6n} \equiv \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n r_{1i} \pi' \left[b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mathbf{V}_b(Y_i; V_j) f_{ij}^{-1} - f_{ij}^{-2} D_{xij} e_2' \bar{\mathbf{S}}_c(X_j, Z_j)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_j) \right] \right\|^2 a_i,$$

$f_{ij} \equiv f(Y_i|V_j)$, and $D_{xij} \equiv D_x F(Y_i|V_j)$. Observe that $\bar{R}_{6n} \leq 2\bar{R}_{6n,1} + 2\bar{R}_{6n,2}$, where

$$\begin{aligned} \bar{R}_{6n,1} &= \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n r_{1i} \pi' b^{-1} e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mathbf{V}_b(Y_i; V_j) f_{ij}^{-1} \right\|^2 a_i, \text{ and} \\ \bar{R}_{6n,2} &= \frac{b^{\frac{d}{2}+2}}{n^2} \sum_{i=1}^n \left\| \sum_{j=1}^n r_{1i} \pi' f_{ij}^{-2} D_{xij} e_2' \bar{\mathbf{S}}_c(V_j)^{-1} \mathbf{V}_c^{(L)}(Y_i; V_j) \right\|^2 a_i. \end{aligned}$$

$$\begin{aligned} E(\bar{R}_{6n,1}) &= \frac{b^{\frac{d}{2}}}{n^4} \sum_{i=1}^n E \left\| \sum_{j=1}^n \sum_{k=1}^n r_{1i} \pi' e_1 \bar{\mathbf{S}}_b(V_j)^{-1} \mu_b(V_k - V_j) K_b(V_k - V_j) \bar{\mathbf{I}}_{Y_i}(W_k) f_{ij}^{-1} \right\|^2 a_i \\ &= \frac{b^{\frac{d}{2}}}{n^4} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \sum_{i_5=1}^n E \left\{ \left[r_{1i_1} \pi' e_1 \bar{\mathbf{S}}_b(V_{i_2})^{-1} \mu_b(V_{i_3} - V_{i_2}) K_b(V_{i_3} - V_{i_2}) \bar{\mathbf{I}}_{Y_{i_1}}(W_{i_3}) f_{i_1 i_2}^{-1} \right] \right. \\ &\quad \left. \times \left[r_{1i_4} \pi' e_1 \bar{\mathbf{S}}_b(V_{i_4})^{-1} \mu_b(V_{i_5} - V_{i_4}) K_b(V_{i_5} - V_{i_4}) \bar{\mathbf{I}}_{Y_{i_4}}(W_{i_5}) f_{i_4 i_4}^{-1} \right] a_{i_1} \right\} \\ &= O\left(b^{\frac{d}{2}} + n^{-1} b^{-\frac{d}{2}} + n^{-2} b^{-\frac{3d}{2}}\right) = o(1). \end{aligned}$$

Similarly, $E(\bar{R}_{6n,2}) = b^{\frac{d}{2}+2} O(1 + n^{-1} c^{-(d+1)} + n^{-2} c^{-2(d+1)}) = o(1)$. Thus $R_{6n} = o_P(1)$ by the Markov inequality, and $b^{\frac{d}{2}+2} \sum_{i=1}^n [\hat{r}_2(Y_i) - r_2(Y_i)]^2 = o_P(1)$. ■

Lemma 8.6 *Suppose that Assumptions C.1-C.7 hold. Then*

- (a) $\Gamma_{2n,5} = 2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' r_i r_0^\pi (\hat{r}_0^\pi - r_0^\pi) a_i = o_P(1)$,
- (b) $\Gamma_{2n,6} = -2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' (\hat{r}_{1i} - r_{1i}) r_0^\pi r_{2i}^\pi a_i = -2\mathbb{B}_{3n} + o_P(1)$,
- (c) $\Gamma_{2n,7} = -2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' r_{1i} r_0^\pi (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i = o_P(1)$,
- (d) $\Gamma_{2n,8} = -2b^{\frac{d}{2}+2} \sum_{i=1}^n r_i' (\hat{r}_{1i} - r_{1i}) (\hat{r}_0^\pi - r_0^\pi) r_{2i}^\pi a_i = o_P(1)$,

$$\begin{aligned}
(e) \Gamma_{2n,9} &= -2b^{\frac{d}{2}+2} \sum_{i=1}^n r'_i r_{1i} (\hat{r}_0^\pi - r_0^\pi) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i = o_P(1), \\
(f) \Gamma_{2n,10} &= 2b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_{1i} - r_{1i})' r_{1i} r_{2i}^\pi (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i = o_P(1), \\
\text{where } \mathbb{B}_{3n} &\equiv n^{-3} b^{\frac{d}{2}+2} r_0^\pi \sum_{i=1}^n \sum_{l=1}^n \zeta_l (W_i)' \sum_{j=1}^n \sum_{k=1}^n \zeta_k (Y_j; X_i, Z_j) r_{2i}^\pi (Y_i) a_i = O_P(b^{(d_z-d_x)/2}).
\end{aligned}$$

Proof. (a) Write $\Gamma_{2n,5} = 2b^{\frac{d}{2}+2} r_0^\pi (\hat{r}_0^\pi - r_0^\pi) \bar{\Gamma}_{2n,5}$ where $\bar{\Gamma}_{2n,5} \equiv \sum_{i=1}^n (\hat{r}_i - r_i)' r_i a_i$. By Lemma 8.3(a), we can show that $\bar{\Gamma}_{2n,5} = \bar{\Gamma}_{2n,51} + o_P(n^{1/2} b^{-(\frac{d}{2}+1)})$, where

$$\begin{aligned}
\bar{\Gamma}_{2n,51} &= \sum_{i=1}^n \left(b^{-1} e_1 \bar{\mathbf{S}}_b (V_i)^{-1} \mathbf{V}_b (Y_i; V_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c (V_i)^{-1} \mathbf{V}_c^{(L)} (Y_i; V_i) f_i^{-2} \right)' r_i a_i \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \zeta_{1j} (Y_i; V_i)' r_i a_i - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \zeta_{2j} (Y_i; V_i)' r_i a_i \equiv R_{8n} + R_{9n}.
\end{aligned}$$

In view of $R_{8n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n [b^{-1} e_1 \bar{\mathbf{S}}_b (V_i)^{-1} \mu_b (V_j - V_i) K_b (V_j - V_i) \bar{\mathbf{I}}_{Y_i} (W_j) f_i^{-1}]' r_i a_i$, it is easy to show that $E(R_{8n}^2) = O(nb^{-2} + b^{-2d-2})$. Thus $R_{8n} = O_P(b^{-d-1} + n^{1/2} b^{-1})$. Similarly, $R_{9n} = O_P(c^{-d-1} + n^{1/2})$. It follows that $\bar{\Gamma}_{2n,51} = O_P(b^{-d-1} + n^{1/2} b^{-1} + c^{-d-1})$, and by Lemma 8.3(b),

$$\begin{aligned}
\Gamma_{2n,5} &= O_P(n^{-1/2} b^{\frac{d}{2}+1}) \left[O_P(b^{-d-1} + n^{1/2} b^{-1} + c^{-d-1}) + o_P(n^{1/2} b^{-(\frac{d}{2}+1)}) \right] \\
&= O_P(n^{-1/2} b^{-d/2} + b^{\frac{d}{2}} + n^{-1/2} b^{\frac{d}{2}+1} c^{-d-1}) + o_P(1) = o_P(1).
\end{aligned}$$

(b) Write $\Gamma_{2n,6} = -2r_0^\pi \bar{\Gamma}_{2n,6}$ where $\bar{\Gamma}_{2n,6} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi a_i$. Then $\bar{\Gamma}_{2n,6} = R_{10n} + R_{11n}$, where

$$\begin{aligned}
R_{10n} &\equiv n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n (\hat{r}_i - r_i)' [\hat{r} (Y_j; X_i, Z_j) - r (Y_j; X_i, Z_j)] r_{2i}^\pi a_i, \\
R_{11n} &\equiv n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n (\hat{r}_i - r_i)' [r (Y_j; X_i, Z_j) - r (X_i)] r_{2i}^\pi a_i.
\end{aligned}$$

Using Lemma 8.3, we can show that $R_{10n} = \bar{R}_{10n} + o_P(1)$, where

$$\begin{aligned}
\bar{R}_{10n} &= n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \left[b^{-1} e_1 \bar{\mathbf{S}}_b (V_i)^{-1} \mathbf{V}_b (Y_i; V_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c (V_i)^{-1} \mathbf{V}_c^{(L)} (Y_i; V_i) f_i^{-2} \right]' \\
&\quad \times \left[b^{-1} e_1 \bar{\mathbf{S}}_b (X_i, Z_j)^{-1} \mathbf{V}_b (Y_j; X_i, Z_j) \bar{f}_{ji}^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c (X_i, Z_j)^{-1} \mathbf{V}_c^{(L)} (Y_j; X_i, Z_j) \bar{f}_{ji}^{-2} \right] \mathbf{1}_i \mathbf{1}_j r_{2i}^\pi a_i \\
&= \mathbb{B}_{3n} / r_0^\pi.
\end{aligned}$$

Noting that $E(\bar{R}_{10n}^2) = O(b^{d+4} (b^{-2d_x-4} + c^{-2d_x}))$, we have $\bar{R}_{10n} = O_P(b^{(d_z-d_x)/2} + b^{\frac{d}{2}+2} c^{-d_x})$ which is $o_P(1)$ under Assumption A.7 if $d_z > d_x$ and otherwise not. Hence $R_{10n} = \mathbb{B}_{3n} / r_0^\pi + o_P(1)$. For R_{11n} , we apply the Cauchy-Schwarz inequality to obtain $R_{11n} \leq \{\alpha_{1n} \alpha_{2n}\}^{1/2}$, where $\alpha_{1n} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n \|(\hat{r}_i - r_i) r_{2i}^\pi\|^2 a_i$ and $\alpha_{2n} \equiv n^{-1} b^{\frac{d}{2}+2} \sum_{i=1}^n \left\| \sum_{j=1}^n [r (Y_j; X_i, Z_j) - r (X_i)] \right\|^2 a_i$. Analogously to the determination of the probability order of $\Gamma_{2n,1}$, we can show that $\alpha_{1n} = O_P(b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}))$. Next, $\alpha_{2n} = O_P(b^{\frac{d}{2}+2})$ by the Markov inequality. It follows that $R_{11n} = O_P(b^{-\frac{d}{4}} + b^{\frac{d}{4}+1} c^{-\frac{d+1}{2}}) O_P(b^{\frac{d}{4}+1}) = O_P(b + b^{\frac{d}{2}+2} c^{-\frac{d+1}{2}}) = o_P(1)$, and $\Gamma_{2n,6} = -2\mathbb{B}_{3n} + o_P(1)$.

(c) Write $\Gamma_{2n,7} = -2r_0^\pi \bar{\Gamma}_{2n,7}$ where $\bar{\Gamma}_{2n,7} \equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (\hat{r}_i - r_i)' r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i$. We further decompose

$\bar{\Gamma}_{2n,7}$ as $\bar{\Gamma}_{2n,7} = R_{12n} + R_{13n}$, where

$$\begin{aligned} R_{12n} &\equiv n^{-1}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n (\hat{r}_i - r_i)' r_{1i} \pi' [\hat{r}(Y_i; V_j) - r(Y_i; V_j)] a_i, \\ R_{13n} &\equiv n^{-1}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n (\hat{r}_i - r_i)' r_{1i}' \pi [r(Y_i; V_j) - r(Y_i)] a_i. \end{aligned}$$

Following the analysis of R_{10} and R_{11n} , we can readily show that $R_{sn} = o_P(1)$ for $s = 12, 13$. It follows that $\Gamma_{2n,7} = o_P(1)$.

(d) Write $\Gamma_{2n,8} = -2b^{\frac{d}{2}+2} (\hat{r}_0^\pi - r_0^\pi) \bar{\Gamma}_{2n,8}$ where $\bar{\Gamma}_{2n,8} \equiv \sum_{i=1}^n r_i' (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi a_i$. Then $\bar{\Gamma}_{2n,8} = R_{14n} + R_{15n}$, where

$$\begin{aligned} R_{14n} &\equiv \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i' [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)] r_{2i}^\pi a_i, \\ R_{15n} &\equiv \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i' [r(Y_j; X_i, Z_j) - r_1(X_i)] r_{2i}^\pi a_i. \end{aligned}$$

By straightforward moment calculations, we can show that $R_{15n} = O_P(n^{1/2})$. By Lemma 8.3(a), we can show that $R_{14n} = \bar{R}_{14n} + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})$, where

$$\begin{aligned} \bar{R}_{14n} &\equiv \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i' \left[b^{-1} e_1 \bar{\mathbf{S}}_b(X_i, Z_j)^{-1} \mathbf{V}_b(Y_j; X_i, Z_j) \bar{f}_{ij}^{-1} - \bar{f}_{ij}^{-2} \bar{D}_{xij} e_2' \bar{\mathbf{S}}_c(X_i, Z_j)^{-1} \mathbf{V}_c^{(L)}(Y_j; X_i, Z_j) \right] \mathbf{1}_j r_{2i}^\pi a_i \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n r_i' \zeta_k(Y_j; X_i, Z_j) r_{2i}^\pi a_i. \end{aligned}$$

Noting that $E(\bar{R}_{14n}^2) = O(nb^{-2} + (b^{-d_x-2} + b^{-2d_z-2} + c^{-d_x-1} + c^{-2d_z-2}) + n^{-1}(b^{-d-2-d_z} + c^{-d-2-d_z}))$, we have $\bar{R}_{14n} = O_P(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1})$. Consequently, $\bar{\Gamma}_{2n,8} = O_P(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1}) + o_P(n^{1/2}b^{-(\frac{d}{2}+1)})$, and by Lemma 8.3(b),

$$\begin{aligned} \Gamma_{2n,8} &= O_P(n^{-1/2}b^{\frac{d}{2}+1}) \left[O_P(n^{1/2}b^{-1} + b^{-d_z-1} + c^{-d_z-1}) + o_P(n^{1/2}b^{-(\frac{d}{2}+1)}) \right] \\ &= O_P(b^{\frac{d}{2}} + n^{-1/2}b^{\frac{d_x-d_z}{2}} + n^{-1/2}b^{\frac{d}{2}+2}c^{-d_z-1}) + o_P(1) = o_P(1). \end{aligned}$$

(e) By the Cauchy-Schwarz inequality and Lemma 8.5, $|\Gamma_{2n,9}| \leq 2(\Gamma_{2n,2}\Gamma_{2n,4})^{1/2} = o_P(1)$.

(f) We first decompose $\Gamma_{2n,10}$ as follows:

$$\begin{aligned}
\Gamma_{2n,10} &= 2n^{-2}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \{[\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)] + [r(Y_j; X_i, Z_j) - r_1(X_i)]\}' r_{1i} r_{2i}^\pi \\
&\quad \times \pi' ([\hat{r}(Y_i; X_k, Z_k) - r(Y_i; X_k, Z_k)] + [r(Y_i; X_k, Z_k) - r_1(Y_i)]) a_i, \\
&= 2n^{-2}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n [r(Y_j; X_i, Z_j) - r_1(X_i)]' r_{1i} r_{2i}^\pi \pi' [r(Y_i; X_k, Z_k) - r_1(Y_i)] a_i \\
&\quad + 2n^{-2}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)]' r_{1i} r_{2i}^\pi \pi' [\hat{r}(Y_i; X_k, Z_k) - r(Y_i; X_k, Z_k)] a_i \\
&\quad + 2n^{-2}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n [\hat{r}(Y_j; X_i, Z_j) - r(Y_j; X_i, Z_j)]' r_{1i} r_{2i}^\pi \pi' [r(Y_i; X_k, Z_k) - r_1(Y_i)] a_i \\
&\quad + 2n^{-2}b^{\frac{d}{2}+2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n [r(Y_j; X_i, Z_j) - r_1(X_i)]' r_{1i} r_{2i}^\pi \pi' [\hat{r}(Y_i; X_k, Z_k) - r(Y_i; X_k, Z_k)] a_i \\
&\equiv 2R_{16n} + 2R_{17n} + 2R_{18n} + 2R_{19n}, \text{ say.}
\end{aligned}$$

By moment calculations, $E(R_{16n}) = O(b^{\frac{d}{2}+2})$ and $E(R_{16n}^2) = O(b^{d+4})$, implying that $R_{16n} = O_P(b^{\frac{d}{2}+2}) = o_P(1)$. For R_{17n} , we can show that $R_{17n} = \bar{R}_{17n} + o_P(1)$, where

$$\begin{aligned}
\bar{R}_{17n} &= n^{-2}b^{\frac{d}{2}+2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n [b^{-1}e_1 \bar{\mathbf{S}}_b(X_{i_1}, Z_{i_2})^{-1} \mathbf{V}_b(Y_{i_2}; X_{i_1}, Z_{i_2}) f_{i_1}^{-1} \\
&\quad - D_{x_{i_1}} e_2' \bar{\mathbf{S}}_c(X_{i_1}, Z_{i_2})^{-1} \mathbf{V}_c^{(L)}(X_{i_1}, Z_{i_2}) f_{i_1}^{-2}]' \mathbf{1}_{i_2} r_{1i_1} r_{2i_1}^\pi \pi' \\
&\quad \times [b^{-1}e_1 \bar{\mathbf{S}}_b(X_{i_3}, Z_{i_3})^{-1} \mathbf{V}_b(Y_{i_1}; X_{i_3}, Z_{i_3}) f_{i_3}^{-1} - D_{x_{i_3}} e_2' \bar{\mathbf{S}}_c(X_{i_3}, Z_{i_3})^{-1} \mathbf{V}_c^{(L)}(X_{i_3}, Z_{i_3}) f_{i_3}^{-2}] \mathbf{1}_{i_1} a_{i_1} \\
&= n^{-4}b^{\frac{d}{2}+2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n \sum_{i_5=1}^n \zeta_{i_4}(Y_{i_2}; X_{i_1}, Z_{i_2})' r_{1i_1} r_{2i_1}^\pi \pi' \zeta_{i_5}(Y_{i_1}; X_{i_3}, Z_{i_3}) a_{i_1}
\end{aligned}$$

Noting that $E(\bar{R}_{17n}^2) = O(b^{d+4}b^{-4}) + n^{-1}b^{d+4}(b^{-d_x-4} + b^{-d_z-4} + c^{-d_x} + c^{-d_z}) = o(1)$, we have $\bar{R}_{17n} = o_P(1)$. Similarly, we can show that $R_{18n} = o_P(1)$ and $R_{19n} = o_P(1)$. Consequently, $\Gamma_{2n,10} = o_P(1)$. ■

Proof of Theorem 3.2

The proof follows closely from that of Theorems 3.1. By (8.3) and the proof of Theorem 3.1. Now $\hat{\Gamma} = n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{1n} + n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{4n} + n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{5n} + o_P(1)$. It is easy to show that $n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{1n} = n^{-1} \sum_{i=1}^n \|r_i r_i^\pi - r_{1i} r_{2i}^\pi\|^2 a_i = \mu_A + o_P(1)$ and $n^{-1}b^{-(\frac{d}{2}+2)}\Gamma_{sn} = o_P(1)$ under \mathbb{H}_A for $s = 4, 5$. In addition, under \mathbb{H}_A , we have $n^{-1}b^{-(\frac{d}{2}+2)}\hat{\mathbb{B}}_n = o_P(1)$ and $\hat{\sigma}_n^2 \xrightarrow{P} \sigma_A^2$. It follows that $n^{-1}b^{-(\frac{d}{2}+2)}T_n = n^{-1}b^{-(\frac{d}{2}+2)}[nb^{(\frac{d}{2}+2)}\hat{\Gamma} - \hat{\mathbb{B}}_n]/\sqrt{\hat{\sigma}_n^2} = \mu_A/\sigma_A + o_P(1)$ and the result follows. ■

Proof of Theorem 3.3

The proof follows closely from that of Theorem 3.1, now keeping the additional terms that do not vanish under $\mathbb{H}_A(\gamma_n)$ with $\gamma_n = n^{-1/2}b^{-\frac{d}{4}-1}$. Noting that $\hat{\mathbb{B}}_n = \mathbb{B}_n + o_P(1)$ and $\hat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$ under $\mathbb{H}_A(\gamma_n)$, it suffices to show that under $\mathbb{H}_A(\gamma_n)$, (i) $\Gamma_{1n} \xrightarrow{P} \mu_0$, (ii) $\Gamma_{4n} = o_P(1)$ and (iii) $\Gamma_{5n} = o_P(1)$, where Γ_{1n} , Γ_{4n} , and Γ_{5n} are defined after (8.3).

(i) holds under $\mathbb{H}_A(\gamma_n)$ because by the weak law of large numbers, we have

$$\Gamma_{1n} = b^{\frac{d}{2}+2} \sum_{i=1}^n \|r_i r_0^\pi - r_{1i} r_{2i}^\pi\|^2 a_i = n^{-1} \sum_{i=1}^n \|\delta_n(Y_i; X_i, Z_i)\|^2 a_i = \mu_0 + o_P(1).$$

For (ii), we decompose Γ_{4n} as $\Gamma_{4n} = \Gamma_{4n,1} + \Gamma_{4n,2} - \Gamma_{4n,3} - \Gamma_{4n,4}$, where

$$\begin{aligned} \Gamma_{4n,1} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_i r_0^\pi - r_{1i} r_{2i}^\pi)' (\hat{r}_i - r_i) r_0^\pi a_i, \\ \Gamma_{4n,2} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_i r_0^\pi - r_{1i} r_{2i}^\pi)' r_i (\hat{r}_0^\pi - r_0^\pi) a_i, \\ \Gamma_{4n,3} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_i r_0^\pi - r_{1i} r_{2i}^\pi)' (\hat{r}_{1i} - r_{1i}) r_{2i}^\pi a_i, \\ \Gamma_{4n,4} &\equiv b^{\frac{d}{2}+2} \sum_{i=1}^n (r_i r_0^\pi - r_{1i} r_{2i}^\pi)' r_{1i} (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i. \end{aligned}$$

It suffices to prove $\Gamma_{4n,s} = o_P(1)$ for $s = 1, 2, 3, 4$. We only prove $\Gamma_{4n,1} = o_P(1)$ as the other cases are similar. Let $\delta_{ni} \equiv \delta_n(Y_i; X_i, Z_i)$. Under $\mathbb{H}_A(\gamma_n)$ we apply Lemma 8.3(a) to obtain

$$\Gamma_{4n,1} = n^{-\frac{1}{2}} b^{\frac{d}{4}+1} r_0^\pi \sum_{i=1}^n \delta'_{ni} (\hat{r}_i - r_i) a_i = \bar{\Gamma}_{4n,1} + n^{\frac{1}{2}} b^{\frac{d}{4}+1} O_P(\nu_{bc}) = \bar{\Gamma}_{4n,1} + o_P(1),$$

where $\bar{\Gamma}_{4n,1} \equiv n^{-\frac{1}{2}} b^{\frac{d}{4}+1} r_0^\pi \sum_{i=1}^n \delta'_{ni} [b^{-1} e_1 \bar{\mathbf{S}}_b(V_i)^{-1} \mathbf{V}_b(Y_i; X_i) f_i^{-1} - D_{xi} e_2' \bar{\mathbf{S}}_c(V_i)^{-1} \mathbf{V}_c^{(L)}(V_i) f_i^{-2}] a_i$. Write $\bar{\Gamma}_{4n,1} = n^{-\frac{3}{2}} b^{\frac{d}{4}+1} r_0^\pi \sum_{i=1}^n \sum_{j=1}^n \delta'_{ni} \zeta_j(Y_i; V_i) a_i$. Then $E(\bar{\Gamma}_{4n,1}^2) = O((b^{\frac{d}{2}} + n^{-1} b^{\frac{d}{2}+2} (b^{-d-2} + c^{-d-1}) + n^{-2} b^{\frac{d}{2}+2} (b^{-2d-2} + c^{-2d-1})) = o(1)$, implying that $\bar{\Gamma}_{4n,1} = o_P(1)$. It follows that $\Gamma_{4n,1} = o_P(1)$.

We now show (iii). Decompose $\Gamma_{5n} = \Gamma_{5n,1} - \Gamma_{5n,2}$, where $\Gamma_{5n,1} = (\hat{r}_0^\pi - r_0^\pi) \Gamma_{4n,1} / r_0^\pi$, and $\Gamma_{5n,2} = \gamma_n b^{\frac{d}{2}+2} \sum_{i=1}^n \delta'_{ni} (\hat{r}_{1i} - r_{1i}) (\hat{r}_{2i}^\pi - r_{2i}^\pi) a_i$. In view of Lemma 8.3(b) and the study of $\Gamma_{4n,1}$, $\Gamma_{5n,1} = O_P(\nu_{bc} + n^{-1/2} b^{-1}) o_P(1) = o_P(1)$. Analogously to the proof of Lemma 8.6(f), we can show that $\Gamma_{5n,2} = o_P(\gamma_n)$. Thus $\Gamma_{5n} = o_P(1)$.

Consequently, $P(T_n \geq z | \mathbb{H}_A(n^{-1/2} b^{-\frac{d}{4}-1})) \rightarrow 1 - \Phi(z - \mu_0 / \sigma_0)$. This concludes the proof of the theorem.

■

Proof of Proposition 5.1

We first prove the ‘‘if’’ part. By the definition of the hazard function, for any values (y, x, ς) on the support of (Y, X, ξ) ,

$$h(y, x, \varsigma) = \frac{f(y|x, \varsigma)}{1 - F(y|x, \varsigma)},$$

where $f(y|x, \varsigma)$ and $F(y|x, \varsigma)$ are conditional PDF and CDF of Y given $(X, \xi) = (x, \varsigma)$, respectively. Then,

$$\begin{aligned} F(y|x, \varsigma) &= P \left[G \left(H_1(X) + \ln \left(\frac{-\ln(1-\varepsilon)}{\xi} \right) \right) \leq y \mid X = x, \xi = \varsigma \right] \\ &= P \left[\frac{-\ln(1-\varepsilon)}{\xi} \leq \exp[G^{-1}(y) - H_1(X)] \mid X = x, \xi = \varsigma \right] \\ &= P[\varepsilon \leq 1 - \exp\{-\varsigma \exp[G^{-1}(y) - H_1(x)]\}] \\ &= 1 - \exp\{-\varsigma \exp[G^{-1}(y) - H_1(x)]\}. \end{aligned}$$

Thus,

$$f(y|x, \varsigma) = \varsigma \exp \left\{ -\varsigma \exp [G^{-1}(y) - H_1(x)] \right\} \exp [G^{-1}(y) - H_1(x)] \frac{d[G^{-1}(y)]}{dy},$$

and

$$\begin{aligned} h(y, x, \varsigma) &= \frac{f(y|x, \varsigma)}{1 - F(y|x, \varsigma)} = \left\{ \frac{\exp [G^{-1}(y)]}{g(G^{-1}(y))} \right\} \exp [-H_1(x)] \varsigma \\ &= \lambda(y) \cdot \theta(x) \cdot \varsigma, \end{aligned}$$

where $\lambda(y) = \exp [G^{-1}(y)] / g(G^{-1}(y))$, $g(s) = dG(s) / ds$, and $\theta(x) = \exp [-H_1(x)]$. This holds for all (y, x, ς) on the support of (Y, X, ξ) , thus the “if” part is proved.

Next, we prove the “only if” part. Define the integrated hazard function $H(Y, X, \xi) = \int_0^Y h(y, X, \xi) dy$. Then

$$H(Y, X, \xi) = \int_0^Y h(y, X, \xi) dy = \int_0^Y \lambda(y) dy \cdot \theta(X) \cdot \xi = \Lambda(Y) \cdot \theta(X) \cdot \xi,$$

where $\Lambda(Y) = \int_0^Y \lambda(y) dy$. Let $F(Y | X, \xi)$ be the conditional CDF of Y given X and ξ . For any distribution function F , the integrated hazard function is related to its distribution function by

$$H(Y, X, \xi) = -\ln(1 - F(Y | X, \xi)).$$

Therefore

$$\Lambda(Y) \theta(X) \xi = -\ln(1 - F(Y | X, \xi)).$$

Define the random variable $\varepsilon = F(Y | X, \xi)$. By construction ε is uniformly distributed on $[0, 1]$ and $\varepsilon \perp (X, \xi)$ and

$$\Lambda(Y) \theta(X) \xi = -\ln(1 - \varepsilon).$$

Thus

$$\ln[\Lambda(Y)] = \ln \left[\frac{-\ln(1 - \varepsilon)}{\xi} \right] + \ln \left[\frac{1}{\theta(X)} \right].$$

That is,

$$Y = G[H_1(X) + U],$$

where $G(\cdot)$ is the inverse function of $\ln[\Lambda(\cdot)]$, $H_1(X) = -\ln[\theta(X)]$ and $U = \ln \left[\frac{-\ln(1 - \varepsilon)}{\xi} \right]$. ■

9 Appendix II: Bootstrap methods

In this appendix, we propose a residual-based bootstrap and a weighted bootstrap, and discuss the relative advantages and disadvantages associated with bootstrapping in the context of our tests.

9.1 Residual-based bootstrap

Here we describe a residual-based bootstrap method.

Step 1. Estimate the restricted model under \mathbb{H}_{10} and obtain the restricted residuals $\hat{\varepsilon}_i$. Let $M(y) = G^{-1}(y)$. Given \mathbb{H}_{10} , we can use the Chiappori et al. (2011, p.14) two-step estimator to estimate $M(\cdot)$ and $H_1(\cdot)$ as follows: (i) Estimate $M(y)$ using:

$$\hat{M}(y) = \int \int w(x, z) \frac{\hat{S}(y, x, z)}{\hat{E}[\hat{S}(Y, x, z)]} dx dz,$$

where $w(x, z)$ is a weighting function satisfying $\int \int w(x, z) dx dz = 1$,

$$\hat{S}(y, x, z) = \int_0^y \frac{\hat{D}_x F_b(y | x, z)}{\hat{f}_c(y | x, z)} dy, \text{ and } \hat{E}[\hat{S}(Y, x, z)] = \frac{1}{n} \sum_{i=1}^n \hat{S}(Y_i, x, z);$$

(ii) Run a series regression of $\hat{M}(Y_i)$ on X_i and Z_i to obtain an estimator of $H_1(\cdot)$, say $\hat{H}_1(\cdot)$. Then the restricted residuals are estimated as

$$\hat{\varepsilon}_i = \hat{M}(Y_i) - \hat{H}_1(X_i), \quad i = 1, 2, \dots, n.$$

Step 2. Obtain the bootstrap error $\varepsilon_i^* = \hat{\varepsilon}_i \eta_i$ for $i = 1, 2, \dots, n$, where η_i 's are IID $N(0, 1)$. Then the bootstrap sample is $\{Y_i^*, X_i\}_{i=1}^n$ where

$$Y_i^* = \hat{M}^{-1} \left(\hat{H}_1(X_i) + \varepsilon_i^* \right), \quad i = 1, 2, \dots, n. \quad (9.1)$$

There are some difficulties associated with this approach. First, estimating \hat{H}_1 and \hat{G}^{-1} is complicated, as it involves several nonparametric objects and integrals. Second, we need to evaluate the inverse function in (9.1) at different points for each bootstrap sample. As a result, this bootstrap is time-consuming. The estimate $\hat{E}[\hat{S}(Y, x, z)]$ is based on marginal integration, so to justify the asymptotic validity of the above bootstrap procedure, we would want to show uniform consistency of $\hat{S}(Y_i, x, z)$ for all (Y_i, x, z) . This may be difficult unless one assumes that $f(\cdot | x, z)$ is bounded away from zero for all (x, z) on the support of $w(\cdot, \cdot)$, which is an additional restriction that our test does not otherwise require.

9.2 Weighted bootstrap

An alternative to the residual based bootstrap is a weighted bootstrap procedure similar to those in Lewbel (1995) and Hansen (1996). Specifically, in the proof of Theorem 3.1 in the appendix, we demonstrate that

$$T_n = 2n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{1 \leq i < j \leq n} \frac{\varphi(W_i, W_j)}{\sigma_0} + o_p(1),$$

where the dominant term on the right hand side is a second-order degenerate U -statistic and asymptotically normally distributed. To approximate the null distribution of T_n , we simulate a large number B of $T_n^{*(k)}$ ($k = 1, \dots, B$):

$$T_n^{*(k)} = 2n^{-1} b^{\frac{d}{2}+2} (r_0^\pi)^2 \sum_{1 \leq i < j \leq n} \frac{\hat{\varphi}(W_i, W_j)}{\sqrt{\hat{\sigma}_n^2}} \xi_i^{(k)} \zeta_j^{(k)},$$

where $\{\xi_i^{(k)}\}_{i=1}^n$ and $\{\zeta_j^{(k)}\}_{j=1}^n$ are IID $N(0, 1)$ draws that are mutually independent of each other. Then the p -value is calculated as

$$p = B^{-1} \sum_{k=1}^B \mathbf{1} \left\{ T_n < \hat{T}_n^{*(k)} \right\}.$$

This bootstrap procedure is easy to implement, particularly in comparison to the residual based bootstrap, however, its finite performance crucially depends on the normal approximation of T_n . In simulations, we found that this procedure tended to over-reject the null, presumably due to finite sample departures from normality.

REFERENCES

- Abbring, J. H., Chiappori, P., Zavadil, T., 2008. Better safe than sorry? Ex ante and ex post moral hazard in dynamic insurance data. Working Paper, VU University Amsterdam and Columbia University.
- Altonji, J., Matzkin, R., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73, 1053-1102.
- Blundell, R., Powell, J., 2003. Endogeneity in nonparametric and semiparametric regression models, In *Advances in Economics and Econometrics*, eds. M. Dewatripont, L. Hansen, and S. Turnovsky, 294-311, Cambridge: Cambridge University Press.
- Blundell, R., Powell, J., 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71, 581-913.
- Cameron, C., Pravin T., 2005. *Microeconometrics: methods and applications*. Cambridge University Press, New York.
- Chesher, A., 2002. Semiparametric identification in duration models. Centre for Microdata Methods and Practice Working Paper.
- Chiappori, P-A., Komunjer, I., 2011. Correct specification and identification of nonparametric transformation models. Working paper, Dept. of Economics, Columbia University.
- Chiappori, P-A., Komunjer, I., Kristensen, D., 2011. Nonparametric identification and estimation of transformation models. Working paper, Dept. of Economics, Columbia University.
- Cox, D. R., 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34, 187-220.
- Dawid, A. P., 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* 41, 1-31.
- Ekeland, I., Heckman, J. J., Nesheim, L., 2004. Identification and estimation of hedonic models. *Journal of Political Economy* 112, 60-109.
- Engle, R. F., 2000. The econometrics of ultra-high-frequency data. *Econometrica* 68, 1-22.
- Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* 83, 189-206.
- Fernandes, M., Grammig, J., 2005. Nonparametric specification tests for conditional duration models. *Journal of Econometrics* 127, 35-68.
- Gozalo, P., Linton, O., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104, 1-48.
- Greene, W., 2011. *Econometric Analysis*. 7th edition. NJ: Prentice Hall.
- 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., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413-430.
- Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726-748.
- Heckman, J. J., Singer, B., (1984), A method for minimizing the impact of distributional assumptions in econometric Models for duration data. *Econometrica*, 52, 271-320.

- Heckman, J. J., Matzkin, R., Nesheim, L., 2005. Estimation and simulation of hedonic models. In T. Kehoe, T. Srinivasan, and J. Whalley (eds), *Frontiers in Applied General Equilibrium*, pp. 277-340. Cambridge University Press, Cambridge.
- Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926-1947.
- Hoderlein, S., Mammen, E., 2007. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75, 1513-1518.
- Hoderlein, S., Su, L., White, H., 2011. Specification testing for nonparametric structural models with monotonicity in unobservables. Working paper, Dept. of Economics, Boston College.
- Horowitz, J. L., 1996. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103-137.
- Horowitz, J. L., 2001. Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica* 69, 499-513.
- Horowitz, J. L., 2013. Nonparametric additive models. Forthcoming in A. Ullah, J. Racine, and L. Su (eds), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press, Oxford.
- Horowitz, J. L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. *Annals of Statistics* 36, 2412-2443.
- Horowitz, J. L., Mammen, E., 2007. Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Annals of Statistics* 35, 2589-2619.
- Horowitz, J. L., Mammen, E., 2011. Oracle-efficient estimation of an additive model with an unknown link function. *Econometric Theory* 27, 582-608.
- Ichimura, H., Lee, S., 2011. Identification and estimation of a nonparametric transformation model. Working paper.
- Imbens, G., Newey, W., 2009. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481-1512.
- Jacho-Chávez, D., Lewbel, A., Linton, O., 2010. Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics* 156, 392-407.
- Kennan, J. E., 1985. The duration of contract strikes in US manufacturing. *Journal of Econometrics* 28, 5-28.
- Keifer, N. M., 1988. Economic duration data and hazard functions. *Journal of Economic Literature* 26, 646-679.
- Lancaster, T., 1979. Econometric models for the duration of unemployment. *Econometrica* 47, 939-956.
- Lewbel, A., 1995. Consistent nonparametric tests with an application to Slutsky symmetry. *Journal of Econometrics* 67, 379-401.
- Li, Q., Lu, X., Ullah, A., 2003. Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics* 15, 607-627.
- Li, Q., Racine, J., 2003. Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, 86, 266-292.
- Lillard, L., Panis, C., 2003. aML Multilevel Multiprocess Statistical Software, Version 2.0. EconWare, Los Angeles, California.
- Lu, X., White, H., 2013. Testing for separability in structural equations. *Journal of Econometrics*, forthcoming.
- Mata, J., Portugal, P., 1994. Life duration of new firms. *Journal of Industrial Economics*, 42, 227-245.
- Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

- 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. 6, pp.5307-5368. Elsevier, B.V.
- Politis, D., Romano, J., Wolf, M., 1999. *Subsampling*. New York: Springer-Verlag.
- Ridder, G., 1990. The nonparametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167-181.
- Su, L., Tu, Y., Ullah, A., 2013. Testing additive separability of error term in nonparametric structural models. *Econometric Reviews*, forthcoming.
- Van den Berg, G., 2001. Duration Models: Specification, Identification and Multiple Durations. In J.J. Heckman and E.E. Leamer (eds), *Handbook of Econometrics*, Vol. 5, pp.3381-3460. Elsevier, B.V.
- White, H., Lu, X., 2011. Causal diagrams for treatment effect estimation with application to selection of efficient covariates. *Review of Economics and Statistics* 93, 1453-1459.