SMOOTH VARYING-COEFFICIENT ESTIMATION AND INFERENCE FOR QUALITATIVE AND QUANTITATIVE DATA

Author(s): Qi Li and Jeffrey S. Racine
Source: *Econometric Theory*, Vol. 26, No. 6 (December 2010), pp. 1607-1637
Published by: Cambridge University Press
Stable URL: https://www.jstor.org/stable/40930669
Accessed: 19-12-2018 20:28 UTC

Econometric Theory, **26**, 2010, 1607–1637.
doi:10.1017/S0266466609990739

SMOOTH VARYING-COEFFICIENT ESTIMATION AND INFERENCE FOR QUALITATIVE AND QUANTITATIVE DATA

Qi Li
Texas A&M University

Jeffrey S. Racine
McMaster University

We propose a semiparametric varying-coefficient estimator that admits both qualitative and quantitative covariates along with a test for correct specification of parametric varying-coefficient models. The proposed estimator is exceedingly flexible and has a wide range of potential applications including hierarchical (mixed) settings, small area estimation, etc. A data-driven cross-validatory bandwidth selection method is proposed that can handle both the qualitative and quantitative covariates and that can also handle the presence of potentially irrelevant covariates, each of which can result in finite-sample efficiency gains relative to the conventional frequency (sample-splitting) estimator that is often found in such settings. Theoretical underpinnings including rates of convergence and asymptotic normality are provided. Monte Carlo simulations are undertaken to assess the proposed estimator's finite-sample performance relative to the conventional semiparametric frequency estimator and to assess the finite-sample performance of the proposed test for correct parametric specification.

1. INTRODUCTION

The seminal work of Aitchison and Aitken (1976) has spawned a rich literature on the kernel smoothing of categorical data and has also spawned numerous advances in the kernel smoothing of data sets comprised of both qualitative and quantitative data. One area in which this approach could be particularly helpful is when confronted with data in groups in which potential sparsity can adversely affect one's analysis but that can be alleviated by "borrowing" information from,

We thank a co-editor and two referees for their valuable comments that led to improvements in the paper. Li's research is partially supported by the Private Enterprise Research Center, Texas A&M University, and the National Science Foundation of China (project 70773005). Racine gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC: www.nserc.ca), the Social Sciences and Humanities Research Council of Canada (SSHRC: www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca). Racine thanks Tristen Hayfield for his exemplary research assistance. Address correspondence to Jeffrey S. Racine, Department of Economics, McMaster University, Hamilton, ON L8S 4M4, Canada; e-mail: racinej@mcmaster.ca.

 0266-4666/10 $15.00

say, neighboring cells in a prescribed manner. By way of example, one might encounter students grouped in classes grouped in schools or crops grouped in plots of land grouped in farms, etc. These situations are frequently encountered in mixed model settings such as when conducting small area estimation, estimating multilevel (hierarchical) models, and the like. The estimator we propose is ideally suited to these settings.

The conventional parametric approach to modeling such problems suffers from a number of drawbacks, and this has given rise to a growing literature on semiparametric and nonparametric approaches. A variety of flexible methods have been proposed for estimating these models including semiparametric and nonparametric smoothing spline approaches (Zhang, Lin, Raz, and Sowers, 1998; Gu and Ma, 2005) and semiparametric kernel-based approaches (Zeger and Diggle, 1994), and a recent *Journal of Multivariate Analysis* (2004) special issue was devoted entirely to such advances. Cai, Fan, and Yao (2000) and Fan, Yao, and Cai (2003) have further extended varying coefficient models to dependent data settings, and Cai, Das, Xiong, and Wu (2006) allowed for endogenous regressors in a varying coefficient model and proposed a semiparametric instrumental variable estimation method to estimate smooth coefficient functions. However, when confronted with the mix of qualitative and quantitative covariates that we often encounter in applied settings, existing semiparametric and nonparametric approaches rely on sample splitting to handle the presence of qualitative covariates, and this can lead to substantial efficiency losses in finite-sample settings. Though existing semiparametric and nonparametric approaches provide the user with much needed flexibility, the methods proposed in this paper possess a number of appealing features not shared by their peers that ought to be quite attractive to practitioners. Most noteworthy is the fact that the proposed method does not rely upon sample splitting because it smooths the qualitative covariates in the manner described in Section 2.

We proceed as follows. In Sections 2 and 3, we provide theoretical underpinnings for the proposed estimator including rates of convergence and asymptotic normality along with a data-driven cross-validatory method of bandwidth selection. Section 4 presents the proposed test for correct specification of parametric varying-coefficient models. In Section 5, Monte Carlo simulations are undertaken to assess the finite-sample performance of the proposed estimator and to assess the finite-sample properties of the proposed test for correct parametric specification, and Section 6 presents some concluding remarks. Proofs of the main theorems are relegated to the Appendixes.

2. KERNEL ESTIMATION OF VARYING-COEFFICIENT MODELS

Consider a varying-coefficient regression model given by

$$Y_i = X_i'\beta(\bar{Z}_i) + u_i, \qquad i = 1, \ldots, n, \tag{1}$$

where X_i is a p-dimensional vector of regressors, $\bar{Z}_i = (\bar{Z}_i^c, \bar{Z}_i^d)$, $\bar{Z}_i^c$ is a continuous covariate of dimension q_1, $\bar{Z}_i^d$ is a categorical covariate of dimension r_1, and n is the sample size. The functional form of $\beta(\cdot)$ is not specified, and u_i is an error term satisfying $\mathrm{E}(u_i|X_i, \bar{Z}_i) = 0$.

We allow for the possibility that the nonparametric varying-coefficient function $\beta(\cdot)$ is over-specified. That is, instead of using only the relevant covariates $\bar{Z}_i$ in (1), one may use a possibly larger set of covariates $Z_i = (Z_i^c, Z_i^d)$, where Z_i^c is a q-dimensional continuous covariate and Z_i^d is an r-dimensional categorical covariate ($q \geq q_1$ and $r \geq r_1$). Therefore, in practice one estimates the following model:

$$Y_i = X_i'\beta(Z_i) + u_i, \qquad i = 1, \dots, n. \tag{2}$$

Of course, we maintain the assumption that (1) is the correctly specified model. Hence, (2) is an overspecified model if $q > q_1$ and/or $r > r_1$.

We use Z_{is}^d to denote the sth component of Z_i^d, and we assume that Z_{is}^d takes c_s unique values ($c_s \geq 2$ is a finite positive integer). We use $\mathcal{S}^d$ to denote the support of Z^d. We allow for one or more of the nonparametric covariates to be "irrelevant" (the specific definition of "irrelevant covariate" will be provided later). Without loss of generality, we assume that the first q_1 $(1 \leq q_1 \leq q)$ components of Z^c and the first r_1 $(0 \leq r_1 \leq r)$ components of Z^d are "relevant" covariates in the sense defined subsequently. Note that we assume there exists at least one relevant continuous covariate ($q_1 \geq 1$). It can be shown that when all of the continuous covariates are irrelevant ($q_1 = 0$), the asymptotic distribution of the cross-validated smoothing parameters will be quite different from the case treated in this paper. Results for the case in which $q_1 = 0$ cannot be obtained as a special case of those derived subsequently; hence we must treat this case separately and intend to do so in future work.

Let $\bar{Z}$ consist of the first q_1 relevant components of Z^c and the first r_1 relevant components of Z^d and let $\tilde{Z} = Z \setminus \{\bar{Z}\}$ denote the remaining irrelevant components of Z.

Following Hall, Li, and Racine (2007) we assume that

$$(Y, X, \bar{Z}) \quad \text{is independent of } \tilde{Z}. \tag{3}$$

Clearly, (3) implies that, for all measurable functions $g(Y, X)$,

$$\mathrm{E}\big(g(Y, X)|Z\big) = \mathrm{E}\big(g(Y, X)|\bar{Z}\big). \tag{4}$$

Obviously (3) is a stronger assumption than (4). A weaker condition would be to require that

$$\text{Conditional on } (X, \bar{Z}), \quad \text{the variables } \tilde{Z} \text{ and } Y \text{ are independent of each other.} \tag{5}$$

However, using (5) would cause some formidable technical difficulties in the proofs that follow that we are unable to handle at this time. Therefore, in this paper we will only consider the unconditional independence assumption (3). Nevertheless, we have also investigated the case of conditional independence as defined by (5) via simulation, and the simulation results reveal that cross-validation can smooth out irrelevant covariates under either unconditional independence (3) or conditional independence (5).[1]

We now discuss the kernel smoothing of discrete covariates. For an unordered discrete covariate, we suggest using a variant of the Aitchison and Aitken (1976) kernel function defined as

$$l(Z_{is}^d, z_s^d, \lambda_s) = \begin{cases} 1, & \text{when } Z_{is}^d = z_s^d, \\ \lambda_s, & \text{otherwise.} \end{cases} \tag{6}$$

For an ordered discrete covariate, we use the following kernel function:

$$l(Z_{is}^d, z_s^d, \lambda_s) = \begin{cases} 1, & \text{when } Z_{is}^d = z_s^d \\ \lambda_s^{|Z_{is}^d - z_s^d|}, & \text{when } Z_{is}^d \neq z_s^d \end{cases}, \tag{7}$$

where $\mathbf{1}(A)$ denotes the usual indicator function, which assumes the value one if A holds true and zero otherwise. Note that for both unordered and ordered discrete covariates, $\lambda_s = 0$ leads to an indicator function whereas $\lambda_s = 1$ leads to a uniform weight function. Therefore, the range of λ_s is $[0, 1]$ for all $s = 1, \dots, r$. Using (6) and (7), we construct a product kernel for the discrete covariates,

$$L(Z_i^d, z^d, \lambda) = \prod_{s=1}^{r} l(Z_{is}^d, z_s^d, \lambda_s). \tag{8}$$

Observe that the kernel weight function we use here does not add up to one when $\lambda_s \neq 0$. However, this does not affect the semiparametric estimator $\hat{\beta}(z)$ defined in (10) in this section as the kernel function appears in both the numerator and the denominator of (10) and thus the kernel function can be multiplied by any nonzero constant, leaving the definition of $\hat{\beta}(z)$ given subsequently.

For the continuous covariates $Z_i^c = (Z_{i1}^c, \dots, Z_{iq}^c)$ we use the product kernel given by

$$W_h\left(\frac{Z_j^c - Z_i^c}{h}\right) = \prod_{s=1}^{q} \frac{1}{h_s} w\left(\frac{Z_{js}^c - Z_{is}^c}{h_s}\right),$$

where $w(\cdot)$ is a symmetric univariate density function and where $0 < h_s < \infty$ is the smoothing parameter for z_s^c, $s = 1, \dots, q$.

The kernel function for the mixed covariate case $z = (z^c, z^d)$ is simply the product of $W_h(\cdot)$ and $L(\cdot)$, i.e.,

$$K_{\gamma,ij} \equiv K_\gamma(Z_i, Z_j) \stackrel{\text{def}}{=} W_h\left(\frac{Z_j^c - Z_i^c}{h}\right) L(Z_j^d, Z_i^d, \lambda), \tag{9}$$

where $\gamma = (h, \lambda) = (h_1, \dots, h_q, \lambda_1, \dots, \lambda_r)$.

We use $\mathcal{S}$ to denote the support of Z_i.[2] For $z \in \mathcal{S}$, from (1) it is fairly straightforward to show that

$$\beta(z) = \left[\mathrm{E}(X_i X_i' | Z_i = z)\right]^{-1} \mathrm{E}(X_i Y_i | Z_i = z).$$

Therefore, we estimate $\beta(z)$ by the local constant kernel method, which is given by

$$\hat{\beta}(z) = \left[n^{-1} \sum_{i=1}^{n} X_i X_i' K_\gamma (Z_i, z)\right]^{-1} \left[n^{-1} \sum_{i=1}^{n} X_i Y_i K_\gamma (Z_i, z)\right]. \tag{10}$$

When $\lambda_s = 0$ for all $s = 1, \ldots, r$, our estimator collapses to the conventional approach that uses a frequency estimator to deal with the categorical covariates (i.e., one resorts to sample splitting), whereas if $\lambda_s = 1$ for some s, then $\hat{\beta}(z)$ becomes unrelated to z_s^d because $l(Z_{is}^d, z_s^d, \lambda_s = 1) \equiv 1$ (a constant that is unrelated to z_s^d). That is, when $\lambda_s = 1$, z_s^d is smoothed out from the regression model (it is deemed to be an "irrelevant" qualitative covariate). Similarly, if h_s is sufficiently large, say, $h_s = \infty$, then $w((Z_{is}^c - z_s^c)/h_s) \sim w(0)$, a constant that is unrelated to z_s^c. Thus, z_s^c is completely smoothed out and $\hat{\beta}(z)$ becomes unrelated to z_s^c when h_s is sufficiently large.

We choose $\gamma = (h, \lambda) = (h_1, \ldots, h_q, \lambda_1, \ldots, \lambda_r)$ to minimize[3]

$$CV(\gamma) = \sum_{i=1}^{n} \left[Y_i - X_i' \hat{\beta}_{-i}(Z_i)\right]^2 M(Z_i), \tag{11}$$

where $0 \leq M(\cdot) \leq 1$ is a weight function that serves to avoid difficulties caused by dividing by zero, or by the slower convergence rate that arises when Z_i lies near the boundary of its support, whereas

$$\hat{\beta}_{-i}(Z_i) = \left[n^{-1} \sum_{j \neq i}^{n} X_j X_j' K_\gamma (Z_j, Z_i)\right]^{-1} \left[n^{-1} \sum_{j \neq i}^{n} X_j Y_j K_\gamma (Z_j, Z_i)\right] \tag{12}$$

is the leave-one-out kernel estimator of $\beta(Z_i)$.

3. ASYMPTOTIC ANALYSIS ALLOWING FOR IRRELEVANT COVARIATES

We will use $f(x, \bar{z})$, $\bar{f}(\bar{z})$, and $\tilde{f}(\tilde{z})$ to denote the joint density of $(X, \bar{Z})$ and the marginal densities of $\bar{Z}$ and $\tilde{Z}$, respectively. Define $\sigma^2(x, \bar{z}) = \mathrm{E}(u_i^2 | X_i = x, \bar{Z}_i = \bar{z})$. We make the following assumptions.

The data are independent and identically distributed (i.i.d.), and u_i has finite moments of any order;

$\beta(\bar{z})$, $f(x, \bar{z})$, and $\sigma^2(x, \bar{z})$ have two continuous derivatives (with respect to $\bar{z}^c$);
$M(\cdot)$ is continuous and nonnegative and has compact support;
f and $\tilde{f}$ are bounded away from zero for $z = (z^c, z^d) \in \mathcal{S} = \mathcal{S}^c \times \mathcal{S}^d$. **(13)**

We impose the following conditions on the bandwidth and kernel functions. Define

$$H = \left(\prod_{s=1}^{q_1} h_s\right) \prod_{s=q_1+1}^{q} \min(h_s, 1). \tag{14}$$

Letting $0 < \epsilon < 1/(q+4)$ and for some constant $c > 0$, we further assume that

$n^{\epsilon-1} \le H \le n^{-\epsilon}$; $\quad n^{-c} < h_s < n^c$ for all $s = 1, \ldots, q$;
the kernel function $w(\cdot)$ is a symmetric, compactly supported,
Hölder-continuous probability density; $\quad w(0) > w(\delta)$ for all $\delta \ne 0$. **(15)**

The preceding conditions basically require that each h_s does not converge to zero, or to infinity, too fast and that $nh_1 \ldots h_{q_1} \to \infty$ as $n \to \infty$.

We use $\mathcal{H}$ to denote the permissible set for $(h_1, \ldots, h_q)$ that satisfies (15). The range for $(\lambda_1, \ldots, \lambda_r)$ is $[0,1]^r$, and we use $\Gamma = \mathcal{H} \times [0,1]^r$ to denote the range for the smoothing parameter vector $\gamma \equiv (h_1, \ldots, h_q, \lambda_1, \ldots, \lambda_r)$.

We expect that, as $n \to \infty$, the smoothing parameters associated with the relevant covariates will converge to zero, whereas those associated with the irrelevant covariates will not. It would be convenient to further assume that $h_s \to 0$ for $s = 1, \ldots, q_1$ and that $\lambda_s \to 0$ for $s = 1, \ldots, r_1$. However, for practical reasons we choose not to assume that the relevant components are known a priori, but rather assume that assumption (17) given later in this section holds. We write $K_{\gamma,ij} = \bar{K}_{\bar{\gamma},ij}\tilde{K}_{\tilde{\gamma},ij}$, where $\bar{\gamma} = (h_1, \ldots, h_{q_1}, \lambda_1, \ldots, \lambda_{r_1})$ and $\tilde{\gamma} = (h_{q_1+1}, \ldots, h_q, \lambda_{r_1+1}, \ldots, \lambda_r)$ so that $\bar{K}$ and $\tilde{K}$ are the product kernel functions associated with the relevant and the irrelevant covariates, respectively.

We define

$$\begin{aligned}\eta_\beta(\bar{z}) &= \mathrm{E}[X_j X_j' K_{\gamma,ij} | z_i = z]^{-1} \mathrm{E}[X_j X_j' \beta(\bar{z}_j) K_{\gamma,ij} | z_i = z] \\ &= \mathrm{E}[X_j X_j' \bar{K}_{\bar{\gamma},ij} | \bar{z}_i = \bar{z}]^{-1} \mathrm{E}[X_j X_j' \beta(\bar{z}_j) \bar{K}_{\bar{\gamma},ij} | \bar{z}_i = \bar{z}],\end{aligned} \tag{16}$$

where the second equality comes from the fact that $\tilde{Z}$ is independent of $(X, \bar{Z})$. Therefore, the term related to $\tilde{z}$ cancels out because $\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij} | \tilde{z}_i = \tilde{z}]^{-1} \mathrm{E}[\tilde{K}_{\tilde{\gamma},ij} | \tilde{z}_i = \tilde{z}] = 1$; hence $\eta_\beta(\bar{z})$ does not depend on $\tilde{z}$ nor does it depend on $(h_{q_1+1}, \ldots, q, \lambda_{r_1+1}, \ldots, \lambda_r)$. We further assume that

$\int \left(\bar{\eta}_\beta(\bar{z}) - \beta(\bar{z})\right)' m(\bar{z}) \left(\bar{\eta}_\beta(\bar{z}) - \beta(\bar{z})\right) \bar{M}(\bar{z}) d\bar{z}$, a function of $h_1, \ldots, h_{q_1}$
and $\lambda_1, \ldots, \lambda_{r_1}$, vanishes if and only if all of the smoothing parameters vanish, **(17)**

where $m(\bar{z}) = \mathrm{E}(X_i X_i' | \bar{Z}_i = \bar{z}) \bar{f}(\bar{z})$, $\bar{M}(\bar{z}) = \int \bar{f}(\tilde{z}) M(z) d\tilde{z}$. In the proof of Lemma A.6 in Appendix A we show that (15) and (17) imply that as $n \to \infty$, $h_s \to 0$ for $s = 1, \dots, q_1$ and $\lambda_s \to 0$ for $s = 1, \dots, r_1$. Therefore, the smoothing parameters associated with the relevant covariates all vanish asymptotically.

Define an indicator function $\mathbf{1}_s(v^d, x^d) = \mathbf{1}(v_s^d \neq x_s^d) \prod_{t \neq s}^r \mathbf{1}(v_t^d = x_t^d)$. Note that $\mathbf{1}_s(v^d, x^d)$ is one if v^d and x^d differ only in their sth component and is zero otherwise. Let $\int d\bar{z} = \sum_{\bar{z}^d \in \bar{S}^d} \int d\bar{z}^c$, recall that $m(\bar{z}) = \mathrm{E}(X_i X_i' | \bar{Z}_i = \bar{z}) \bar{f}(\bar{z})$, and let m_s and m_{ss} denote the first and second derivatives of $m(\cdot)$ with respect to z_s^c, respectively, with β_s and β_{ss} similarly defined. Now define

$$B_{1s}(\bar{z}) = m(\bar{z})^{-1} \kappa_2 \left[m_s(\bar{z}^c, \bar{z}^d) \beta_s(\bar{z}^c, \bar{z}^d) + (1/2) m(\bar{z}^c, \bar{z}^d) \beta_{ss}(\bar{z}^c, \bar{z}^d) \right],$$

$$B_{2s}(\bar{z}) = m(\bar{z})^{-1} \sum_{\bar{v}^d \in \bar{S}_d} \mathbf{1}_s(\bar{z}^d, \bar{v}^d) m(\bar{z}^c, \bar{v}_d) \left[\beta(\bar{z}^c, \bar{v}^d) - \beta(\bar{z}^c, \bar{z}^d) \right], \tag{18}$$

where $\kappa_2 = \int w(v) v^2 dv$. In Appendix A we show that the leading term of $CV(\gamma)$ is

$$\int \left\{ \left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z}) \right]' m(\bar{z}) \left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z}) \right] \right\} \bar{M}(\bar{z}) d\bar{z}$$
$$+ \frac{\kappa^{q_1}}{n h_1 \dots h_{q_1}} \int \bar{f}(\bar{z}) \tilde{f}(\tilde{z}) \delta(\bar{z}) \tilde{R}(\tilde{z}) M(z) dz + o_p\left(\zeta_n^2 + \left(n h_1 \dots h_{q_1} \right)^{-1} \right), \tag{19}$$

where $\delta(\bar{Z}_i) = \mathrm{tr}\left\{ \mathrm{E}[X_i X_i' \sigma^2(X_i, \bar{Z}_i) | \bar{Z}_i] m(\bar{Z}_i)^{-1} \right\}$ (if $\mathrm{E}(u_i^2 | X_i, \bar{Z}_i) = \mathrm{E}(u_i^2 | \bar{Z}_i) = \sigma^2(\bar{Z}_i)$, then $\delta(\bar{z}) = \sigma^2(\bar{z}) \mathrm{tr}\{I_p\} / \bar{f}(\bar{z}) = p\sigma^2(\bar{z}) / \bar{f}(\bar{z})$), $\zeta_n = \sum_{s=1}^{q_1} h_s^2 + \sum_{s=1}^{r_1} \lambda_s$, $\kappa = \int w(v)^2 dv$, and $\kappa_2 = \int w(v) v^2 dv$ and where $\tilde{R}(\tilde{z}) = \tilde{R}(z, h_{q_1+1}, \dots, h_q, \lambda_{r_1+1}, \dots, \lambda_r)$ is given by

$$\tilde{R}(\tilde{z}) = \frac{\nu_2(\tilde{z})}{[\nu_1(\tilde{z})]^2}, \tag{20}$$

where for $l = 1, 2$, $\nu_l(\tilde{z}) = \mathrm{E}([\prod_{s=q_1+1}^q h_s^{-1} w\left((z_{is}^c - z_s^c)/h_s \right) \prod_{s=r_1+1}^r \lambda_s^{\mathbf{1}(z_{is}^d \neq z_s^d)}]^l)$.

In (19) the irrelevant covariate $\tilde{z}$ appears in $\tilde{R}(\tilde{z})$. By Hölder's inequality, $\tilde{R}(\tilde{z}) \geq 1$ for all choices of z, $h_{q_1+1}, \dots, h_q$, and $\lambda_{q_1+1}, \dots, \lambda_q$. Also, $\tilde{R}(\tilde{z}) \to 1$ as $h_s \to \infty$ $(q_1 + 1 \leq s \leq q)$ and $\lambda_s \to 1$ $(r_1 + 1 \leq s \leq r)$. Therefore, to minimize (19), one needs to select h_s $(s = q_1 + 1, \dots, q)$ and λ_s $(s = r_1 + 1, \dots, r)$ to minimize $\tilde{R}(\tilde{z})$. In fact, we show that the only smoothing parameter values for which $\tilde{R}(\tilde{z}, h_{q_1+1}, \dots, h_q, \lambda_{r_1+1}, \dots, \lambda_r) = 1$ are $h_s = \infty$ for $q_1 + 1 \leq s \leq q$, and $\lambda_s = 1$ for $r_1 + 1 \leq s \leq r$. To see this, let us define $\mathcal{V}_n = \prod_{s=q_1+1}^q w\left((z_s^c - Z_{is}^c)/h_s \right) \prod_{s=r_1+1}^r l(z_s^d, Z_{is}^d, \lambda_s)$, where $l(z_s^d, Z_{is}^d, \lambda_s) = \lambda_s^{\mathbf{1}(z_s^d \neq Z_{is}^d)}$ if z_s^d is an unordered covariate and $l(z_s^d, Z_{is}^d, \lambda_s) = \lambda_s^{|Z_{is}^d - z_s^d|}$ if z_s^d is an ordered covariate. If at least one

h_s is finite (for $q_1 + 1 \le s \le q$), or one $\lambda_s < 1$ (for $r_1 + 1 \le s \le r$), then by (15) ($w(0) > w(\delta)$ for all $\delta > 0$) we know that $\mathrm{var}(\mathcal{V}_n) = \mathrm{E}[\mathcal{V}_n^2] - [E(\mathcal{V}_n)]^2 > 0$ so that $R = \mathrm{E}[\mathcal{V}_n^2]/[E(\mathcal{V}_n)]^2 > 1$. Only when, in the definition of $\mathcal{V}_n$, all $h_s = \infty$ and all $\lambda_s = 1$ do we have $\mathcal{V}_n \equiv w(0)^{q-q_1}$ (a constant) and $\mathrm{var}(\mathcal{V}_n) = 0$ so that $\tilde{R}(\tilde{z}) = 1$ only in this case.

Therefore, to minimize (19), the smoothing parameters corresponding to the irrelevant covariates must all converge to their upper bounds so that $\tilde{R}(\tilde{z}) \to 1$ as $n \to \infty$ for all $\tilde{z} \in \tilde{\mathcal{S}}$ ($\tilde{\mathcal{S}}$ is the support of $\tilde{Z}$). Thus, irrelevant components are asymptotically smoothed out.

To analyze the behavior of smoothing parameters associated with the relevant covariates, we replace $\tilde{R}(\tilde{z})$ by 1 in (19); thus the second term on the right-hand side of (19) becomes

$$\frac{\kappa^{q_1}}{nh_1 \dots h_{q_1}} \int \bar{f}(\bar{z})\delta(\bar{z})\bar{M}(\bar{z})\,d\bar{z}, \tag{21}$$

where $\bar{M}(\bar{z}) = \int \tilde{f}(\tilde{z})M(z)\,d\tilde{z}$.

Next, defining $a_s = h_s n^{1/(q_1+4)}$ and $b_s = \lambda_s n^{2/(q_1+4)}$, then (19) (with (21) as its first term because $\tilde{R}(\tilde{z}) \to 1$) becomes $n^{-4/(q_1+4)}\bar{\chi}(a_1, \dots, a_{q_1}, b_1, \dots, b_{r_1})$, where

$$\begin{aligned}\bar{\chi}(a_1, \dots, b_{r_1}) = {} & \frac{\kappa^{q_1}}{a_1 \dots a_{q_1}} \int \bar{f}(\bar{z})\delta(\bar{z})\bar{M}(\bar{z})\,d\bar{z} \\ & + \int \left\{ \left[\sum_{s=1}^{q_1} a_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} b_s B_{2s}(\bar{z}) \right]' m(\bar{z}) \right. \\ & \left. \times \left[\sum_{s=1}^{q_1} a_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} b_s B_{2s}(\bar{z}) \right] \right\} \bar{M}(\bar{z})\,d\bar{z}. \end{aligned} \tag{22}$$

Let $a_1^0, \dots, a_{q_1}^0, b_1^0, \dots, b_{r_1}^0$ denote values of $a_1, \dots, a_{q_1}, b_1, \dots, b_{r_1}$ that minimize $\bar{\chi}$ subject to each of them being nonnegative. We require that

Each a_s^0 is positive and each b_s^0 nonnegative; all are finite and uniquely defined. **(23)**

The approach of Li and Zhou (2005) can be used to obtain primitive necessary and sufficient conditions that ensure that (23) holds true. The preceding analysis leads to the main result of this paper, which is stated in the following theorem.

THEOREM 3.1. *Assume that conditions (13), (15), (17), and (23) hold and let $\hat{h}_1, \dots, \hat{h}_q, \hat{\lambda}_1, \dots, \hat{\lambda}_r$ denote the smoothing parameters that minimize $CV(\gamma)$. Then*

$n^{1/(q_1+4)}\hat{h}_s \to a_s^0$ *in probability for* $1 \le s \le q_1$,

$$P\left(\hat{h}_s > C\right) \to 1 \quad \textit{for } q_1+1 \le s \le q \quad \textit{and for all } C > 0,$$

$$n^{2/(q_1+4)}\hat{\lambda}_s \to b_s^0 \quad \textit{in probability for } 1 \le s \le r_1,$$

$$\hat{\lambda}_s \to 1 \quad \textit{in probability for } r_1+1 \le s \le r, \tag{24}$$

$$\textit{and } \left\{n^{-4/(q_1+4)}CV(\hat{h}_1,\dots,\hat{h}_q,\hat{\lambda}_1,\dots,\hat{\lambda}_r) - n^{-1}\textstyle\sum_{i=1}^n u_i^2\right\} \to \inf \bar{\chi} \textit{ in probability.}$$

The proof of Theorem 3.1 is given in Appendix A. Theorem 3.1 states that the smoothing parameters associated with the irrelevant covariates all converge to their upper bounds, so that, asymptotically, all irrelevant covariates are smoothed out, whereas the smoothing parameters associated with the relevant covariates all converge to zero at a rate that is optimal for minimizing asymptotic mean square error (MSE) (i.e., without the presence of irrelevant covariates). From Theorem 3.1 one can easily obtain the following result.

THEOREM 3.2. *Under the same conditions given in Theorem 3.1, then*

$$\sqrt{n\hat{h}_1\dots\hat{h}_{q_1}}\left(\hat{\beta}(z) - \beta(\bar{z}) - \sum_{s=1}^{q_1}\hat{h}_s^2 B_{1s}(\bar{z}) - \sum_{s=1}^{r_1}\hat{\lambda}_s B_{2s}(\bar{z})\right)$$

$$\to N(0,\Omega(\bar{z})) \textit{ in distribution,}$$

where $\Omega(\bar{z}) = m(\bar{z})^{-1}\kappa^{q_1}\mathrm{E}[X_iX_i'\sigma^2(X_i,\bar{Z}_i)|\bar{Z}_i = \bar{z}]\bar{f}(\bar{z})m(\bar{z})^{-1}$, $m(\bar{z}) = \mathrm{E}[X_i X_i|\bar{Z}_i = \bar{z}]\bar{f}(\bar{z})$, $\kappa = \int w^2(v)dv$, $\sigma^2(X_i,\bar{Z}_i) = \mathrm{E}(u_i^2|X_i,\bar{Z}_i)$. *Moreover,* $\Omega(\bar{z})$ *can be consistently estimated by* $\hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}\hat{\mathcal{A}}^{-1}$, *where* $\hat{\mathcal{A}} = n^{-1}\sum_{i=1}^n X_iX_i'\bar{K}_{\hat{\gamma},iz}$, $\hat{\mathcal{B}} = n^{-1}\sum_{i=1}^n X_iX_i'\hat{u}_i^2\bar{K}_{\hat{\gamma},iz}^2$, $\bar{K}_{\hat{\gamma},iz} = \bar{K}_{\hat{\gamma}}(\bar{Z}_i,\bar{z})$, *and* $\hat{u}_i = Y_i - X_i'\hat{\beta}(Z_i)$.

A sketch of the proof of Theorem 3.2 is provided in Appendix A. We mention that when there are only continuous covariates ($r = 0$) and all of them are relevant covariates ($q = q_1$) the result of Theorem 3.2 has been obtained by Escanciano and Jacho-Chávez (2009)

4. TESTING FOR CORRECT VARYING-COEFFICIENT FUNCTIONAL SPECIFICATION

In this section we propose a test for correct parametric specification of $\beta(\cdot)$. Because all irrelevant covariates can be smoothed out asymptotically (see Theorem 3.1), in this section we shall restrict attention to the relevant covariates case. That is, we will omit the bar notation in $\bar{z}$ and simply write z to simplify the notation in what follows.

If one believes that $\beta(z)$ may be of a simple parametric functional form, say, $\beta(z) = \beta_0$, where β_0 is a $p \times 1$ vector of parameter constants, then naturally one ought to test this hypothesis. Furthermore, if the hypothesis is believed to be true then the semiparametric varying-coefficient model becomes a fully parametric regression model, and one should therefore estimate the parametric model.

Let $\beta_0(z,\alpha_0)$ denote a parametric varying-coefficient model of interest, where $\beta_0(\cdot,\cdot_0)$ has a known parametric functional form and where $\alpha_0 \in \mathbb{R}^k$ is a finite k-dimensional unknown parameter vector. Formally, we state the null hypothesis as

$$H_0: \quad \Pr[\beta(Z) = \beta_0(Z,\alpha_0)] = 1 \quad \text{for some } \alpha_0 \in \mathcal{B}, \tag{25}$$

where $\mathcal{B}$ is a compact subset in $\mathbb{R}^k$.

There is a rich literature on consistent model specification testing, and we direct the interested reader to Bierens (1982), Bierens (1990), Härdle and Mammen (1993), Bierens and Ploberger (1997), and Li and Racine (2007, Chs. 12 and 13) for more detailed discussion. Subsequently we will construct an easy to implement testing procedure following the approach suggested by Li, Huang, Li, and Fu (2002).

Let $\hat{\alpha}$ denote a $\sqrt{n}$-consistent estimator of α^*, where $\alpha^* = \alpha_0$ under H_0. We will use the shorthand notation $\hat{\beta}_0(z)$ to denote $\beta(z,\hat{\alpha}_0)$. Following Li, Huang, Li, and Fu (2002), we construct our test statistic based on $\int \left[\hat{\beta}(z) - \hat{\beta}_0(z)\right]' \left[\hat{\beta}(z) - \hat{\beta}_0(z)\right] dz$, where $\hat{\beta}(z) = \left[n^{-1}\sum_{i=1}^n X_i X_i' K_{\hat{\gamma},iz}\right]^{-1} \left[n^{-1}\sum_{i=1}^n X_i Y_i K_{\hat{\gamma},iz}\right]$ and where $K_{\hat{\gamma},iz} = \prod_{s=1}^{q_1} \hat{h}_s^{-1} w((Z_{is}^c - z_s^c)/\hat{h}_s) \prod_{s=1}^{r_1} l(Z_{is}^d, z_s^d, \hat{\lambda}_s)$, i.e., $\hat{\beta}(z)$ is the semiparametric estimator of $\beta(z)$ using cross-validated smoothing parameter selection. To avoid a random denominator problem in our test statistic, we modify the test statistic by inserting a positive definite matrix $D_n(z)'D_n(z)$ between $\left(\hat{\beta}(z) - \hat{\beta}_0(z)\right)'$ and $\left(\hat{\beta}(z) - \hat{\beta}_0(z)\right)$, where $D_n(z) = n^{-1}\sum_{i=1}^n X_i X_i' K_{\hat{\gamma},iz}$, which leads to the weighted statistic

$$\int \left[D_n(z)\left(\hat{\beta}(z) - \hat{\beta}_0(z)\right)\right]' \left[D_n(z)\left(\hat{\beta}(z) - \hat{\beta}_0(z)\right)\right] dz$$

$$= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n X_i' X_j \hat{u}_i \hat{u}_j \int K_{\hat{\gamma},iz} K_{\hat{\gamma},jz}\, dz, \tag{26}$$

where $\hat{u}_i = Y_i - X_i'\hat{\beta}_0(Z_i)$. To avoid a nonzero center term in the test statistic, we will remove the $i = j$ term in (26). Also, note that for $i \neq j$,

$$\int K_{\hat{\gamma},iz} K_{\hat{\gamma},jz}\, dz = \left[\int W_{\hat{h},iz^c} W_{\hat{h},jz^c}\, dz^c\right] \left[\sum_{z^d \in \mathcal{S}^d} L(Z_i^d, z^d, \hat{\lambda}) L(Z_j^d, z^d, \hat{\lambda})\right]$$

$$= \mathcal{W}_{\hat{h},ij} \mathcal{L}_{\hat{\lambda},ij},$$

where $\mathcal{W}_{\hat{h},ij} = \prod_{s=1}^{q_1} \hat{h}_s^{-1} w^0\left((Z_{is}^c - Z_{js}^c)/\hat{h}_s\right)$, $w^0(u) = \int w(v)\, w(u+v)\, dv$ is the twofold convolution kernel defined from $w(\cdot)$, and $\mathcal{L}_{\hat{\gamma},ij} = \sum_{z^d \in \mathcal{S}^d} L(Z_i^d, z^d, \hat{\lambda}) L(Z_j^d, z^d, \hat{\lambda})$. Because the convolution kernel $\mathcal{W}(\cdot)$ is also a (product) second-order kernel function, there is no need to use a convolution kernel in practice.

Rather, one can simply replace the convolution kernel $\mathcal{W}(\cdot)$ by the original kernel $\bar{W}(\cdot)$. Similarly, one can replace $\mathcal{L}_{\hat{\gamma},ij}$ by $L_{\hat{\gamma},ij} = \prod_{s=1}^{r_1} l(Z_{is}^d, Z_{js}^d, \hat{\gamma}_s)$. Our modified test statistic is then given by

$$\hat{I}_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} X_i' X_j \hat{u}_i \hat{u}_j K_{\hat{\gamma},ij}, \tag{27}$$

where $K_{\hat{\gamma},ij} = \prod_{s=1}^{q_1} \hat{h}_s^{-1} w((Z_{is}^c - Z_{js}^c)/\hat{h}_s) \prod_{s=1}^{r_1} l(Z_{is}^d, Z_{js}^d, \hat{\lambda}_s)$.

The asymptotic null distribution of $\hat{I}_n$ is given in the following theorem.

THEOREM 4.1. *Under the conditions given in Theorem 3.1, then, under H_0, we have*

$$\hat{T}_n \stackrel{\text{def}}{=} n\sqrt{\hat{h}_1 \ldots \hat{h}_{q_1}}\, \hat{I}_n / \hat{\sigma}_0 \xrightarrow{d} N(0,1),$$

where $\hat{\sigma}_0^2 = 2n^{-2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (X_i' X_j)^2 \hat{u}_i^2 \hat{u}_j^2 K_{\hat{\gamma},ij}^2$ is a consistent estimator of $\sigma_0^2 = 2\kappa^{q_1} \mathrm{E}[f(X_i, \bar{Z}_i)(X_i' X_j)^2 \sigma^4(X_i, \bar{Z}_i)]$ and where $\kappa = \int w^2(v)dv$ and $\sigma^2(X_i, \bar{Z}_i) = E(u_i^2 | X_i, \bar{Z}_i)$.

The proof of Theorem 4.1 is similar to the proof of Theorem 2.1 of Hsiao, Li, and Racine (2007). A sketch of the proof of Theorem 4.1 is provided in Appendix B.

It can also be shown that if the null hypothesis is false then $\hat{T}_n$ converges to $+\infty$ at the rate $n\sqrt{\hat{h}_1 \ldots \hat{h}_{q_1}} = O_p\left(n^{(8+q_1)/(8+2q_1)}\right)$. Hence, the test statistic $\hat{T}_n$ provides a consistent basis for detecting a parametric null model. However, it is known that kernel-based consistent model specification tests such as $\hat{T}_n$ that use asymptotic null distributions (such as that given in Theorem 4.1) are plagued by severe size distortions in finite-sample applications. Therefore, we propose a nonparametric bootstrap procedure to approximate the null distribution of $\hat{T}_n$.

We advocate using the following residual-based wild bootstrap method to approximate the null distribution of $\hat{T}_n$. The wild bootstrap error u_i^* is generated via a two point distribution $u_i^* = [(1-\sqrt{5})/2]\hat{u}_i$ with probability $(1+\sqrt{5})/[2\sqrt{5}]$, and $u_i^* = [(1+\sqrt{5})/2]\hat{u}_i$ with probability $(\sqrt{5}-1)/[2\sqrt{5}]$. From $\{u_i^*\}_{i=1}^n$, we generate the bootstrap sample dependent variable $Y_i^* = X_i'\beta_0(Z_i, \hat{\alpha}_0) + u_i^*$ for $i = 1, \ldots, n$. Calling $\{Y_i^*, X_i, Z_i\}_{i=1}^n$ the "bootstrap sample," we use this bootstrap sample to estimate the parametric null model and obtain $\hat{\alpha}_0^*$.[4] Next, we compute the bootstrap residual given by $\hat{u}_i^* = Y_i^* - X_i'\beta_0(Z_i, \hat{\alpha}^*)$. The bootstrap test statistic $\hat{T}_n^*$ is obtained from $\hat{T}_n$ with $\hat{u}_i \hat{u}_j$ being replaced by $\hat{u}_i^* \hat{u}_j^*$. Note that we use the same cross-validated smoothing parameters $\hat{h}_s$ $(s = 1, \ldots, q_1)$ and $\hat{\lambda}_s$ $(s = 1, \ldots, r_1)$ when computing the bootstrap statistics. That is, there is no need to rerun cross-validation for each bootstrap resample. Hence our bootstrap test is computationally quite simple. In practice, we repeat the preceding steps a large number of times, say, $B = 399$ times; see Davidson and MacKinnon (2000) for the appropriate number of bootstrap replications in these settings. The B bootstrap test statistics approximate the empirical finite-sample null distribution of $\hat{T}_n$, and these B statistics plus the original test statistic $\hat{T}_n$ can be used to compute nonparametric P-values.

Note that in the preceding bootstrap test procedure, we generate the bootstrap sample Y_i^* according to the null model. Therefore, even if the null hypothesis is false, our test statistic $\hat{T}_n^*$ still mimics the *null* distribution of $\hat{T}_n$. Hence, the empirical distribution function of $\hat{T}_n^*$ can be used to approximate the null distribution of $\hat{T}_n$ in practice, regardless of the validity of the null hypothesis.

The next theorem shows that the wild bootstrap works for the cross-validation-based $\hat{T}_n^*$ test.

THEOREM 4.2. *Under the same conditions as in Theorem 4.1 except that we do not require that the null hypothesis be true, we then have*

$$\sup_{v \in \mathbb{R}} \left| P\left(\hat{T}_n^* \leq v | \{Y_i, X_i, Z_i\}_{i=1}^n\right) - \Phi(v) \right| = o_p(1), \tag{28}$$

where $\hat{T}_n^* = n\sqrt{\hat{h}_1 \ldots \hat{h}_q}\, \hat{I}_n^* / \hat{\sigma}_0^*$, I_n^* *and* $\hat{\sigma}_0^*$ *are defined the same way as* $\hat{I}_n$ *and* $\hat{\sigma}_0$ *except that* $\hat{u}_i \hat{u}_j$ *is replaced by* $\hat{u}_i^* \hat{u}_j^*$ *whenever it occurs, and* $\Phi(\cdot)$ *is the cumulative distribution function of a standard normal random variable.*

The proof of Theorem 4.2 is similar to the proof of Theorem 2.2 of Hsiao, Li, and Racine (2007). We provide a sketch of the proof of Theorem 4.2 in Appendix B. In the next section we report Monte Carlo simulations designed to examine the finite-sample performance of both the test statistic $\hat{T}_n^*$ and the semiparametric estimator $\hat{\beta}(\cdot)$.

5. FINITE-SAMPLE PERFORMANCE

5.1. The Finite-Sample Performance of $\hat{\beta}(z)$

In this section we outline a modest Monte Carlo simulation designed to highlight the finite-sample behavior of the proposed estimator $\hat{\beta}(\cdot)$. In what follows, we shall consider the behavior of two estimators, namely, the proposed kernel method and the conventional frequency-based kernel method that breaks the data into subsets.

We simulate data from

$$Y_i = \beta_{0j} + \beta_{1ji} X_{i1} + \beta_2 X_{i2} + u_i, \qquad i = 1, \ldots, n, \qquad j = 0, \ldots, c-1,$$

where X_{i1} and X_{i2} are Uniform[0, 1], $u_i \sim N(0,1)$,

$$\beta_{0j} = \beta_0 + \eta_{0j}, \qquad \eta_{0j} \sim N(0,1), \qquad \beta_0 = 1, \qquad j = 0, \ldots, c-1,$$

$$\beta_{1ji} = \beta_{1j} + (Z_i^c)^2, \qquad \beta_{1j} = j+1, \qquad j = 0, \ldots, c-1,$$

with $Z_{i1}^c \sim N(0,1)$.

We draw c subsets of size n/c ($n > c$ and divides evenly) so that if, say, $n = 100$ and, say, $c = 2$, then we would have two subsets consisting of 50 observations for which $Y_i = \beta_{00} + \beta_{10i}X_{i1} + \beta_2 X_{i2} + u_i$ and 50 observations for which $Y_i = \beta_{01} + \beta_{11i}X_{i1} + \beta_2 X_{i2} + u_i$. Next we let Z_{i1}^d denote "group membership"; i.e., if, say, $n = 100$ and, say, $c = 2$, then for the 50 observations for which $Y_i = \beta_{00} + \beta_{10i}X_{i1} + \beta_2 X_{i2} + u_i$ we set $Z_{i1}^d = 0$, whereas for the 50 observations for which $Y_i = \beta_{01} + \beta_{11i}X_{i1} + \beta_2 X_{i2} + u_i$ we set $Z_{i1}^d = 1$. Finally, we generate another discrete covariate $Z_{i2}^d \in \{0, \dots, c-1\}$ that is uncorrelated with Y_i, i.e., is "irrelevant" though we do not presume this is known a priori. One can interpret Z_1^d as a discrete covariate (say, "group 1 membership") where the model changes with respect to Z_1^d, and Z_2^d can be interpreted as a discrete covariate (say, "group 2 membership") where there is no variation in the model with respect to this group.

In other words, the true data generating process (DGP) is a function of X_1, X_2, Z_1^c, and Z_1^d only, namely,

$$Y_i = \beta_0(Z_{i1}^d) + \beta_1(Z_{i1}^d, Z_{i1}^c)X_{i1} + \beta_2 X_{i2} + u_i.$$

However, Z_2^d is an irrelevant covariate, but this is not known a priori. Hence the user includes all covariates in the specification. The semiparametric model is of the form

$$Y_i = \beta_0(Z_{i1}^d, Z_{i2}^d, Z_{i1}^c) + \beta_1(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d)X_{i1} + \beta_2(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d)X_{i2} + u_i.$$

We draw $M = 1{,}000$ Monte Carlo replications from this DGP, estimate each model, consider settings having 4, 16, and 25 cells of data, and consider samples of size $n = 100, 200, 300, 400, 500$. In Table 1 we report the relative median MSE for the smooth and frequency-based kernel approaches (i.e., the ratio of the median MSE of the conventional sample-splitting approach to the median MSE of the proposed approach). Numbers greater than one indicate a loss of efficiency relative to the proposed method.

A quick scan of Table 1 reveals that the conventional semiparametric approach that breaks data into subsets results in substantial efficiency losses that worsen as the number of subsets increases and/or the sample size falls.

TABLE 1. Relative median MSE of the frequency estimator versus the proposed smooth estimator

	$c = 2$	$c = 4$	$c = 5$
100	1.92	4.41	5.50
200	1.82	3.91	4.96
300	1.80	3.65	4.62
400	1.79	3.63	4.51
500	1.79	3.60	4.38

Next we consider the performance of the cross-validated bandwidths for Z_1^d ($\hat{\lambda}_1$), Z_2^d ($\hat{\lambda}_2$), and Z_1^c ($\hat{h}$). Table 2 summarizes the median bandwidths over the $M = 1{,}000$ Monte Carlo replications.

Table 2 reveals that the cross-validated bandwidths for the relevant covariates ($\hat{\lambda}_1$ and $\hat{h}$) behave as expected, converging to zero as n increases, whereas that for the irrelevant covariate ($\hat{\lambda}_2$) converges instead to its upper bound value of one in probability as n increases.

This modest Monte Carlo simulation highlights the fact that the proposed method that smooths both the discrete and continuous covariates is more efficient than the conventional frequency-based approach. It also illustrates how cross-validated bandwidth selection delivers appropriate bandwidths for both relevant and irrelevant covariates. Next we consider the finite-sample performance of the proposed test for correct specification of parametric varying-coefficient models.

5.2. The Finite-Sample Performance of the Test $\hat{T}_n$

In this section we report Monte Carlo simulations designed to examine the finite-sample performance of the proposed bootstrap-based test for the statistic $\hat{T}_n$. For what follows, the null model is a parametric model of the form

$$\begin{aligned} Y_i &= \beta_0(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d) + \beta_1(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d)X_i + u_i \\ &= (1 + Z_{i1}^c + Z_{i1}^d + Z_{i2}^d) + X_i + u_i, \end{aligned}$$

whereas the alternative model is given by

$$\begin{aligned} Y_i &= \beta_0(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d) + \beta_1(Z_{i1}^c, Z_{i1}^d, Z_{i2}^d)X_i + u_i \\ &= (1 + Z_{i1}^c + Z_{i1}^d + Z_{i2}^d + \cos(Z_{i1}^c)/2) + (1 + Z_{i1}^c/2)X_i + u_i, \end{aligned}$$

where $X_i \sim \text{Uniform}[0,1]$, $Z_i^c \sim \text{Uniform}[-\pi, \pi]$, Z_{1i}^d and Z_{2i}^d are draws from the binomial distribution with c trials ($z^d \in \{0, 1, \ldots, c-1\}$ with probability of success $\frac{1}{2}$), and $u \sim N(0,1)$.

TABLE 2. Median bandwidths for the proposed estimator

	$c=2$			$c=4$			$c=5$		
	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{h}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{h}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{h}$
100	0.46	1.00	0.98	0.30	0.99	1.12	0.24	0.98	1.26
200	0.34	1.00	0.78	0.19	0.99	0.91	0.16	0.99	0.95
300	0.29	1.00	0.69	0.15	0.99	0.80	0.12	1.00	0.84
400	0.28	1.00	0.62	0.14	1.00	0.73	0.10	1.00	0.77
500	0.25	1.00	0.59	0.12	1.00	0.68	0.09	1.00	0.73

We consider two variants of the test, namely, the proposed test and that using the traditional frequency (sample-splitting) approach ($\lambda_1 = \lambda_2 = 0$). By way of comparison, we also consider two popular alternative tests, those of Härdle and Mammen (1993) and Bierens and Ploberger (1997). The Härdle and Mammen (HM) test takes the following form:

$$HM = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j}^{n} \hat{u}_i \hat{u}_j K_{\hat{\gamma},ij}, \tag{29}$$

where $\hat{\gamma}$ is the cross-validation selected smoothing parameters.

Bierens-type tests are nonsmoothing tests. Fan and Li (2000) have shown that the Bierens' test is equivalent to a kernel test but with fixed values for the smoothing parameters. Therefore, we will use the same test statistic as given in (29), but we replace $\hat{h}$ by z_{sd} (z_{sd} is the sample standard deviation of $\{Z_i^c\}_{i=1}^n$) and use $\lambda_1 = \lambda_2 = 0$ to obtain a Bierens and Ploberger (1997) (BP) type test. For both the BP and HM tests we obtain the null distributions via a wild bootstrap resampling procedure identical to that described for the proposed test.

In the simulations that follow, the estimated parametric model is always the null model. We simulate data under the null to assess the test's empirical size and then under the alternative to assess its empirical power. We draw $M = 1{,}000$ Monte Carlo replications. For each replication we compute $\hat{T}_n$ and HM using cross-validated bandwidth selection and then conduct $B = 399$ bootstrap replications and compute the empirical P-value. We then report the empirical rejection frequencies for $\alpha = 0.10, 0.05$, and 0.01. Results for $c = 2$ (four subsamples) and $c = 4$ (16 subsamples) are presented in Tables 3–6.

Tables 3–6 reveal that (i) the proposed test is correctly sized, (ii) it has power that increases with n, and (iii) it has substantially higher power than its frequency-based counterpart and it compares favorably to the tests of Bierens and Ploberger (1997) and Härdle and Mammen (1993).

TABLE 3. Empirical size for the proposed test, $c = 2$

	$\hat{T}_n$			$\hat{T}_n, \lambda_1 = \lambda_2 = 0$			HM			BP		
n	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	0.119	0.061	0.015	0.062	0.023	0.003	0.102	0.041	0.001	0.097	0.047	0.010
200	0.103	0.048	0.011	0.069	0.030	0.005	0.092	0.046	0.002	0.104	0.046	0.008
300	0.106	0.052	0.015	0.079	0.038	0.004	0.104	0.047	0.004	0.100	0.046	0.004
400	0.117	0.064	0.014	0.091	0.044	0.007	0.104	0.054	0.006	0.101	0.046	0.011
500	0.107	0.052	0.010	0.086	0.033	0.001	0.108	0.054	0.008	0.085	0.035	0.009

TABLE 4. Empirical size for proposed test, $c = 4$

	$\hat{T}_n$			$\hat{T}_n, \lambda_1 = \lambda_2 = 0$			HM			BP		
n	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	0.119	0.065	0.013	0.059	0.020	0.003	0.063	0.021	0.000	0.094	0.039	0.004
200	0.104	0.061	0.012	0.071	0.035	0.005	0.066	0.019	0.001	0.091	0.038	0.008
300	0.100	0.060	0.010	0.083	0.037	0.003	0.065	0.027	0.002	0.095	0.054	0.012
400	0.105	0.052	0.008	0.078	0.031	0.002	0.076	0.024	0.002	0.099	0.044	0.008
500	0.096	0.054	0.016	0.080	0.043	0.007	0.077	0.030	0.003	0.093	0.052	0.013

TABLE 5. Empirical power for proposed test, $c = 2$

	$\hat{T}_n$			$\hat{T}_n, \lambda_1 = \lambda_2 = 0$			HM			BP		
n	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	0.861	0.765	0.539	0.755	0.603	0.263	0.464	0.309	0.069	0.433	0.281	0.076
200	0.993	0.976	0.911	0.980	0.957	0.812	0.847	0.725	0.433	0.790	0.660	0.308
300	0.999	0.998	0.991	0.997	0.995	0.984	0.974	0.941	0.761	0.953	0.888	0.623
400	1.000	1.000	0.999	1.000	1.000	0.997	0.992	0.985	0.915	0.992	0.970	0.848
500	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.982	1.000	0.998	0.963

TABLE 6. Empirical power for proposed test, $c = 4$

	$\hat{T}_n$			$\hat{T}_n, \lambda_1 = \lambda_2 = 0$			HM			BP		
n	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	0.837	0.737	0.445	0.614	0.414	0.098	0.191	0.101	0.017	0.356	0.208	0.032
200	0.979	0.952	0.845	0.955	0.896	0.642	0.396	0.251	0.079	0.715	0.566	0.214
300	1.000	0.994	0.974	0.996	0.992	0.940	0.574	0.399	0.137	0.926	0.837	0.520
400	1.000	1.000	0.996	1.000	1.000	0.991	0.752	0.593	0.299	0.987	0.959	0.779
500	1.000	1.000	1.000	1.000	1.000	1.000	0.861	0.747	0.468	1.000	0.994	0.935

6. SUMMARY

In this paper we propose a novel varying-coefficient model that is capable of handling the mix of qualitative and quantitative covariates commonly encountered in applied work. A data-driven bandwidth selection method is proposed, theoretical underpinnings are provided, and a Monte Carlo experiment is undertaken to examine the finite-sample behavior of the proposed estimator. We also propose a consistent test for correct parametric specification of varying-coefficient models. In closing, we note that results for the case in which all elements of Z are qualitative cannot be obtained as a special case of those derived herein (which require the presence of at least one continuous variable). Hence we must treat this case separately and intend to do so in future work.

NOTES

1. We do not report these simulations here for space considerations. These results are available upon request.
2. Similarly, we use $\bar{\mathcal{S}}^d$, $\bar{\mathcal{S}}^c$, $\tilde{\mathcal{S}}^d$, and $\tilde{\mathcal{S}}^c$ to denote the support of $\bar{Z}_i^d$, $\bar{Z}_i^c$, $\tilde{Z}_i^d$, and $\tilde{Z}_i^c$, respectively.
3. For related work that uses least squares cross-validation for selecting smoothing parameters in a nonparametric regression model with continuous covariates, see Härdle and Marron (1985), Härdle, Hall, and Marron (1988), and Härdle, Hall, and Marron (1992).
4. The computation of $\hat{\alpha}_0^*$ is the same as for computing $\hat{\alpha}_0$ except that we use the bootstrap sample $\{Y_i^*, X_i, Z_i\}_{i=1}^n$ rather than the original sample $\{Y_i, X_i, Z_i\}_{i=1}^n$.

REFERENCES

Aitchison, J. & C.G.G. Aitken (1976) Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413–420.

Bhattacharya, R.N. & R.R. Rao (1986) *Normal Approximations and Asymptotic Expansions*. Krieger.

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

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1433–1458.

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

Cai, Z., M. Das, H. Xiong, & X. Wu (2006) Functional coefficient instrumental variables models. *Journal of Econometrics* 133, 207–241.

Cai, Z., J. Fan, & Q.W. Yao (2000) Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95, 941–956.

Davidson, R. & J.G. MacKinnon (2000) Bootstrap tests: How many bootstraps? *Econometric Reviews* 19, 55–68.

de Jong, P. (1987), A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

Escanciano, J.C. & D. Jacho-Chávez (2009) Uniform in bandwidth consistency of smooth varying coefficient estimators. *Economics Bulletin* 29, 1892–1898.

Fan, J., Q.W. Yao, & Z. Cai (2003) Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society, Series B* 65, 57–80.

Fan, Y. & Q. Li (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.

Gu, C. & P. Ma (2005) Generalized nonparametric mixed-effect models: Computation and smoothing parameter selection. *Journal of Computational and Graphical Statistics* 14, 485–504.

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

Hall, P., Q. Li, & J.S. Racine (2007) Nonparametric estimation of regression functions in the presence of irrelevant regressors, *Review of Economics and Statistics* 89, 784–789.

Hall, P., J.S. Racine, & Q. Li (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99, 1015–1026.

Härdle, W., P. Hall, & J.S. Marron (1988) How far are automatically chosen regression smoothing parameters from their optimum? *Journal of The American Statistical Association* 83, 86–101.

Härdle, W., P. Hall, & J.S. Marron (1992) Regression smoothing parameters that are not far from their optimum. *Journal of the American Statistical Association* 87, 227–233.

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

Härdle, W. & J. Marron (1985) Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics* 13, 1465–1481.
Hsiao, C., Q. Li, & J.S. Racine (2007) A consistent model specification test with mixed categorical and continuous data. *Journal of Econometrics* 140, 802–826.
Journal of Multivariate Analysis (2004) Vol. 91. Special issue on semiparametric and nonparametric mixed models.
Lee, J. (1990), *U-Statistics: Theory and Practice*. Marcel Dekker.
Li, Q., C.J. Huang, D. Li, & T.T. Fu (2002) Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics* 20, 412–422.
Li, Q. & J. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Li, Q. & J. Zhou (2005) The uniqueness of cross-validation selected smoothing parameters in kernel estimation of nonparametric models. *Econometric Theory* 21, 1017–1025.
Masry, E. (1996) Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81–101.
Rosenthal, H.P. (1970) On the subspace of l^p $(p \geq 1)$ spanned by sequences of independent random variables. *Israel Journal of Mathematics* 8, 273–303.
Zeger, S.L. & P.J. Diggle (1994) Semiparametric models for longitudinal data with application to cd4 cell numbers in HIV seroconverters. *Biometrics* 50, 689–699.
Zhang, D., X. Lin, J. Raz, & M. Sowers (1998) Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* 93, 710–719.

APPENDIX A: Proofs of Theorems 3.1 and 3.2

The proof of Theorem 3.1 is quite tedious. Therefore, it is necessary to introduce some shorthand notation and preliminary manipulations to simplify the derivations that follow. For the reader's convenience we list most of the notation used in this Appendix next.

1. We will write z_i for Z_i and use β_i to denote $\beta(Z_i)$ and $\hat{\beta}_{-i}$ to denote $\hat{\beta}_{-i}(Z_i)$.
2. We use $\sum_i = \sum_{i=1}^n$, $\sum\sum_{j\neq i} = \sum_{i=1}^n \sum_{j=1,j\neq i}^n$, $\sum\sum\sum_{l\neq j\neq i} = \sum_{i=1}^n \sum_{j=1,j\neq i}^n \sum_{l=1,l\neq i,l\neq j}^n$.
3. We write $A_n = B_n + (s.o.)$ to denote the fact that B_n is the leading term of A_n, where $(s.o.)$ denotes terms that have probability orders smaller than B_n. In what follows $A_i = B_i + (s.o.)$ always means that $n^{-1}\sum_i A_i = n^{-1}\sum_i B_i + (s.o.)$, and $A_{ij} = B_{ij} + (s.o.)$ means that $n^{-2}\sum_i \sum_j A_{ij} = n^{-2}\sum_i \sum_j B_{ij} + (s.o.)$. Also, we write $A_n \sim B_n$ to mean that A_n and B_n have the same order of magnitude in probability.
4. For notational simplicity we often ignore the difference between n^{-1} and $(n-1)^{-1}$ simply because this will have no effect on the asymptotic analysis.

In the proofs that follow we make use of Rosenthal's inequality (Rosenthal, 1970) and the U-statistic H-decomposition. We present these results next for the readers' convenience.

Rosenthal's Inequality. Let $p \geq 2$ be a positive constant and let $X_1, \dots, X_n$ denote i.i.d. random variables for which $\mathrm{E}(X_i) = 0$ and $\mathrm{E}(|X_i|^p) < \infty$. Then there exists a positive constant (which may depend on p) $C(p)$ such that

$$\mathrm{E}\left(\left|\sum_{i=1}^n X_i\right|^p\right) \leq C(p)\left\{\sum_{i=1}^n \mathrm{E}\left(|X_i|^p\right) + \left[\sum_{s=1}^n \mathrm{E}\left(X_i^2\right)\right]^{p/2}\right\}. \qquad \textbf{(A.1)}$$

Equation (A.1) is widely known as Rosenthal's inequality (see Rosenthal, 1970).

The H-Decomposition for U-Statistics. Let $(n,k) = n!/[k!(n-k!)]$ denote the number of combinations obtained by choosing k items from n (distinct) items. Then a general kth-order U-statistic $U_{(k)}$ is defined by

$$U_{(k)} = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k} H_n(X_{i_1}, \ldots, X_{i_k}),$$

where $H_n(X_{i_1}, \ldots, X_{i_k})$ is symmetric in its arguments and $\mathrm{E}[H_n^2(X_{i_1}, \ldots, X_{i_k})] < \infty$. In our proofs we will often use the following H-decomposition for a second-order U-statistic:

$$U_{(2)} = \theta + \frac{1}{n}\sum_i (H_{ni} - \theta) + \frac{2}{n(n-1)}\sum_i \sum_{j>i} [H_{n,ij} - H_{ni} - H_{nj} + \theta], \tag{A.2}$$

where $H_{n,ij} = H_n(X_i, X_j)$, $H_{ni} = \mathrm{E}[H_{n,ij}|X_i]$, and $\theta = \mathrm{E}[H_{n,ij}]$. We will also make use of the H-decomposition for a third-order U-statistic,

$$\begin{aligned} U_{(3)} &= \theta + \frac{1}{n}\sum_i (H_{ni} - \theta) + \frac{2}{n(n-1)}\sum \sum_{j>i} \left(H_{n,ij} - H_{ni} - H_{nj} + \theta\right) \\ &\quad + \frac{6}{n(n-1)(n-2)} \sum\sum \sum_{l>j>i} \\ &\quad \times \left(H_{n,ijl} - H_{n,ij} - H_{n,jl} - H_{n,li} + H_{ni} + H_{nj} + H_{nl} - \theta\right), \end{aligned} \tag{A.3}$$

where $H_{nijl} = H_n(X_i, X_j, X_l)$, $H_{nij} = \mathrm{E}[H_{nijl}|X_i, X_j]$, $H_{ni} = \mathrm{E}[H_{nij}|X_i]$, and $\theta = \mathrm{E}[H_{nijl}]$. For a derivation of the preceding formulas and an H-decomposition for a general kth-order U-statistic, see Lee (1990, p. 26).

Proof of Theorem 3.1. Using (11) and $Y_i = X_i'\beta_i + u_i$, we have (where $\beta_i = \beta(Z_i)$, $\hat\beta_{-i} = \hat\beta_{-i}(Z_i)$, and $M_i = M(Z_i)$)

$$\begin{aligned} CV(\gamma) &= \frac{1}{n}\sum_i [Y_i - X_i'\hat\beta_{-i}]^2 M_i \\ &= \frac{1}{n}\sum_i [X_i'(\beta_i - \hat\beta_{-i})]^2 M_i + \frac{2}{n}\sum_i u_i X_i'(\beta_i - \hat\beta_{-i}) M_i + n^{-1}\sum_i u_i^2 M_i. \end{aligned} \tag{A.4}$$

In what follows we obtain the leading terms of $CV(\gamma)$. We use $CV_0(\gamma)$ to denote the first two terms on the right-hand side of (A.4). Minimizing $CV(\gamma)$ over $\gamma = (h,\lambda)$ is equivalent to minimizing $CV_0(\gamma)$ as $n^{-1}\sum_i u_i^2$ does not depend on $\gamma = (h,\lambda)$, where

$$\begin{aligned} CV_0(\gamma) &\stackrel{\text{def}}{=} n^{-1}\sum_i \left[X_i'(\beta_i - \hat\beta_{-i})\right]^2 M_i + 2n^{-1}\sum_i u_i X_i'(\beta_i - \hat\beta_{-i}) M_i \\ &= CV_{0,1} + CV_{0,2}, \end{aligned} \tag{A.5}$$

where the definitions of $CV_{0,1}$ and $CV_{0,2}$ should be apparent.

Replacing $Y_j = X_j'\beta_j + u_j = X_j'\beta_i + X_j'(\beta_j - \beta_i) + u_j$ in (12), we get

$$\hat{\beta}_{-i} = \beta_i + \left[n^{-1}\sum_{j\neq i}^{n} X_j X_j' K_{\gamma,ij}\right]^{-1}$$

$$\times \left[n^{-1}\sum_{j\neq i} X_j X_j'(\beta_j - \beta_i)K_{\gamma,ij} + n^{-1}\sum_{j\neq i} X_j u_j K_{\gamma,ij}\right]$$

$$= \beta_i + \left[\hat{A}_i\right]^{-1}\left[\hat{B}_i + \hat{C}_i\right], \tag{A.6}$$

where $\hat{A}_i = n^{-1}\sum_{j\neq i}^{n} X_j X_j' K_{\gamma,ij}$, $\hat{B}_i = n^{-1}\sum_{j\neq i}^{n} X_j X_j'(\beta_j - \beta_i)K_{\gamma,ij}$, and $\hat{C}_i = n^{-1}\sum_{j\neq i}^{n} X_j u_j K_{\gamma,ij}$.

Substituting (A.6) into $CV_{0,1}$ we get

$$CV_{0,1} = n^{-1}\sum_i \left\{X_i'\hat{A}_i^{-1}\left[\hat{B}_i + \hat{C}_i\right]\right\}^2 M_i$$

$$= n^{-1}\sum_i \left[X_i'\hat{A}_i^{-1}\hat{B}_i\right]^2 M_i + n^{-1}\sum_i \left[X_i'\hat{A}_i^{-1}\hat{C}_i\right]^2 M_i$$

$$+ 2n^{-1}\sum_i \left[X_i'\hat{A}_i^{-1}\hat{B}_i\right]\left[X_i'\hat{A}_i^{-1}\hat{C}_i\right] M_i$$

$$\equiv CV_1 + CV_2 + 2CV_3,$$

where the definitions of CV_j $(j = 1,2,3)$ should be apparent.

In Lemmas A.1–A.3 in this Appendix we show, uniformly in $(h,\lambda) \in \Gamma$, that

$$CV_1 = \int \left\{\left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z})\right]' m(\bar{z})\left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z})\right]\right\}$$

$$\times \bar{M}(\bar{z})d\bar{z} + o_p\left(\zeta_n^2 + (nh_1\ldots h_{q_1})^{-1}\right), \tag{A.7}$$

$$CV_2 = \frac{\kappa^{q_1}}{nh_1\ldots h_{q_1}}\int \bar{f}(\bar{z})\tilde{f}(\tilde{z})\delta(\bar{z})\tilde{R}(\tilde{z})M(z)\,dz, \tag{A.8}$$

$$CV_3 = o_p\left(\zeta_n^2 + (nh_1\ldots h_{q_1})^{-1}\right), \tag{A.9}$$

where $B_{1s}(\cdot)$ and $B_{2s}(\cdot)$ are defined in (18), $\bar{M}(\bar{z}) = \int \tilde{f}(\tilde{z})M(z)d\tilde{z}$, $m(\bar{z}_i) = \mathrm{E}[X_i X_i'|\bar{z}_i]$ $\bar{f}(\bar{z}_i)$, $\delta(\bar{z}_i) = \mathrm{tr}\{\mathrm{E}[X_i' X_i \sigma^2(X_i,\bar{z}_i)|\bar{z}_i]m(\bar{z}_i)^{-1}\}$, $\tilde{R}(\tilde{z}) = \nu_2(\tilde{z})/\nu_1(\tilde{z})^2$, and $\zeta_n = \sum_{s=1}^{q_1} h_s^2 + \sum_{s=1}^{r_1}\lambda_s$. Also, in Lemma A.4 we show that

$$CV_{0,2} = o_p\left(\zeta_n^2 + (nh_1\ldots h_{q_1})^{-1}\right). \tag{A.10}$$

Hence, the leading term of $CV_0(\gamma)$ is given by $CV_1 + CV_2$. This proves (19). The remaining steps for proving Theorem 3.1 were stated in Section 3; see (20)–(22). ∎

Proof of Theorem 3.2. By Theorem 3.1 we know that $\hat{h}_s \xrightarrow{p} +\infty$ for $s = q_1+1,\dots,q$ and $\hat{\lambda}_s \xrightarrow{p} 1$ for $s = r_1+1,\dots,r$. Therefore, we need only consider the case with all irrelevant covariates removed; i.e., we consider $\hat{\beta}(\bar{z}) = \left[\sum_i X_i X_i \bar{K}_{\hat{\gamma},i\bar{z}}\right]^{-1} \sum_i X_i Y_i \bar{K}_{\hat{\gamma},i\bar{z}}$, where $\bar{K}_{\hat{\gamma},iz} = \left[\prod_{s=1}^{q_1} \hat{h}_s^{-1} w((z_{is}^c - z_s^c)/\hat{h}_s)\right]\left[\prod_{s=1}^{r_1} l(z_{is}^d, z_s^d, \hat{\lambda}_s)\right]$.

We first consider the benchmark case whereby we use nonstochastic smoothing parameters. Define $h_s^0 = a_s^0 n^{-1/(4+q_1)}$ for $s = 1,\dots,q_1$ and $\lambda_s^0 = b_s^0 n^{-2/(4+q_1)}$ for $s = 1,\dots,r_1$, where a_s^0 and b_s^0 are defined in (23). Also, define

$$\bar{\beta}(\bar{z}) = \left[\sum_{i=1}^n X_i X_i' \bar{K}_{\gamma^0,i\bar{z}}\right]^{-1} \sum_{i=1}^n X_i Y_i \bar{K}_{\gamma^0,i\bar{z}},$$

where $\bar{K}_{\gamma^0,iz} = \left[\prod_{s=1}^{q_1} (h_s^0)^{-1} w((z_{is} - z_s)/h_s^0)\right]\left[\prod_{s=1}^{r_1} l(z_{is}^d, z_s^d, \lambda_s^0)\right]$. Then, using $Y_i = X_i'\beta(\bar{z}_i) + u_i$ and by adding and subtracting terms in $\bar{\beta}(\bar{z})$, we obtain

$$\begin{aligned}
\bar{\beta}(\bar{z}) &= \left[\frac{1}{nh_1^0\dots h_{q_1}^0}\sum_i X_i X_i' \bar{K}_{\gamma^0,iz}\right]^{-1} \frac{1}{nh_1^0\dots h_{q_1}^0} \\
&\quad \times \sum_i X_i[X_i'\beta(\bar{z}) + X_i'(\beta(\bar{z}_i) - \beta(\bar{z})) + u_i]\bar{K}_{\gamma^0,iz} \\
&= \beta(\bar{z}) + \left[\frac{1}{nh_1^0\dots h_{q_1}^0}\sum_i X_i X_i' \bar{K}_{\gamma^0,iz}\right]^{-1} \frac{1}{nh_1^0\dots h_{q_1}^0} \\
&\quad \times \sum_i X_i[X_i'(\beta(\bar{z}_i) - \beta(\bar{z})) + u_i]\bar{K}_{\gamma^0,iz} \\
&\equiv \beta(\bar{z}) + \left[A^0(\bar{z})\right]^{-1}\left[B^0(\bar{z}) + C^0(\bar{z})\right],
\end{aligned} \tag{A.11}$$

where $A^0(\cdot) = (nh_1^0\dots h_{q_1}^0)^{-1}\sum_i X_i X_i' \bar{K}_{\gamma^0,iz}$, $B^0(\bar{z}) = (nh_1^0\dots h_{q_1}^0)^{-1}\sum_i X_i X_i'(\beta(\bar{z}_i) - \beta(\bar{z}))\bar{K}_{\gamma^0,iz}$, and $C^0(\bar{z}) = (nh_1^0\dots h_{q_1}^0)^{-1}\sum_i X_i u_i \bar{K}_{\gamma^0,iz}$.

By the same arguments as we used in the proof of Lemma A.1, one can show that $A^0(\bar{z}) = \mathrm{E}[X_i X_i' | \bar{z}_i = \bar{z}]\bar{f}(\bar{z}) + o_p(1) \equiv m(\bar{z}) + o_p(1)$ and that $B^0(\bar{z}) = m(\bar{z})\left[\sum_{s=1}^{q_1}(h_s^0)^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1}\lambda_s^0 B_{2s}(\bar{z})\right] + o_p(\zeta_n^0)$, where $\zeta_n^0 = \sum_{s=1}^{q_1}(h_s^0)^2 + \sum_{s=1}^{r_1}\lambda_s^0$. (It can be easily seen that $A^0(\bar{z})$ is a local constant kernel estimator of $\mathrm{E}[X_i X_i' | \bar{z}_i = \bar{z}]\bar{f}(\bar{z})$.) Obviously, $C^0(\bar{z})$ has zero mean, and its asymptotic variance, namely, $n^{-1}\left(h_1^0\dots h_{q_1}^0\right)^{-2}\mathrm{E}\left[X_i X_i' \sigma^2 (X_i, \bar{z}_i)\ \bar{K}_{\gamma^0,i\bar{z}}^2\right]$, is given by

$$\begin{aligned}
&(nh_1^0\dots h_{q_1}^0)^{-1}\left\{\kappa^{q_1}\bar{f}(\bar{z})\mathrm{E}[X_i X_i' \sigma^2(X_i, \bar{z}_i) | \bar{z}_i = \bar{z}] + o(1)\right\} \\
&\quad \equiv (nh_1^0\dots h_{q_1}^0)^{-1}[V(\bar{z}) + o(1)],
\end{aligned}$$

where $V(\bar{z}) = \kappa^{q_1} \bar{f}(\bar{z}) \mathrm{E}[X_i X_i' \sigma^2(X_i, \bar{z}_i) | \bar{z}_i = \bar{z}]$. By applying a triangular-array central limit theorem (CLT), we know that

$$\sqrt{n h_1^0 \dots h_{q_1}^0} \left[\bar{\beta}(\bar{z}) - \beta(\bar{z}) - \sum_{s=1}^{q_1} (h_s^0)^2 B_{1s}(\bar{z}) - \sum_{s=1}^{r_1} \lambda_s^0 B_{2s}(\bar{z}) \right] \xrightarrow{d} N\left(0, m_z^{-1} V_z m_z^{-1}\right), \tag{A.12}$$

where $m_z = m(\bar{z})$ and $V_z = V(\bar{z})$. It is obvious that $m_z^{-1} V_z m_z^{-1} = \Omega(\bar{z})$ as given in Theorem 3.2.

Next, we consider $\hat{\beta}(\bar{z}) = \left[\sum_i X_i X_i \bar{K}_{\hat{\gamma},iz}\right]^{-1} \sum_i X_i Y_i \bar{K}_{\hat{\gamma},iz}$ with cross-validation selected smoothing parameters, where $\bar{K}_{\hat{\gamma},iz} = \left[\prod_{s=1}^{q_1} \hat{h}_s^{-1} w((z_{is}^c - z_s^c)/\hat{h}_s)\right] \left[\prod_{s=1}^{r_1} l(z_{is}^d, z_s^d, \hat{\lambda}_s)\right]$. Therefore, the only difference between $\hat{\beta}(\bar{z})$ and $\bar{\beta}(z)$ is that the former uses the cross-validated smoothing parameters, whereas the latter uses some benchmark nonstochastic smoothing parameters. By Theorem 3.1 we know that $\hat{h}_s/h_s^0 \xrightarrow{p} 1$ for $s = 1, \dots, q_1$ and $\hat{\lambda}_s/\lambda_s^0 \xrightarrow{p} 1$ for $s = 1, \dots, r_1$. By using stochastic equicontinuity arguments as in Hall, Racine, and Li (2004) or in Hsiao, Li, and Racine (2007), one can show that $\hat{\beta}(\bar{z})$ and $\bar{\beta}(\bar{z})$ have the same asymptotic distribution, i.e.,

$$\sqrt{n \hat{h}_1 \dots \hat{h}_{q_1}} \left[\hat{\beta}(\bar{z}) - \beta(\bar{z}) - \sum_{s=1}^{q_1} \hat{h}_s^2 B_{1s}(\bar{z}) - \sum_{s=1}^{r_1} \hat{\lambda}_s B_{2s}(\bar{z}) \right] \xrightarrow{d} N\left(0, m_z^{-1} V_z m_z^{-1}\right). \tag{A.13}$$

Finally, we show that $\hat{\mathcal{A}}^{-1} \hat{\mathcal{B}} \hat{\mathcal{A}}^{-1}$ (defined in Theorem 3.2) is a consistent estimator of $\Omega(\bar{z})$. It is easy to show that $\hat{\mathcal{A}} = A^0(\bar{z}) + o_p(1)$. Then it is straightforward to show that $A^0(\bar{z}) = m_z + o_p(1)$. Hence, $\hat{\mathcal{A}} = m_z + o_p(1)$. Similarly, one can show that $\hat{\mathcal{B}} = V_z + o_p(1)$. Hence, $\hat{\mathcal{A}}^{-1} \hat{\mathcal{B}} \hat{\mathcal{A}}^{-1} = m_z^{-1} V_z m_z^{-1} + o_p(1) \equiv \Omega(\bar{z}) + o_p(1)$. This completes the proof of Theorem 3.2. ■

In the remaining part of this Appendix we prove some lemmas that are used in the proof of Theorem 3.1. For notational convenience we will write z_i for Z_i.

LEMMA A.1. *Equation (A.7) holds true.*

Proof. By Lemma A.5 we know that $\hat{A}(z)$ is the kernel estimator of $\mu(z) = m(\bar{z}) \nu_1(\tilde{z})$, where $m(\bar{z}) = \mathrm{E}[X_i X_i' | \bar{Z}_i = \bar{z}] \bar{f}(\bar{z})$ and $\nu_1(\tilde{z}) = \mathrm{E}[\tilde{K}_{\tilde{\gamma},ij} | \tilde{z}_i = \tilde{z}]$. Therefore, we know that (see Lemma A.5) the leading term of $\hat{A}(z_i)^{-1}$ is $\mu(z_i)^{-1}$. Define CV_1^0 by replacing $\hat{A}(z_i)^{-1}$ in CV_1 by its leading term $\mu(z_i)^{-1}$. Then using the result of Lemma A.5, it is easy to show that $CV_1 = CV_1^0 + (s.o.)$. Hence, we only need to consider CV_1^0, which is defined by (recall that $\hat{B}_i = n^{-1} \sum_{j \neq i} X_j X_j' (\beta_j - \beta_i) K_{\gamma,ij}$)

$$\begin{aligned} CV_1^0 &= n^{-1} \sum_i \left[X_i' \mu(z_i)^{-1} \hat{B}_i \right]^2 M_i \\ &= n^{-3} \sum_i \sum_{j \neq i} \sum_{l \neq i} X_i' \mu(z_i)^{-1} X_j X_j' (\beta_j - \beta_i) K_{\gamma,ij} X_i' \mu(z_i)^{-1} X_l X_l' (\beta_l - \beta_i) K_{\gamma,il} M_i \\ &= n^{-3} \sum_i \sum_{j \neq i} \left[X_i' \mu(z_i)^{-1} X_j X_j' (\beta_j - \beta_i) K_{\gamma,ij} \right]^2 M_i \end{aligned}$$

$$+n^{-3}\sum_i \sum_{j\neq i} \sum_{l\neq i, l\neq j} X_i'\mu(z_i)^{-1}X_jX_j'(\beta_j-\beta_i)K_{\gamma,ij}X_i'\mu(z_i)^{-1}$$

$$\times X_lX_l'(\beta_l-\beta_i)K_{\gamma,il}M_i$$

$$\equiv CV_{1,1}+CV_{1,2},$$

where the definitions for $CV_{1,1}$ and $CV_{1,2}$ should be apparent.

By noting that $\mu(z_i)=m(\bar{z}_i)\nu_1(\tilde{z})$ and that $\nu_1(\tilde{z})=\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i]$, it is fairly straightforward to show that (note that $|M(\cdot)|\leq 1$)

$$\mathrm{E}\left[|CV_{1,1}|\right]\leq n^{-3}n(n-1)\mathrm{E}\left\{\left[X_i'm(\bar{z}_i)^{-1}X_jX_j'(\beta_j-\beta_i)\bar{K}_{\bar{\gamma},ij}\right]^2\right\}\mathrm{E}\left\{\tilde{K}^2_{\tilde{\gamma},ij}/\nu_1(\tilde{z}_i)^2\right\}$$

$$\leq C(nh_1\ldots h_{q_1})^{-1}\left(\sum_{i=1}^{q_1}h_s^2+\sum_{s=1}^{r_1}\lambda_s\right)=o\left((nh_1\ldots h_{q_1})^{-1}\right), \qquad \textbf{(A.14)}$$

because it is easy to see that $\mathrm{E}\left\{\left[X_i'm(\bar{z}_i)^{-1}X_jX_j'(\beta_j-\beta_i)\bar{K}_{\bar{\gamma},ij}\right]^2\right\}=O\left(\sum_{i=1}^{q_1}h_s^2 + \sum_{s=1}^{r_1}\lambda_s\right)$ and that $\mathrm{E}\{[\tilde{K}_{\tilde{\gamma},ij}]^2/\nu_1(\tilde{z}_i)^2\}=\mathrm{E}\{[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i]^2/(\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i])^2\}=O(1)$. Hence, (A.14) implies that $CV_{1,1}=o_p\left((nh_1\ldots h_{q_1})^{-1}\right)$.

Next, $CV_{1,2}$ can be written as a third-order U-statistic. By the U-statistic H-decomposition, one can show that $CV_{1,2}=\mathrm{E}[CV_{1,2}]+(s.o.)$. Before evaluating $\mathrm{E}[CV_{1,2}]$, we first compute an intermediate quantity. Recall that $\mu(z)=m(\bar{z})\nu_1(\tilde{z})$ and that $m(\bar{z})=\mathrm{E}[X_jX_j'|\bar{z}_j=\bar{z}]\bar{f}(\bar{z})$. We have (noting that $\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}/\nu_1(\tilde{z}_i)|\tilde{z}_i]=\nu_1(\tilde{z})/\nu_1(\tilde{z}_i)=1$)

$$\mathrm{E}\left[X_i'\mu(z_i)^{-1}X_jX_j'(\beta_j-\beta_i)K_{\gamma,ij}|X_i,z_i\right]$$

$$=X_i'm(\bar{z}_i)^{-1}\mathrm{E}\left[\mathrm{E}(X_jX_j'|\bar{z}_j)(\beta_j-\beta_i)\bar{K}_{\bar{\gamma},ij}|X_i,\bar{z}_i\right]\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}/\nu_1(\tilde{z}_i)|\tilde{z}_i]$$

$$=X_i'm(\bar{z}_i)^{-1}\int\bar{f}(\bar{z}_j)\mathrm{E}(X_jX_j'|\bar{z}_j)(\beta_j-\beta_i)\bar{K}_{\gamma,ij}d\bar{z}_j$$

$$\equiv X_i'm(\bar{z}_i)^{-1}\int m(\bar{z}_j)(\beta_j-\beta_i)\bar{K}_{\gamma,ij}d\bar{z}_j$$

$$=X_i'm(\bar{z}_i)^{-1}\sum_{\bar{z}_j^d\in\bar{S}^d}\int m(\bar{z}_i^c+hv,\bar{z}_j^d)(\beta(\bar{z}_i^c+hv,\bar{z}_j^d)-\beta(\bar{z}_i^c,\bar{z}_i^d))W(v)L(\bar{z}_i^d,\bar{z}^d,\lambda)dv$$

$$=X_i'm(\bar{z}_i)^{-1}\left\{\kappa_2\sum_{s=1}^{q_1}h_s^2\left[m_s(z_i^c,z_i^d)\beta_s(\bar{z}_i^c,\bar{z}_i^d)+(1/2)m(\bar{z}_i^c,\bar{z}_i^d)\beta_{ss}(\bar{z}_i^c,\bar{z}_i^d)\right]\right.$$

$$\left.+\sum_{s=1}^{r_1}\lambda_s\sum_{\bar{v}^d\in\bar{S}^d}\mathbf{1}_s(\bar{z}_i^d,\bar{v}^d)m(\bar{z}_i^c,\bar{v}_d)\left[\beta(\bar{z}_i^c,\bar{v}^d)-\beta(\bar{z}_i^c,\bar{z}_i^d)\right]\right\}+o_p(\zeta_n)$$

$$\equiv X_i'\left[\sum_{s=1}^{q_1}h_s^2B_{1s}(\bar{z}_i)+\sum_{s=1}^{r_1}\lambda_sB_{2s}(\bar{z}_i)\right]+o_p(\zeta_n) \qquad \textbf{(A.15)}$$

uniformly in $(h,\lambda)\in\Gamma$ and $\zeta_n=\sum_{s=1}^{q_1}h_s^2+\sum_{s=1}^{r_1}\lambda_s$.

Substituting (A.15) into $CV_{1,2}$ and letting $\xi(\bar{z}) = \mathrm{E}[X_i X_i' | \bar{z}_j = \bar{z}]$ (so $m(\bar{z}) = \xi(\bar{z})\bar{f}(\bar{z})$), we immediately obtain (where $\bar{M}(\bar{z}) = \int \tilde{f}(\tilde{z}) M(z) d\tilde{z}$)

$$\begin{aligned}
\mathrm{E}[CV_{1,2}] &= \mathrm{E}\left\{ \left[\sum_{s=1}^{q_1} h_s^2 B_{1s,i} + \sum_{s=1}^{r_1} \lambda_s B_{2s,i} \right]' \xi(\bar{z}_i) \right. \\
&\quad \left. \times \left[\sum_{s=1}^{q_1} h_s^2 B_{1s,i} + \sum_{s=1}^{r_1} \lambda_s B_{2s,i} \right] \bar{M}(\bar{z}_i) \right\} + (s.o.) \\
&= \int \left\{ \left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z}) \right]' m(\bar{z}) \right. \\
&\quad \left. \times \left[\sum_{s=1}^{q_1} h_s^2 B_{1s}(\bar{z}) + \sum_{s=1}^{r_1} \lambda_s B_{2s}(\bar{z}) \right] \right\} \bar{M}(\bar{z}) d\bar{z} + (s.o.),
\end{aligned} \tag{A.16}$$

where in the last equality we used the fact that $\mathrm{E}(\cdot) = \int (\cdot) \bar{f}(\bar{z}) d\bar{z}$ and $\xi(\bar{z})\bar{f}(\bar{z}) = m(\bar{z})$. Hence, the leading term of CV_1 is given by (A.16). Note that in the preceding discussion we have only shown that for all fixed values of $(h, \lambda) \in \Gamma$ that (A.16) holds true. By utilizing Rosenthal's and Markov's inequalities, it is straightforward to show that (A.16) holds true uniformly in $(h, \lambda) \in \Gamma$. ■

LEMMA A.2. *Equation (A.8) holds true.*

Proof. We use CV_2^0 to denote CV_2 with $\hat{A}(z_i)^{-1}$ being replaced by its leading term $\mu(z_i)^{-1}$. Using Lemma A.5 it is fairly straightforward to show that $CV_2 = CV_2^0 + (s.o.)$. Hence, we only need to consider CV_2^0, which is given by

$$\begin{aligned}
CV_2^0 &= n^{-3} \sum_i \sum_{j \neq i} \sum_{l \neq i} X_i' \mu(z_i)^{-1} X_j u_j K_{\gamma,ij} X_i' \mu(z_i)^{-1} X_l u_l K_{\gamma,il} M_i \\
&= n^{-3} \sum_i \sum_{j \neq i} (X_i' \mu(z_i)^{-1} X_j)^2 u_j^2 K_{\gamma,ij}^2 M_i \\
&\quad + n^{-3} \sum_i \sum_{j \neq i} \sum_{l \neq j, l \neq j} X_i' \mu(z_i)^{-1} X_j u_j K_{\gamma,ij} X_i' \mu(z_i)^{-1} X_l u_l K_{\gamma,il} M_i \\
&\equiv D_1 + D_2.
\end{aligned}$$

We first consider D_2, which has mean zero. By noting that terms related to $\tilde{K}_{\tilde{\gamma},ij}$ are bounded because the kernel functions $W(\cdot)$ and $L(\cdot)$ are bounded and that $h_{q_1+1} \dots h_q$ from the $\tilde{K}$ and from $\mu(\cdot)^{-1}$ cancel, then it is fairly straightforward to show that $\mathrm{E}[D_2^2] = n^{-6} O\left(n^4 (h_1 \dots h_{q_1})^{-1}\right) = O\left(\left(n^2 h_1 \dots h_{q_1}\right)^{-1}\right)$. Hence, $D_2 = O_p\left(n^{-1}(h_1 \dots h_{q_1})^{-1/2}\right) = o_p\left(\left(n h_1 \dots h_{q_1}\right)^{-1}\right)$.

Next, we consider D_1. The term D_1 can be written as a second-order U-statistic. By the U-statistic H-decomposition it is straightforward, though tedious, to show that

$D_1 = \mathrm{E}(D_1) + (s.o.)$. By Lemma A.5 we know that $h_s \to 0$ for $s = 1, \ldots, q_1$ and $\lambda_s \to 0$ for $s = 1, \ldots, r_1$. We first compute the following intermediate result:

$$
\begin{aligned}
&\mathrm{E}\left[(X_i' m(\bar{z}_i)^{-1} X_j)^2 \sigma^2(X_j, \bar{z}_j) | \bar{z}_i, \bar{z}_j\right] \\
&= \mathrm{E}\left[(X_j' m(\bar{z}_i)^{-1} \mathrm{E}(X_i X_i' | \bar{z}_i) \bar{m}(\bar{z}_i)^{-1} X_j \sigma^2(X_j, \bar{z}_j) | \bar{z}_j\right] \\
&= \mathrm{E}\left[(X_j' m(\bar{z}_i)^{-1} X_j \sigma^2(X_j, \bar{z}_j) | \bar{z}_j\right] \bar{f}(\bar{z}_i)^{-1} \\
&= \mathrm{tr}\{\mathrm{E}\left[X_j' X_j \sigma^2(X_j, \bar{z}_j) | \bar{z}_j\right] m(\bar{z}_i)^{-1}\} \bar{f}(\bar{z}_i)^{-1} \\
&= \mathrm{tr}\{G(\bar{z}_j) m(\bar{z}_i)^{-1}\} \bar{f}(\bar{z}_i)^{-1},
\end{aligned} \tag{A.17}
$$

where $G(\bar{z}_j) = \mathrm{E}[X_j' X_j \sigma^2(X_j, \bar{z}_j) | \bar{z}_j]$.

By recalling that $\mu(z) = m(\bar{z}) \nu_1(\tilde{z})$ and $\nu_2(\tilde{z}) = \mathrm{E}[\tilde{K}^2_{\tilde{i}j} | \tilde{z}_i = \tilde{z}]$, using (A.17) and the law of iterated expectations, $\mathrm{E}(D_1) = \mathrm{E}\{\mathrm{E}[\mathrm{E}(D_1 | z_i, z_j) | z_i]\}$, we obtain

$$
\begin{aligned}
\mathrm{E}(D_1) &= n^{-1} \mathrm{E}\left\{\mathrm{E}[(X_i' m(\bar{z}_i)^{-1} X_j)^2 \sigma^2(X_j, \bar{z}_j) | \bar{z}_i, \bar{z}_j] \bar{K}^2_{\bar{\gamma},ij} \mathrm{E}[\tilde{K}^2_{\tilde{\gamma},ij} | \tilde{z}_i] \nu_1(\tilde{z}_i)^{-2} M_i\right\} \\
&= n^{-1} \mathrm{E}\left\{\mathrm{E}\left[\mathrm{tr}\left(G(\bar{z}_j) m(\bar{z}_i)^{-1}\right) \bar{f}(\bar{z}_i)^{-1} \bar{K}^2_{\bar{\gamma},ij} | \bar{z}_i\right] M_i \nu_2(\tilde{z}_i) / \nu_1(\tilde{z}_i)^2\right\} \\
&= (nh_1 \ldots h_{q_1})^{-1} \kappa^{q_1} \mathrm{E}\left\{\mathrm{tr}\left[G(\bar{z}_i) m(\bar{z}_i)^{-1}\right] M_i \tilde{R}(\tilde{z}_i)\right\} \\
&\quad + O\left(\zeta_n^{1/2} (nh_1 \ldots h_{q_1})^{-1}\right) \\
&= (nh_1 \ldots h_{q_1})^{-1} \kappa^{q_1} \int \bar{f}(\bar{z}) \tilde{f}(\tilde{z}) \delta(\bar{z}) \tilde{R}(\tilde{z}) M(z) dz + O\left(\zeta_n^{1/2} (nh_1 \ldots h_{q_1})^{-1}\right),
\end{aligned}
$$

where $\delta(\bar{z}) = \mathrm{tr}[G(\bar{z}) m(\bar{z})^{-1}]$, $\tilde{R}(\tilde{z}) = \nu_2(\tilde{z}) / \nu_1(\tilde{z})^2$, and $\kappa^{q_1} = \int \bar{W}(v)^2 dv \equiv [\int w^2(v) dv]^{q_1}$. Summarizing the preceding discussion we have shown that

$$
\begin{aligned}
CV_2 &= CV_2^0 + (s.o.) = \mathrm{E}(D_1) + (s.o.) \\
&= (nh_1 \ldots h_{q_1})^{-1} \kappa^{q_1} \int \bar{f}(\bar{z}) \tilde{f}(\tilde{z}) \delta(\bar{z}) \tilde{R}(\tilde{z}) M(z) dz + (s.o.).
\end{aligned}
$$

Moreover, by utilizing Rosenthal's and Markov's inequalities, one can show that the preceding result holds uniformly in $(h, \lambda) \in \Gamma$. This completes the proof of Lemma A.2. ■

LEMMA A.3. *Equation (A.9) holds true.*

Proof. By Lemma A.5 we know that $\hat{A}(z) = \mu(z)$ uniformly in $z \in \mathcal{M}$ and $(h, \lambda) \in \Gamma$, where $\mathcal{M}$ denotes the support of the weight function $M(\cdot)$. Let CV_3^0 denote CV_3 with

$\hat{A}(z_i)$ being replaced by its leading term $\mu(z_i)$. Then it can be shown that $CV_3 = CV_3^0 + (s.o.)$. Hence, we only need to consider CV_3^0, which is defined by

$$
\begin{aligned}
CV_3^0 &= n^{-3}\sum_i \sum_{j\neq i} \sum_{l\neq i} X_i'\mu(z_i)^{-1}X_jX_j'(\beta_j-\beta_i)K_{\gamma,ij}X_i'\mu(z_i)^{-1}X_lu_lK_{\gamma,il} \\
&= n^{-3}\sum_i \sum_{j\neq i} X_i'\mu(z_i)^{-1}X_jX_j'(\beta_j-\beta_i)X_i'\mu(z_i)^{-1}X_ju_jK_{\gamma,ij}^2 \\
&\quad + n^{-3}\sum_i \sum_{j\neq i} \sum_{l\neq i, l\neq j} X_i'\mu(z_i)^{-1}X_jX_j'(\beta_j-\beta_i)K_{\gamma,ij}X_i'\mu(z_i)^{-1}X_lu_lK_{\gamma,il} \\
&\equiv CV_{3,1} + CV_{3,2},
\end{aligned}
$$

where the definitions of $CV_{3,1}$ and $CV_{3,2}$ should be apparent.

We first consider $CV_{3,1}$. The term $CV_{3,1}$ has zero mean. By following the same arguments as we did in the proof of Lemma A.1, one can show that the part in $CV_{1,2}$ that is related to the irrelevant covariates is bounded. This occurs because we have $\nu_1(\tilde{z}_i)^{-2} = \{\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i]\}^{-2}$ from $\mu(z_i)^{-2} = m(\bar{z}_i)^{-2}\nu_1(\tilde{z}_i)^{-2}$, which when combined with $\tilde{K}_{\tilde{\gamma},ij}^2$ always gives us a bounded quantity when computing the moments of $CV_{3,1}$. Therefore, when evaluating the order of $CV_{3,1}$ we can ignore the irrelevant covariates part and need only consider

$$
\bar{CV}_{3,1} \stackrel{\text{def}}{=} n^{-3}\sum_i \sum_{j\neq i} X_i'm(\bar{z}_i)^{-1}X_jX_j'(\beta_j-\beta_i)X_i'm(\bar{z}_i)^{-1}X_ju_j\bar{K}_{\bar{\gamma},ij}^2.
$$

Note that $\bar{CV}_{3,1}$ only depends on $(h_1,\ldots,h_{q_1},\lambda_1,\ldots,\lambda_{r_1})$. By Lemma A.6 we know that these smoothing parameters all converge to zero as $n\to\infty$. Hence, we can use standard change-of-variable and Taylor expansion arguments to deal with the continuous covariates' kernel function and use the polynomial expansion for the discrete kernel functions. Then, it is straightforward to show that the second moment of $\bar{CV}_{3,1}$ is of order (by noting that u_j has zero mean)

$$
\begin{aligned}
\mathrm{E}[\bar{CV}_{3,1}^2] &= n^{-6}\sum_i \sum_{j\neq i} \sum_{i'\neq j} \mathrm{E}\Big[X_i'm(\bar{z}_i)^{-1}X_jX_j'(\beta_j-\beta_i)X_i'm(\bar{z}_i)^{-1}X_l\sigma^2(X_j,\bar{z}_j)\bar{K}_{\bar{\gamma},ij}^2 \\
&\qquad \times X_{i'}'m(\bar{z}_{i'})^{-1}X_jX_j'(\beta_j-\beta_{i'})X_{i'}'m(\bar{z}_{i'})^{-1}X_j\bar{K}_{\bar{\gamma},i'j}^2\Big] \\
&= n^{-6}O\left(n^3\zeta_n^2\left(h_1\ldots h_{q_1}\right)^{-2} + n^2\zeta_n\left(h_1\ldots h_{q_1}\right)^{-3}\right), \qquad \textbf{(A.18)}
\end{aligned}
$$

where the first term corresponds to three indexes, i, j, i', each different from the other, and the second term corresponds to the case where $i' = i$ so that there are only two summations in the second case ($\zeta_n = \sum_{s=1}^{q_1} h_s^2 + \sum_{s=1}^{r_1}\lambda_s$).

Hence,

$$
\begin{aligned}
CV_{3,1} &= O_P\left(\zeta_n n^{-1/2}\left(nh_1\ldots h_{q_1}\right)^{-1} + \zeta_n^{1/2}n^{-1/2}\left(nh_1\ldots h_{q_1}\right)^{-3/2}\right) \\
&= o_p\left(\zeta_n^2 + \left(nh_1\ldots h_{q_1}\right)^{-1}\right).
\end{aligned}
$$

Similarly, when evaluating the moments of $CV_{3,2}$, we can ignore the irrelevant covariate part of $CV_{3,2}$ and need only consider

$$\bar{CV}_{3,2} = n^{-3} \sum_i \sum_{j \neq i} \sum_{l \neq i, l \neq j} X_i' m(\bar{z}_i)^{-1} X_j X_j' (\beta_j - \beta_i) \bar{K}_{\bar{\gamma},ij} X_i' m(\bar{z}_i)^{-1} X_l u_l \bar{K}_{\bar{\gamma},il}.$$

Note that $\bar{CV}_{3,1}$ also has zero mean and it can be shown that its second moment is

$$\begin{aligned} \mathrm{E}[\bar{CV}_{3,2}^2] = n^{-6} &\sum_i \sum_{j \neq i} \sum_{i'} \sum_{j' \neq i'} \sum_{l \neq i, j, i', j' j} \\ &\times \mathrm{E}\Big[X_i' m(\bar{z}_i)^{-1} X_j X_j' (\beta_j - \beta_i) \bar{K}_{\bar{\gamma},ij} X_i' m(\bar{z}_i)^{-1} X_l u_l^2 \bar{K}_{\bar{\gamma},il} \\ &\quad \times X_{i'}' m(\bar{z}_{i'})^{-1} X_{j'} X_{j'}' (\beta_{j'} - \beta_{i'}) \bar{K}_{\bar{\gamma},i'j'} X_{i'}' m(\bar{z}_{i'})^{-1} X_l \bar{K}_{\bar{\gamma},i'l}\Big]. \end{aligned} \quad \textbf{(A.19)}$$

The second moment of $\bar{CV}_{3,2}$ only has five summations because of the fact that u_l has zero mean. If the five summation indexes i, j, l, i', j' all differ from one another, then by the independent data assumption, we can compute $\mathrm{E}[X_i' m(\bar{z}_i)^{-1} X_j X_j' (\beta_j - \beta_i) \bar{K}_{\bar{\gamma},ij} X_i'$ $m(\bar{z}_i)^{-1} X_l u_l^2 \bar{K}_{\bar{\gamma},il}]$ and $\mathrm{E}[X_{i'}' m(\bar{z}_{i'})^{-1} X_{j'} X_{j'}' (\beta_{j'} - \beta_{i'}) \bar{K}_{\bar{\gamma},i'j'} X_{i'}' m(\bar{z}_{i'})^{-1} X_l \bar{K}_{\bar{\gamma},i'l}]$ separately. Each term will give us an order $\zeta_n = \sum_{s=1}^{q_1} h_s^2 + \sum_{s=1}^{r_1} \lambda_s$. Hence, we have an order $n^5 \zeta_n^2$ for this case. Similarly, one can easily show that if the five indexes take only four different values, we will get an order $\left(n^4 \zeta_n^2 + n^4 \zeta_n^2 \left(h_1 \dots h_{q_1}\right)^{-1}\right)$. Other cases give even smaller orders because they involve fewer summations. Therefore, we conclude that

$$\mathrm{E}[\bar{CV}_{3,2}^2] = n^{-6} O\left(\zeta_n^2 n^5 + n^4 \zeta_n \left(h_1 \dots h_{q_1}\right)^{-1}\right) = O\left(n^{-1} \zeta_n^2\right).$$

Hence, $CV_{3,2} = O_p\left(n^{-1/2} \zeta_n^2\right) = o_p\left(\zeta_n^2\right)$. Thus, we have shown that

$$CV_3 = CV_3^0 + (s.o) = o_p\left(\zeta_n^2 + \left(n h_1 \dots h_{q_1}\right)^{-1}\right). \quad \textbf{(A.20)}$$

Moreover, by utilizing Rosenthal's and Markov's inequalities, one can show that (A.20) holds uniformly in $(h, \lambda) \in \Gamma$. Hence, (A.16) holds true. ■

LEMMA A.4. $CV_{0,2} = o_p\left(\zeta_n^2 + \left(n h_1 \dots h_{q_1}\right)^{-1}\right)$ *uniformly in* $(h, \lambda) \in \Gamma$.

Proof. Letting $CV_{0,2}^0$ denote $CV_{0,2}$ (defined in (A.5)) with $\mu(z_i)$ replacing $\hat{A}(z_i)$ in $CV_{0,2}$, we get

$$\begin{aligned} CV_{0,2}^0 &= n^{-1} \sum_i u_i X_i' \mu(z_i)^{-1} \left[\hat{B}_i + \hat{C}_i\right] \\ &= n^{-2} \sum_i \sum_{j \neq i} u_i X_i' \mu(z_i)^{-1} X_j X_j' (\beta_j - \beta_i) K_{\gamma,ij} \\ &\quad + n^{-2} \sum_i \sum_{j \neq i} u_i X_i' \mu(z_i)^{-1} X_j u_j K_{\gamma,ij} \\ &= F_1 + F_2. \end{aligned}$$

By following exactly the same arguments as we did in the proof of Lemma A.3, one can show that $\mathrm{E}[F_1^2] = n^{-4}O\left(n^3\zeta_n^2 + n^2\zeta_n\left(h_1\ldots h_{q_1}\right)^{-1}\right) = O\left(n^{-1}\zeta_n^2 + n^{-1}\zeta_n\right.$ $\left.\left(nh_1\ldots h_{q_1}\right)^{-1}\right)$. Hence, $F_1 = O_p\left(n^{-1/2}\zeta_n + n^{-1/2}\zeta_n^{1/2}\left(nh_1\ldots h_{q_1}\right)^{-1/2}\right) = o_p$ $\left(\zeta_n^2 + \left(nh_1\ldots h_{q_1}\right)^{-1}\right)$ because $n^{-1/2} = o\left(\zeta_n + \left(nh_1\ldots h_{q_1}\right)^{-1/2}\right)$.

Also, F_2 is a second-order degenerate U-statistic, and so it is fairly straightforward to show that $\mathrm{E}\left(F_2^2\right) = n^{-4}O\left(n^2\left(h_1\ldots h_{q_1}\right)^{-1}\right) = O\left(n^{-1}\left(nh_1\ldots h_{q_1}\right)^{-1}\right)$. Hence, $F_1 = O_p\left(n^{-1/2}\left(nh_1\ldots h_{q_1}\right)^{-1/2}\right) = o_p\left(\left(nh_1\ldots h_{q_1}\right)^{-1}\right)$ because $n^{-1/2} = o\left(\left(nh_1\ldots h_{q_1}\right)^{-1/2}\right)$. Summarizing the preceding discussion, we have shown that

$$CV_{0,2} = CV_{0,2}^0 + (s.o.) = o_p\left(\zeta_n^2 + \left(nh_1\ldots h_{q_1}\right)^{-1}\right). \tag{A.21}$$

By utilizing Rosenthal's and Markov's inequalities, one can show that (A.21) holds uniformly in $(h,\lambda) \in \Gamma$. This completes the proof of Lemma A.4. ■

LEMMA A.5. *Defining* $m(\bar{z}) = \mathrm{E}[X_j X_j' | \bar{z}_j = \bar{z}]\bar{f}(\bar{z})$, $\nu_1(\tilde{z}) = \mathrm{E}[\tilde{K}_{\tilde{y},ij}|\tilde{z}_i = \tilde{z}]$, *and* $\mu(z) = m(\bar{z})\nu_1(\tilde{z})$, *then*

$$\hat{A}(z)^{-1} = \mu(z)^{-1} + O_p\left(|\bar{h}|^2 + |\bar{\lambda}| + \ln^{1/2}(n)(nh_1\ldots h_{q_1})^{-1/2}\right)$$

uniformly in $z \in \mathcal{M}$ *and* $(h,\lambda) \in \Gamma$.

Proof. Defining $\hat{\mu}(z_i) = \mathrm{E}[\hat{A}(z_i)|z_i = z]$, then by the independence of $\tilde{z}_i$ and $(y_i, x_i, \bar{z}_i)$, we have

$$\begin{aligned}\hat{\mu}(z) &= \mathrm{E}\left[X_j X_j' \bar{K}_{\bar{y},ij}|\bar{z}_i = \bar{z}\right]\mathrm{E}\left[\tilde{K}_{\tilde{y},ij}|\tilde{z}_i = \tilde{z}\right] \\ &= \{m(\bar{z}) + O(\zeta_n)\}\mathrm{E}\left[\tilde{K}_{\tilde{y},ij}|\tilde{z}_i = \tilde{z}\right] \\ &= \mu(z) + O_p(\zeta_n),\end{aligned} \tag{A.22}$$

where $\zeta_n = \sum_{s=1}^{q_1} h_s^2 + \sum_{s=1}^{r_1}\lambda_s$. Also, $\hat{A}(z) - \hat{\mu}(z)$ has zero mean. Following standard arguments used when deriving uniform convergence rates for nonparametric kernel estimators (e.g., Masry, 1996), we know that

$$\hat{A}(z) - \hat{\mu}(z) = O_p\left(\frac{(\ln(n))^{1/2}}{(nh_1\ldots h_{q_1})^{1/2}}\right) \tag{A.23}$$

uniformly in $z \in \mathcal{M}$ and $(h,\lambda) \in \Gamma$.

Combining (A.22) and (A.23) we obtain

$$\hat{A}(z) - \mu(z) = O_p\left(\zeta_n + (\ln(n))^{1/2}\left(nh_1\ldots h_{q_1}\right)^{-1/2}\right) \tag{A.24}$$

uniformly in $z \in \mathcal{M}$ ($\mathcal{M}$ is the support of the trimming function $M(\cdot)$) and $(h,\lambda) \in \Gamma$.

Using (A.24) we obtain

$$\hat{A}(z)^{-1} = \left[\mu(z) + \hat{A}(z) - \mu(z)\right]^{-1}$$

$$= \mu(z)^{-1} - \mu(z)^{-1}\left[\hat{A}(z) - \mu(z)\right]\mu(z)^{-1} + O_p\left(\left|\hat{A}(z) - \mu(z)\right|^2\right)$$

$$= \mu(z)^{-1} + O_p\left(\zeta_n + \ln^{1/2}(n)\left(nh_1 \dots h_{q_1}\right)^{-1/2}\right),$$

completing the proof of Lemma A.5. ∎

LEMMA A.6. *$\hat{h}_s = o_p(1)$ for $s = 1, \dots, q_1$ and $\hat{\lambda}_s = o_p(1)$ for $s = 1, \dots, r_1$.*

Proof. Without assuming that any of the smoothing parameters converge to zero, then the only possible non-$o_p(1)$ term in $CV(\gamma)$ is CV_1. It is fairly straightforward to see that $CV_1 = n^{-3}\sum\sum\sum_{l\neq j\neq i} X_i'\hat{\mu}(z_i)^{-1}X_jX_j'(\beta_j - \beta_i)K_{\gamma,ij}X_i'\hat{\mu}(z_i)^{-1}X_lX_l'(\beta_l - \beta_i)K_{\gamma,il} M_i + o_p(1) \equiv G_1 + o_p(1)$, where $\hat{\mu}(z_i) = \mathrm{E}\left[X_jX_j'\bar{K}_{\bar{\gamma},ij}|\bar{z}_i\right]\mathrm{E}\left[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i\right]$ is defined in the proof of Lemma A.5.

Note that G_1 can be written as a third-order U-statistic; hence by the H-decomposition of a U-statistic it is fairly straightforward to show that $G_1 = E(G_1) + o_p(1)$. Furthermore, by the law of iterated expectations we have

$$E(G_1) = \mathrm{E}\left\{\left[X_i'\hat{\mu}(z_i)^{-1}\mathrm{E}\left(X_jX_j'(\beta_j - \beta_i)K_{\gamma,ij}|X_i, z_i\right)\right]^2 M(z_i)\right\}$$

$$= \mathrm{E}\left\{\left\{X_i'[\mathrm{E}\left(X_jX_j'\bar{K}_{\bar{\gamma},ij}|\bar{z}_i\right)]^{-1}\mathrm{E}\left(X_jX_j'(\beta_j - \beta_i)\bar{K}_{\bar{\gamma},ij}|X_i, \bar{z}_i\right)\right\}^2 \bar{M}(\bar{z}_i)\right\}$$

$$= \mathrm{E}\left\{\left\{X_i'[\eta_\beta(\bar{z}_i) - \beta(\bar{z}_i)]\right\}^2 \bar{M}(\bar{z}_i)\right\}$$

$$= \mathrm{E}\left\{(\eta_\beta(\bar{z}_i) - \beta(\bar{z}_i))' E(X_iX_i'|\bar{z}_i)(\eta_\beta(\bar{z}_i) - \beta(\bar{z}_i))\bar{M}(\bar{z}_i)\right\}$$

$$= \int (\eta_\beta(\bar{z}) - \beta(\bar{z}))' m(\bar{z})(\eta_\beta(\bar{z}) - \beta(\bar{z}))\bar{M}(\bar{z})\,d\bar{z}, \tag{A.25}$$

where $m(\bar{z}) = \mathrm{E}(X_iX_i'|\bar{z}_i = \bar{z})\bar{f}(\bar{z})$, $\eta_\beta(\bar{z})$ is defined in equation (16), and $\bar{M}(\bar{z}) = \int \tilde{f}(\tilde{z})M(z)d\tilde{z}$. Note that the right-hand side of (A.25) does not depend on $(h_{q_1+1}, \dots, h_q, \lambda_{r_1+1}, \dots, \lambda_r)$ because $\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i]$ in the numerator cancels with the same quantity in the denominator (from $\hat{\mu}(z_i)^{-1} = \mathrm{E}[X_jX_j'\bar{K}_{\bar{\gamma},ij}|\bar{z}_i]^{-1}\mathrm{E}[\tilde{K}_{\tilde{\gamma},ij}|\tilde{z}_i]^{-1}$).

If the smoothing parameters $h_1, \dots, h_{q_1}, \lambda_1, \dots, \lambda_{r_1}$ that minimize $CV(\gamma)$ do not all converge in probability to zero, then by (17), $\mathrm{E}(G_1)$ (or CV_1) does not converge to zero, which implies the probability that the minimum of CV_1 (over the smoothing parameters) exceeds δ, which does not converge to zero as $n \to \infty$ (for some $\delta > 0$).

However, choosing $h_1, \dots, h_{q_1}$ to be of size $n^{-1/(q_1+4)}$, and $\lambda_1, \dots, \lambda_{r_1}$ to be of size $n^{-2/(q_1+4)}$, letting $h_{q_1+1}, \dots, h_q$ diverge to infinity, and letting $\lambda_{r_1+1}, \dots, \lambda_r$ converge to ones, one can easily show that CV_1 converges in probability to zero. This contradicts the result obtained in the previous paragraph and thus demonstrates that, at the minimum of

$CV(\gamma)$, the smoothing parameters $h_1, \ldots, h_{q_1}, \lambda_1, \ldots, \lambda_{r_1}$, for the relevant components of Z, all converge in probability to zero. ■

APPENDIX B: Proofs of Theorems 4.1 and 4.2

Proof of Theorem 4.1. First, we consider the case where the smoothing parameters are non-stochastic, i.e., $h_s = h_s^0 = a_s^0 n^{-1/(4+q_1)}$ for $s = 1, \ldots, q_1$ and $\lambda_s = \lambda_s^0 = b_s^0 n^{-2/(4+q_1)}$ for $s = 1, \ldots, r_1$. We use $\hat{I}_n^0$ ($\hat{T}_n^0$) to denote the test statistic that uses the smoothing parameter γ^0, where $\gamma^0 = (h_1^0, \ldots, h_{q_1}^0, \lambda_1^0, \ldots, \lambda_{r_1}^0)$. Under H_0 we conduct a Taylor expansion of $\beta_0(z_i, \hat{\alpha}_0)$ at $\hat{\alpha}_0 = \alpha_0$ to get $\beta(z_i, \hat{\alpha}_0) + \beta^{(1)}(z_i, \alpha_0)(\hat{\alpha} - \alpha_0) + O_p\left(n^{-1}\right)$, where $\beta^{(1)}(z, \alpha_0) = [(\partial/\partial\alpha)\beta(z, \alpha)]|_{\alpha=\alpha_0}$.

Then from $\hat{u}_i = Y_i - X_i'\beta(z_i, \hat{\alpha}_0) = Y_i - X_i'\beta(z_i, \alpha_0) - m^{(1)}(z_i, \alpha_0)(\hat{\alpha} - \alpha_0) + O_p(n^{-1})$ and the fact that $\hat{\alpha}_0 - \alpha_0 = O_p(n^{-1/2})$, one can show that the leading term in $\hat{I}_n^0$ is obtained by replacing $\hat{u}_i$ with $Y_i - X_i'\beta(z_i, \alpha_0) = u_i$. We use I_{n1}^0 to denote it, i.e., $I_{n1}^0 = \frac{1}{n^2}\sum_i \sum_{j \neq i} X_i' X_j u_i u_j K_{\gamma^0, ij}$. Here I_{n1}^0 is a second-order degenerate U-statistic. It is straightforward to show that I_{n1}^0 has zero mean and asymptotic variance $(n^2 h_1^0 \ldots h_{q_1}^0)^{-1}$ $(\sigma_0^2 + o(1))$. By the same arguments as in the Hsiao, Li, and Racine (2007), one can show that the conditions for Hall (1984) CLT hold. Therefore, we have $n(h_1^0 \ldots h_{q_1}^0)^{1/2} \hat{I}_n^0 \xrightarrow{d} N(0, \sigma_0^2)$, where $\sigma_0^2 = 2\kappa^{q_1} \mathrm{E}[f(X_i, z_i)(X_i' X_j)^2 \sigma^4(X_i, z_i)]$. It is easy to show that $\hat{\sigma}_0^2$ (with smoothing parameter γ^0) is a consistent estimator of σ_0^2. Hence, we have

$$\hat{T}_n^0 = n(h_1^0 \ldots h_{q_1}^0)^{1/2} \hat{I}_n^0 / \hat{\sigma}_0 \xrightarrow{d} N(0, 1). \quad \textbf{(B.1)}$$

Next, from the result of Theorem 3.2 we know that $\hat{h}_s / h_s^0 - 1 = o_p(1)$ for $s = 1, \ldots, q_1$ and $\hat{\lambda}_s / \lambda_s^0 - 1 = o_p(1)$ for $s = 1, \ldots, r_1$. By stochastic equicontinuity arguments similar to those in Hsiao, Li, and Racine (2007), one can show that

$$\hat{T}_n - \hat{T}_n^0 = o_p(1). \quad \textbf{(B.2)}$$

Equations (B.1) and (B.2) lead to $\hat{T}_n \xrightarrow{d} N(0, 1)$ under H_0, completing the proof of Theorem 4.1. ■

Proof of Theorem 4.2. Because the standard normal random variable's distribution function ($\Phi(\cdot)$) is a continuous distribution, by Polyā's theorem (Bhattacharya and Rao, 1986), we known that (28) is equivalent to the following expression for any given $v \in \mathbb{R}$:

$$\left| P\left(\hat{T}_n^* \leq v | \{Y_i, X_i, Z_i\}_{i=1}^n\right) - \Phi(v) \right| = o_p(1). \quad \textbf{(B.3)}$$

The proof of (B.3) is similar to the proof of Theorem 4.1. Using $\hat{\alpha}_0^* - \hat{\alpha}_0 = O_p(n^{-1/2})$ and a Taylor expansion argument, one can show that the leading term of $\hat{I}_n^*$ is given by

$$I_{n1}^* = \frac{1}{n^2} \sum_i \sum_{j \neq i} X_i' X_j u_i^* u_j^* K_{\hat{\gamma}, ij}.$$

It is easy to show that, conditional on $\{Y_i, X_i, Z_i\}_{i=1}^n$, I_{n1}^* has zero mean and asymptotic variance $(n^2\hat{h}_1 \ldots \hat{h}_{q_1})^{-1}(\hat{\sigma}_0^2 + o_p(1))$. Also, $\hat{\sigma}_0^* = \hat{\sigma}_0 + o_p(1)$. Hence, conditional on $\{Y_i, X_i, Z_i\}$, $T_{n1}^* \equiv n(\hat{h}_1 \ldots \hat{h}_{q_1})^{1/2} I_{n1}^* / \hat{\sigma}_0^*$ has zero mean and asymptotic unit variance. To show that T_{n1}^* has an asymptotic normal distribution conditional on the random sample, we apply the CLT of de Jong (1987) for generalized quadratic forms to derive the asymptotic distribution of $T_{n1}^* | \{Y_i, X_i, Z_i\}_{i=1}^n$. The reason for using the de Jong (1987) CLT instead of the Hall (1984) CLT is that in the bootstrap world, i.e., conditional on the random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, the u_i^* are not identically distributed as their distribution depends on i. Therefore, the de Jong (1987) CLT is the appropriate one to use. By following the same arguments as in the proof of Theorem 2.2 of Hsiao, Li, and Racine (2007), one can show that the sufficient conditions for the de Jong (1987) CLT hold. Hence, (B.3) holds true for T_{n1}^*. It also holds true for T_n^* because $T_n^* = T_{n1}^* + o_p(1)$. ■