Journal of Statistical Planning and Inference 174 (2016) 129–152

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Minimum distance partial linear regression model checking with Berkson measurement errors[☆]

Hira L. Koul, Weixing Song [*]

Michigan State University, United States
Kansas State University, United States

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 18 November 2015
Accepted 19 January 2016
Available online 9 February 2016

MSC:
primary 62G08
secondary 62G20

Keywords:
Partial linear model
Minimum distance
Berkson measurement error
Consistency and local power

ABSTRACT

We propose a class of tests for fitting a parametric model to the nonparametric part in partial linear regression models in the presence of Berkson measurement errors in the covariates. The proposed tests are based on certain minimized L_2 distances between a semi-parametric regression function estimator and the parametric regression model being fitted. We establish asymptotic normality of the proposed test statistics and that of the corresponding minimum distance estimators under the fitted model. The consistency of the tests and their asymptotic power against certain local alternatives are also investigated. Simulation and comparison studies are included to evaluate the finite sample performance of the proposed tests.

1. Introduction

Let $p,\ q$ be known positive integers. Consider the partial linear regression model

$$Y = \beta' X + \nu(T) + \varepsilon, \quad X = Z + \xi,\ T = S + \eta, \tag{1.1}$$

where Y is a scalar response, X and T are, respectively, $p \times 1$ and $q \times 1$ unobservable covariates, β is a p-dimensional unknown regression parameter vector, and ν is an unknown real valued function of the q-variables T with $\nu(T)$ having finite expectation. The observable surrogates Z and S relate to X and T in an additive way with Berkson measurement errors ξ and η, respectively. The random errors $\varepsilon,\ \xi$ and η are assumed to be mutually independent having zero means and finite variances, and are independent of $Z,\ S$. For the sake of the model identifiability, we also assume that the distribution of η is known. As one of the two error structures in measurement error model literature, Berkson measurement modeling is commonly used in health, agriculture and bioassay studies. A typical example is given in Rudemo et al. (1989) where a bioassay experiment with plants is considered. For each of eight herbicides combinations, six nonzero doses were applied to the plant and the dry weight of five plants grown in the same pot was measured. The predictor of interest is the amount of the herbicide absorbed by the plants which is not observable. Instead, the nominal concentration of herbicide applied to the plants was available, and the authors assumed that the true amount was linearly related to the nominal amount, resulting in Berkson measurement error modeling. As an illustrative example, a partial linear regression model with Berkson measurement error is investigated by Wang (2004), where he developed a minimum distance estimation procedure.

☆ Research supported in part by the collaborative grants NSF DMS 1205271, 1205276.

* Corresponding author at: Michigan State University, United States.
E-mail address: weixing@ksu.edu (W. Song).

http://dx.doi.org/10.1016/j.jspi.2016.01.007
0378-3758/

Extensive studies on the classical partial linear regression models, that is, when $\xi \equiv 0$ and $\eta \equiv 0$, or their variants, have been conducted in the past several decades. Engle et al. (1986) were among the first to consider these models when they analyzed the relationship between average daily temperature and electricity usage. Much work has been focused on estimating the unknown regression parameters β and the nonparametric function g. Early important contributions on the estimation theory for this model can be found in Heckman (1986), Rice (1986), Speckman (1988), Robinson (1988), Cuzick (1992), Mammen and van de Geer (1997) and the references therein. A comprehensive overview of the statistical inference for partially linear regression models can be found in the monograph by Härdle et al. (2000). Estimation theories for the parametric and nonparametric Berkson measurement regression models have been developed, see Wang (2003, 2004) and Schennach (2013) and the references therein.

Here, we are interested in developing lack-of-fit tests for checking the adequacy of a parametric form of the nonparametric component ν in the above partial linear regression model (1.1). In the classical regression setup, the lack-of-fit testing problem has been well studied as is evidenced in the monograph of Hart (1997), and the papers of Stute et al. (1998), Stute and Zhu (2002, 2005) and Khmaladze and Koul (2004), and the references therein. Relatively, few works are available for the classical partial linear regression model, and to our best knowledge, there is even less published work available on the model checking procedures in the literature for the above Berkson measurement errors models.

For the classical partial linear regression model, Zhu and Ng (2003) developed a procedure to test the hypothesis $E(Y|X = x, T = t) = \beta' x + \nu(t)$, for some β and ν. Their test is not distribution free, and a variant of wild bootstrap approximation is used to implement their method. Liang (2006) developed two tests based on a residual-marked empirical process and a linear mixed effect framework to check the linearity of the nonparametric component. Again, bootstrap approximation is used to implement the procedure due to the complexity of the asymptotic distribution of the test statistic. Koul and Song (2010) developed a lack of fit test for the nonparametric component in the partial linear regression model (1.1). Their test is based on the supremum of a martingale transform of a certain partial sum process of calibrated residuals. Although the test is superior to some existing counterparts, it is only applicable when the covariate T in the nonparametric part is one-dimensional. In this paper, we remove this unpleasant restriction by basing tests on a class of minimum distance (MD) statistics, inspired by the work of Koul and Ni (2004).

A direct extension of Koul and Ni (2004) testing procedures would consider a transformed regression model based on $E(Y|Z, S) = Z'\beta + E[\nu(T)|S]$. In this case a nonparametric kernel estimator of $E(Y|Z, S)$ is needed to construct the test statistics. Most often T has lower dimension than X, thus the augmented predictors (Z, S) would have a higher dimension. Consequently, we would be dealing with a kernel regression estimator with possibly higher dimensional predictors, and will inevitably fall into the "curse of dimensionality" trap. Also, by doing this, we totally neglect the pre-assumed linear dependence between Y and X, thus artificially making the problem much more complicated. The methodology proposed in this paper allows us to construct a test statistic based only on a regression model with S as the only predictor vector.

The paper is organized as follows. The MD estimators of the unknown parameters under the null hypothesis, and the MD testing procedures are described in Section 2. Technical assumptions and the main results, including the consistency and asymptotic normality of the MD estimators and the asymptotic distribution of the test statistics, together with the consistency and local power discussion, are also described in Section 2. Section 3 contains extensive simulation studies in order to evaluate the finite sample performance of the proposed estimation and testing procedure. All the proofs are deferred to Section 4.

2. Main results

In this section we shall describe assumptions, the class of testing procedure, and the main results. Accordingly, let Θ be a compact subset of $\mathbb{R}^k$ for some integer $k \geq 1$, $m(\cdot; \vartheta)$, $\vartheta \in \Theta$, be a known family of parametric functions, $\mathcal{I}$ be a compact subset of $\mathbb{R}^q$, and consider the problem of testing

$$H_0 : \nu(t) = m(t; \vartheta), \text{ for some } \vartheta \in \Theta, \text{and for all } t \in \mathcal{I}, \quad \text{versus} \quad H_a : H_0 \text{ is not true.}$$

For any r.v. U, let f_U denote its Lebesgue density. Let $\sigma^2 = \mathrm{Var}(\varepsilon)$, $\Sigma_\xi = \mathrm{Cov}(\xi)$, the covariance matrix of ξ, and let $\mu(s) = E[\nu(T)|S = s]$.

The above testing problem is relatively simple when Z and S are independent. In this case $E(Y|S) = \gamma + \mu(S)$, where $\gamma = (EZ)'\beta$. Consider $\gamma + \mu(S)$ as the regression function in Koul and Ni (2004). Then under some regularity conditions, their tests can be directly applied to test for the above H_0. A significant feature in this scenario is that we do not have to estimate β itself, and only the estimate of the scalar γ is needed. Because of the high similarity to Koul and Ni (2004)'s model, we will only state some key theoretical results for this case in Section 2.5 for the sake of completeness, without any proofs. From now on, unless mentioned otherwise, we shall assume that Z and S are dependent. Therefore, $E(Z|S = s)$ is a non-constant function of s.

Let β_0 denote the true value of the parameter β. Then (1.1), together with the mutual independence assumption of ε, ξ, η, and (Z, S), imply

$$E[Y - \beta_0' Z|S = s] = E[\varepsilon + \beta_0'\xi + \nu(T)|S = s] = E[\nu(T)|S = s] = \mu(s).$$

This leads to the calibrated partial linear regression model,

$$Y - \beta_0' Z = \mu(S) + e, \tag{2.1}$$

with $Y - \beta_0' Z$ as a pseudo-response, $e = \varepsilon + \beta_0'\xi + \nu(T) - \mu(S)$, and $E(e|S) = 0$, a.s., so that e is uncorrelated with S. Moreover, the conditional variance of e, given S, is

$$\tau^2(s) = E(e^2|S = s) = \sigma_\varepsilon^2 + \beta_0'\Sigma_\xi\beta_0 + \mathrm{Var}(\nu(T)|S = s). \tag{2.2}$$

The idea of taking $Y - \beta_0' Z$ as a pseudo-response is commonly adopted when estimating ν.

Let $\mu(s;\vartheta) = E(m(T;\vartheta)|S = s) = \int m(s + v;\gamma) f_\eta(v)dv$. From (2.1), it follows that testing for H_0 vs. H_a amounts to testing for

$$\mathcal{H}_0 : \mu(s) = \mu(s;\vartheta),\ \text{for some } \vartheta \in \Theta, \text{and for all } t \in \mathcal{I}, \quad \text{versus} \quad \mathcal{H}_a : \mathcal{H}_0 \text{ is not true.}$$

As described in Koul and Song (2010), the two hypotheses H_0 and $\mathcal{H}_0$ are not equivalent in general. The null hypothesis H_0 clearly implies $\mathcal{H}_0$, but the converse may not be true, since, for any two functions m_1 and m_2, $\int m_1(s + v;\gamma) f_\eta(v)dv = \int m_2(s + v;\vartheta) f_\eta(v)dv$ need not imply $m_1 = m_2$. But if the location family of densities $\{f_\eta(\cdot - s) : s \in \mathbb{R}^q\}$ is complete, then $m_1 = m_2$ holds almost everywhere, and hence $\mathcal{H}_0$ implies H_0, almost everywhere.

To introduce the MD procedures, let K be a q-dimensional symmetric kernel density function around 0 and h, w be two bandwidths depending on the sample size. Let $\theta = (\beta', \vartheta')'$, $\theta_0 = (\beta_0', \vartheta_0')'$ denote the true value of θ under $\mathcal{H}_0$, and for a general bandwidth h, define $K_h(\cdot) = K(\cdot/h)/h^q$. Let G be a σ-finite measure supported on $\mathcal{I}$ and having a continuous Lebesgue density function g. Define, analogous to the definition given in Koul and Ni (2004),

$$\hat f_w(s) = \frac{1}{n}\sum_{i=1}^n K_w(s - S_i), \quad w = w_n \sim (\log(n)/n)^{1/(q+4)}, \tag{2.3}$$

$$M_n(\theta) = \int_{\mathcal{I}} \Big[\frac{1}{n\hat f_w(s)}\sum_{i=1}^n K_h(s - S_i)[Y_i - Z_i'\beta - \mu(S_i;\vartheta)]\Big]^2 dG(s),$$

$$\hat\theta_n = (\hat\beta_n', \hat\vartheta_n')' = \mathrm{argmin}_{\theta\in\mathbb{R}^p\times\Theta} M_n(\theta).$$

Note that under $\mathcal{H}_0$, the ith summand inside the squared integrand of $M_n(\theta_0)$ is conditionally centered, given S_i, for each $i = 1, \ldots, n$. The proposed class of tests, one for each G, is to be based on $M_n(\hat\theta_n)$. The needed assumptions for obtaining asymptotic distributions of $\hat\theta_n$ and $M_n(\hat\theta_n)$ are given in the next subsection.

2.1. Assumptions

Some technical assumptions needed for deriving the consistency and asymptotic normality of the proposed minimum distance tests and estimators are stated in this section. Let $Z(s) = E(Z|S = s)$ and f denote the Lebesgue density of S.

About the random errors, the design variables, and the weighting measure G, we assume the following:

(z). The function $Z(s)$ is continuous in s and the matrix $\mathcal{Z} = \int Z(s)Z'(s)dG(s)$ is positive definite.
(f1). Density function f is uniformly continuous and bounded away from zero on $\mathcal{I}$.
(f2). Density function f is twice continuously differentiable on $\mathcal{I}$.
(g). The integrating measure G has a continuous Lebesgue density g.

About the kernel function K, we assume the following:

(k). The kernel function K is positive symmetric square integrable density on $[-1, 1]^q$ and satisfies a Lipschitz condition.

About the parametric family of functions to be fitted for the nonparametric component in (1.1) we shall assume the following:

(m1). The nonparametric component $\mu(s)$ satisfies $\int \mu^2(s)dG(s) < \infty$, and almost surely, $\mathrm{Var}(\nu(T)|S = s)$ is a continuous function of s on $\mathcal{I}$ w.r.t. G.
(m2). $E(\mu^{2+\delta}(S)) < \infty$, for some $\delta > 0$.
(m3). $E(\mu^4(S)) < \infty$.
(m4). For each ϑ, $\mu(s;\vartheta)$ is a.s. continuous in s w.r.t. the weighting measure G.
(m5). The parametric family $\{H(s;\theta) = Z'(s)\beta - \mu(s;\vartheta) : \beta \in \mathbb{R}^p, \vartheta \in \Theta\}$ is identifiable w.r.t. ϑ and β, i.e., if $H(s;\theta_1) = H(s;\theta_2)$ holds almost everywhere in (s) w.r.t. G, then $\beta_1 = \beta_2$ and $\vartheta_1 = \vartheta_2$.
(m6). For some positive continuous function L on $\mathcal{I}$ and for some $\alpha > 0$,

$$|\mu(s;\vartheta_1) - \mu(s;\vartheta_2)| \le \|\vartheta_1 - \vartheta_2\|^\alpha L(s), \quad \forall\, \vartheta_1, \vartheta_2 \in \Theta,\ s \in \mathcal{I}.$$

(m7). For every s, $\mu(s;\vartheta)$ is differentiable in ϑ with the vector of derivatives $\dot\mu(s,\vartheta)$ satisfying the following. For any consistent estimator of $\hat\vartheta_n$ of ϑ_0,

$$\text{(a)}\quad \max_{1\le i\le n} \frac{\mu(S_i;\hat\vartheta_n) - \mu(S_i;\vartheta_0) - (\hat\vartheta_n - \vartheta_0)'\dot\mu(S_i;\vartheta_0)}{\|\hat\vartheta_n - \vartheta_0\|} = o_p(1),$$

$$\text{(b)}\quad \max_{1\le i\le n} h^{-q/2}\|\dot\mu(S_i;\hat\vartheta_n) - \dot\mu(S_i;\vartheta_0)\| = o_p(1).$$

(m8). Define $Q(s) = \int \dot\mu(t;\vartheta_0)Z'(t)dG(t)\mathcal{Z}^{-1}Z(s)$. Then in a neighborhood of ϑ_0, $\dot\mu(s;\vartheta)$ and $Z(s)$ are linearly independent, and $\int \dot\mu(s;\vartheta_0)(Q(s) - \dot\mu(s;\vartheta_0))'dG(s)$ is nonsingular.

About the bandwidth, we assume that

(h1). $h \to 0$, $nh^{2q} \to \infty$ as $n \to \infty$.
(h2). $h \sim n^{-a}$, where $a < \min\{1/2q, 4/q(q+4)\}$.

Condition (z) is needed to guarantee the uniqueness of the MD estimate for β in the linear term, and it also implies that $E(Z|S=s)$ is a nonconstant function of s. Conditions (m1), (f1), (k), (m4)–(m6), and (h1) are required for the consistency of $\hat{\theta}_n$, but to obtain asymptotic normality for $\hat{\theta}_n$, we also need (m2), (f2), (h2), and (m7). To derive the desired asymptotic distribution for the test statistic, we need all the conditions stated above except for (m2). If Z and S are independent, then condition (z) reduces to $E(ZZ')$ being nonsingular, and the identifiability condition (m5) becomes

(m5)′. The parametric family $\{H(s;\theta) = \gamma - \mu(s;\vartheta) : \gamma \in \mathbb{R}, \vartheta \in \Theta\}$ is identifiable w.r.t. $\theta = (\gamma, \vartheta)$, i.e., if $H(s;\theta_1) = H(s;\theta_2)$ holds almost surely in (s) w.r.t. G, then $\gamma_1 = \gamma_2$ and $\vartheta_1 = \vartheta_2$.

Condition (m8) in fact partly overlaps with the identifiability condition (m5). To see this, let us assume $p=q=k=1$ temporary. If there exists a function $a(\vartheta) \neq 0$ such that $\dot{\mu}(s,\vartheta) = a(\vartheta)Z(s)$, then we can find ϑ_1, ϑ_2 from the neighborhood of ϑ_0, such that $\vartheta_1 \neq \vartheta_2$, and $\mu(s,\vartheta_2) - \mu(s,\vartheta_1) = \int_{\vartheta_1}^{\vartheta_2} a(\vartheta)d\vartheta \cdot Z(s)$. Let $\beta_2 = \int_{\vartheta_1}^{\vartheta_2} a(\vartheta)d\vartheta$, and $\beta_1 = 0$, then $\mu(s,\vartheta_2) - \beta_2 Z(s) = \mu(s,\vartheta_1) - \beta_1 Z(s)$, which contradicts (m5). The requirement of the integral in (m8) not being singular is needed to guarantee the nonsingularity of certain matrix which will be specified later, but the appropriateness of such seemingly strange condition could be justified by noting that $Q(s)$ is a linear combination of $Z(s)$. A statistical model satisfying all of the above conditions can be found in Section 3.

2.2. Consistency and asymptotic distribution of MD estimators

This section states the consistency of $\hat{\theta}_n$. The method of proof here is similar to that of Koul and Ni (2004). A typical application of Lindeberg–Feller central limit theorem shows the asymptotic normality of $\hat{\theta}_n$.

The proof of the consistency of $\hat{\theta}_n$ is facilitated by first proving the consistency of θ_n^* defined as follows. Let

$$M_n^*(\theta) = \int_{\mathcal{I}} \Big[\frac{1}{n\hat{f}_w(s)} \sum_{i=1}^n K_h(s-S_i)(Y_i - Z_i'\beta) - \mu(s;\vartheta)\Big]^2 dG(s),$$
$$\theta_n^* = (\beta_n^{*'}, \vartheta_n^{*'})' = \operatorname{argmin}_{(\beta,\vartheta)\in\mathbb{R}^p\times\Theta} M_n^*(\theta).$$

For the sake of brevity, let $K_{hi}(s) = K_h(s-S_i)$,

$$Z_n(s) = \frac{\sum_{i=1}^n K_{hi}(s)Z_i}{n\hat{f}_w(s)}, \qquad Y_n(s) = \frac{\sum_{i=1}^n K_{hi}(s)Y_i}{n\hat{f}_w(s)}, \qquad \mu_n(s;\vartheta) = \frac{\sum_{i=1}^n K_{hi}(s)\mu(S_i;\vartheta)}{n\hat{f}_w(s)}. \tag{2.4}$$

We shall use the profile MD procedure to derive the MD estimates for θ. Since $\mu(s;\vartheta)$ is differentiable w.r.t. ϑ, θ_n^* satisfies $\partial M_n^*(\theta)/\partial\theta\,|_{\theta=\theta_n^*} = 0$, i.e.,

$$\int \Big[Y_n(s) - Z_n'(s)\beta_n^* - \mu(s;\vartheta_n^*)\Big] \cdot Z_n(s)dG(s) = 0,$$
$$\int \Big[Y_n(s) - Z_n'(s)\beta_n^* - \mu(s;\vartheta_n^*)\Big] \cdot \dot{\mu}(s;\vartheta_n^*)dG(s) = 0.$$

Here, and in the sequel, the integration is understood to be over $\mathcal{I}$. By condition (z), when n is large enough, the matrix $\mathcal{Z}_n := \int Z_n(s)Z_n'(s)dG(s)$ is positive definite in probability. Then from the above equations, for sufficiently large n, we obtain

$$\beta_n^*(\vartheta_n^*) = \mathcal{Z}_n^{-1} \int [Y_n(s) - \mu(s;\vartheta_n^*)]Z_n(s)dG(s),$$
$$\int \Big[Y_n(s) - Z_n'(s)\beta_n^*(\vartheta_n^*)\Big] \cdot \dot{\mu}(s;\vartheta_n^*)dG(s) = \int \mu(s;\vartheta_n^*) \cdot \dot{\mu}(s;\vartheta_n^*)dG(s).$$

Now verify that ϑ_n^* is a minimizer of

$$T_n^*(\vartheta) = \int \Big[\frac{1}{n\hat{f}_w(s)} \sum_{i=1}^n K_h(s-S_i)[Y_i - Z_i'\beta_n^*(\vartheta)] - \mu(s;\vartheta)\Big]^2 dG(s), \tag{2.5}$$

with $\beta_n^*(\vartheta) = \mathcal{Z}_n^{-1}\int [Y_n(s) - \mu(s;\vartheta)]Z_n(s)dG(s)$. Similarly, $\hat{\theta}_n = (\hat{\beta}_n', \hat{\vartheta}_n')'$ is the solution of the following two sets of equations.

$$\hat{\beta}_n(\hat{\vartheta}_n) = \mathcal{Z}_n^{-1} \int [Y_n(s) - \mu_n(s;\hat{\vartheta}_n)]Z_n(s)dG(s), \tag{2.6}$$

$$\int \Big[Y_n(s) - Z_n(s)'\hat{\beta}_n(\hat{\vartheta}_n)\Big] \cdot \dot{\mu}_n(s;\hat{\vartheta}_n)dG(s) = \int \mu_n(s;\hat{\vartheta}_n) \cdot \dot{\mu}_n(s;\hat{\vartheta}_n)dG(s), \tag{2.7}$$

and one can also verify that $\hat{\vartheta}_n$ is a minimizer of

$$T_n(\vartheta) = \int \Big[\frac{1}{n\hat{f}_w(s)} \sum_{i=1}^{n} K_h(s - S_i)[Y_i - Z_i' \hat{\beta}_n(\vartheta) - \mu(S_i; \vartheta)] \Big]^2 dG(s) \tag{2.8}$$

with $\hat{\beta}_n(\vartheta) = \mathcal{Z}_n^{-1} \int [Y_n(s) - \mu_n(s; \vartheta)] Z_n(s) dG(s)$.

The following theorem states the consistency of the MD estimators θ_n^* and $\hat{\theta}_n$ under the null hypothesis.

Theorem 2.1. *Under* H_0, (z), (f1), (k), (m1), (m4)–(m6), (h1), $\theta_n^* \to_p \theta_0$ *and* $\hat{\theta}_n \to_p \theta_0$, *i.e., both* θ_n^* *and* $\hat{\theta}_n$ *are consistent estimators of* θ_0.

To state the asymptotic normality of the MD estimators, the following entities are needed, where $f^{-1} = 1/f$.

$$Q(s) = \int \dot{\mu}(s; \vartheta_0) Z'(s) dG(s) \cdot \mathcal{Z}^{-1} Z(s), \tag{2.9}$$

$$\Sigma = \int \tau^2(s) [\dot{\mu}(s; \vartheta_0) - Q(s)][\dot{\mu}(s; \vartheta_0) - Q(s)]' g^2(s) f^{-1}(s) ds,$$

$$\Sigma_1 = \int \dot{\mu}(s; \vartheta) Q'(s) dG(s), \quad \Sigma_2 = \int \dot{\mu}(s; \vartheta) \dot{\mu}'(s; \vartheta) dG(s),$$

$$M(s) = \mathcal{Z}^{-1} \Big[Z(s) - \int Z(s) \dot{\mu}'(s; \vartheta_0) dG(s) \cdot (\Sigma_2 - \Sigma_1)^{-1} \big(\dot{\mu}(s; \vartheta_0) + Q(s) \big) \Big] f(s),$$

$$\Sigma_3 = \int \tau^2(s) M(s) M'(s) g^2(s) f^{-1}(s) ds,$$

where $\tau^2(s)$ is as in (2.2). By assumption (m8), both Σ and $\Sigma_2 - \Sigma_1$ are positive definite. The asymptotic normality of the MD estimators is summarized in the following theorem.

Theorem 2.2. *Under* H_0, (z), (f1), (f2), (g), (k), (m1), (m2), (m4)–(m8), *and* (h2),

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \to_D N(0, (\Sigma_2 - \Sigma_1)^{-1} \Sigma (\Sigma_2 - \Sigma_1)^{-1}), \quad \sqrt{n}(\hat{\beta}_n - \beta_0) \to_D N(0, \Sigma_3).$$

2.3. Asymptotic distribution of $M_n(\hat{\theta}_n)$

The minimum distance test statistic for H_0 will be built upon $M_n(\hat{\theta}_n)$ defined in (2.3). In this section we describe the asymptotic normality of this statistic more precisely.

Let $e_i = Y_i - Z_i' \beta_0 - \mu(S_i; \vartheta_0)$, $\hat{e}_i = Y_i - Z_i' \hat{\beta}_n - \mu(S_i; \hat{\vartheta}_n)$, and

$$d\phi(s) = dG(s)/f^2(s), \quad d\hat{\phi}_w(s) = dG(s)/\hat{f}_w^2(s), \tag{2.10}$$

$$\tilde{M}_n(\theta_0) = \int \Big[n^{-1} \sum_{i=1}^{n} K_h(s - S_i) e_i \Big]^2 d\phi(s),$$

$$\tilde{C}_n = n^{-2} \sum_{i=1}^{n} \int K_h^2(s - S_i) e_i^2 d\phi(s), \qquad \hat{C}_n = n^{-2} \sum_{i=1}^{n} \int K_h^2(s - S_i) \hat{e}_i^2 d\hat{\phi}_w(s),$$

$$\tilde{\Gamma}_n = 2h^q n^{-2} \sum_{i \neq j} \Big(\int K_h(s - S_i) K_h(s - S_j) e_i e_j d\phi(s) \Big)^2,$$

$$\hat{\Gamma}_n = 2h^q n^{-2} \sum_{i \neq j} \Big(\int K_h(s - S_i) K_h(s - S_j) \hat{e}_i \hat{e}_j d\hat{\phi}_h(s) \Big)^2,$$

$$\Gamma_n = 2h^q (n-1) n^{-1} \iint [E K_h(x - S) K_h(y - S) \tau^2(S)]^2 d\phi(x) d\phi(y),$$

$$\Gamma = 2 \int \tau^4(s) g(s) d\phi(s) \cdot \int \Big[\int K(u) K(u + v) du \Big]^2 dv.$$

Let $T_n(\hat{\theta}_n) = nh^{q/2} \hat{\Gamma}_n^{-1/2} (M_n(\hat{\theta}_n) - \hat{C}_n)$. Then we have the following theorem.

Theorem 2.3. *Suppose* (z), (f1), (f2), (g), (k), (m1), (m3), (m4)–(m8), (h2), *and* H_0 *hold. Then* $T_n(\hat{\theta}_n) \to_D N(0, 1)$.

This theorem readily implies that the test that reject H_0 whenever $\big| T_n(\hat{\theta}_n) \big| \geq z_{\alpha/2}$ is of the asymptotic size α, where z_α is the $100(1 - \alpha)\%$ percentile of the standard normal distribution.

2.4. Consistency and local power of the MD test

A basic requirement for any test procedure is the consistency, i.e., the power of the test at a fixed alternative should tend to 1 as the sample size tends to infinity. This indeed is the case for the proposed tests, under suitable conditions. Let $\ell(t)$ be a measurable function such that $\ell(\cdot) \notin \{m(\cdot;\vartheta) : \vartheta \in \Theta\}$ and $E\ell^2(T) < \infty$. Consider the alternative hypothesis $H_a : \nu(t) = \ell(t),\ t \in \mathcal{I}$. Still denote $\mu(s) = E(\ell(T)|S = s)$. We proved that under H_0, the estimator $\hat\theta_n$ is $\sqrt{n}$-consistent for the true parameter θ_0 and possesses the asymptotic normality. We now need this estimator to have similar properties under H_a. Under some regularity conditions, Jennrich (1969) and White (1981, 1982) showed that the nonlinear least squares estimator in classical regression models converges in probability and is asymptotically normal even in the presence of model misspecification. Let β_a, ϑ_a denote minimizers of $\int[E_{H_a}(Y|S = s) - \beta^T E(Z|S = s) - \mu(s;\vartheta)]^2 dG(s)$ with respect to β, ϑ, respectively, under H_a. After a slight modification of the regularity conditions, similar to the proof of consistency and asymptotic normality of the minimum distance estimators under the null hypothesis, we can show that the MD estimators defined by (2.3) satisfy $\sqrt{n}(\hat\beta_n - \beta_a) = O_p(1)$ and $\sqrt{n}(\hat\vartheta_n - \vartheta_a) = O_p(1)$, under H_a. However, we will not justify this claim rigorously here.

Theorem 2.4. *Suppose all the conditions in Theorem 2.3 hold with ϑ_0 replaced by ϑ_a. If $\int[Z'(s)(\beta_0 - \beta_a) + (\mu(s) - \mu(s;\vartheta_a))]^2 dG(s) \neq 0$, then the above test based on $T_n(\hat\theta_n)$ is consistent for H_a.*

In real applications, it is often desirable to investigate how sensitive the test is to local alternatives. For this purpose, let $\delta(\cdot)$ be a measurable function such that $\delta(\cdot) \notin \{m(\cdot;\vartheta) : \vartheta \in \Theta\}$, $E\delta^2(T) < \infty$, and a_n be a sequence of positive real numbers tending to zero as $n \to \infty$. Consider the local alternatives

$$H_{loc} : \nu(t) = m(t;\theta_0) + a_n\delta(t), \quad t \in \mathcal{I}.$$

Let $L(s) = E[\delta(T)|S = s]$. We shall assume that under H_{loc}, $\sqrt{n}(\hat\beta_n - \beta_0) = O_p(1)$, and $\sqrt{n}(\hat\vartheta_n - \vartheta_0) = O_p(1)$. The following theorem states that the proposed test has nontrivial asymptotic power against a sequence of local alternatives which approaches to the null hypothesis at the rate of $1/\sqrt{nh^{q/2}}$.

Theorem 2.5. *Under the same conditions as in Theorem 2.3 and local alternative hypothesis H_{loc} with $a_n = 1/\sqrt{nh^{q/2}}$, $T_n(\hat\theta_n) \to_D N\Big(\Gamma^{-1/2}\int L^2(s)dG(s), 1\Big)$.*

Optimal G: From the above theorem, we conclude that the asymptotic power of the asymptotic level α test is

$$1 - \Phi\Big(z_{\alpha/2} - \Gamma^{-1/2}\int L^2(s)g(s)ds\Big) + \Phi\Big(-z_{\alpha/2} - \Gamma^{-1/2}\int L^2(s)g(s)ds\Big).$$

An optimal G would maximize this power. Since this power is an increasing function of $\Psi(g) = \Gamma^{-1/2}\int L^2(s)g(s)ds$, the function g that will maximize the power is the one that maximizes $\Psi(g)$. Let $c := 2\int\Big[\int K(u)K(u+v)du\Big]^2 dv$. By the Cauchy–Schwarz inequality,

$$\Psi(g) = \frac{\int L^2(s)g(s)ds}{\sqrt{c\int \tau^4(s)g^2(s)/f^2(s)ds}} \le c^{-1/2}\Big(\int \frac{L^4(s)f^2(s)}{\tau^2(s)}ds\Big)^{1/2},$$

with equality holding, if and only if, $g(s) \propto L^2(s)f^2(s)/\tau^4(s)$, for all s. Since $\Psi(ag) = \Psi(g)$ for all $a > 0$, we may take the optimal weight function g to be $g(s) = L^2(s)f^2(s)/\tau^4(s)$. Clearly this g is unknown because of f, β_0, ϑ_0, but one can estimate it by $g_n(s)$, the analogue of $g(s)$ where these unknown parameters and function are replaced by their estimators.

2.5. Asymptotic results for independent Z and S

As mentioned in Introduction, when Z and S are independent, $E(Y|Z,S) = Z'\beta + E(\nu(T)|S)$. This relatively simple structure of the regression function avoids the need to estimate β and enables us to directly adopt the methodology developed in Koul and Ni (2004) to test the null hypothesis. Let $(\hat\gamma_n, \hat\vartheta_n) = \text{argmin}_{\theta=\{\gamma,\vartheta\}\in\mathbb{R}\times\Theta} M_n(\theta)$, where

$$M_n(\theta) = \int\left[\frac{1}{n}\sum_{i=1}^n K_h(s - S_i)[Y_i - \gamma - \mu(S_i;\vartheta)]\right]^2 d\hat\phi_w(s).$$

The asymptotic normality of $(\hat\gamma, \hat\vartheta)$ is summarized in the following theorem.

Theorem 2.6. *Under* H_0, *if* (f1), (f2), (g), (k), (m1), (m2), (m4), (m5)′, (m6), (m7), *and* (h2) *hold, then*

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0, (\hat{\vartheta}_n - \vartheta_0)')' \to_D N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1}),$$

where

$$\Sigma_0 = \int \begin{bmatrix} 1 & \dot{\mu}'(s;\vartheta_0) \\ \dot{\mu}'(s;\vartheta_0) & \dot{\mu}(s;\vartheta_0)\dot{\mu}'(s;\vartheta_0) \end{bmatrix} dG(s),$$

$$\Sigma = \int \begin{bmatrix} 1 & \dot{\mu}'(s;\vartheta_0) \\ \dot{\mu}(s;\vartheta_0) & \dot{\mu}(s;\vartheta_0)\dot{\mu}'(s;\vartheta_0) \end{bmatrix} \frac{\sigma^2(s)g^2(s)}{f(s)} ds,$$

and $\sigma^2(s) = E(Y - \gamma_0 - \mu(S;\vartheta_0)^2|S = s)$.

Let $\widetilde{e}_i = Y_i - \hat{\gamma}_n - \mu(S_i;\hat{\vartheta}_n)$. Let $\widetilde{C}_n, \widetilde{\Gamma}_n$ denote the $\hat{C}_n, \hat{\Gamma}_n$ defined in (2.10) after replacing $\hat{e}_i$ there by $\widetilde{e}_i$. Then we have the following theorem.

Theorem 2.7. *Suppose* (f1), (f2), (g), (k), (m1), (m3), (m4), (m5)′, (m6), (m7), (h2), *and* H_0 *hold. Then* $T_n(\hat{\theta}_n) = nh^{q/2}\widetilde{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \widetilde{C}_n) \to_D N(0, 1)$.

3. Numerical studies

Extensive simulation studies are conducted in this section to evaluate the finite sample performance of some members of the proposed class of tests. Z and S are assumed to be dependent in the first simulation. In the second simulation, optimal weight discussed in Section 2.4 and non-optimal weight on the performance of the proposed test are compared for some local models, while Z and S are assumed to be independent. Finally, a comparison study is conducted between Koul and Ni (2004)'s test and the proposed test in the third simulation.

Simulation 1: The data are generated from the following four models:

$$\begin{aligned}
\text{M0:}\quad & Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + e,\\
\text{M1:}\quad & Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + 2\sin(T_1) + 2\sin((T_2 + 0.5)^2) + e,\\
\text{M2:}\quad & Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + 2(T_1 - 0.5)(T_2 + 0.5)^2 + e,\\
\text{M3:}\quad & Y = X_1 + X_2 + T_1(T_2 + 0.5)^2 I_{(T_2 > -0.3)} + e.
\end{aligned}$$

The unobserved X_1, X_2, T_1, T_2 and their surrogates are generated from the Berkson measurement error model $X_1 = Z_1 + \xi_1, X_2 = Z_2 + \xi_2, T_1 = S_1 + \eta_1, T_2 = S_2 + \eta_2$, and

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \qquad \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}\right).$$

$$\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \qquad \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}\right),$$

and $[Z_1, Z_2]' = [0.5S_1 + U_1, 0.5S_2 + U_2]'$. The random error e is generated from the standard normal distribution. The compact set $\mathcal{I}$ is chosen to be $[-3, 3] \times [-3, 3]$, which covers the majority values of (S_1, S_2), and G is chosen to be the uniform distribution function over $\mathcal{I}$. We can check that all the assumptions stated in Section 2.1 hold for this model. All integrals are evaluated by Riemann sum with 200 equally spaced subintervals in each direction. The nominal level is taken to be 0.05. For each sample size, the simulation is repeated 200 times, and the empirical sizes and powers in each scenario are calculated by $\#\{|T_n| \geq z_{0.975}\}/200$.

The data from the model M0 are used to study the empirical level, while the data from models M1 to M3 are used to study the empirical power of the test. Note that the theory of the present paper is not applicable to Model M3, which is included here to see the effect of the discontinuity in the regression function on the power of the proposed test. Under the null hypothesis, $\nu(t) = t_1 + 2(t_2 + 0.5)^2$, and $\beta_0 = (1, 1)'$, $\vartheta_0 = (1, 2)'$, $p = q = 2$. The bivariate Epanechnikov kernel $K(u)K(v)$ with $K(u) = 3(1 - u^2)I(|u| \leq 1)/4$ is used for smoothing, and the bandwidth h is chosen to be $h = an^{-1/4.5}$ for $a = 0.5, 0.8, 1$. The main purpose of choosing various a values is to check the sensitivity of the proposed MD test to different smoothness. The bandwidth w is chosen to be $(\log(n)/n)^{1/6}$.

Table 1 reports the finite sample levels and powers of the proposed MD test. The empirical levels appear slightly unstable, and for all the chosen alternatives, the power approaches unity as the sample size increases. From the simulation, we see that the values of a do have an effect on the finite sample performance of the proposed tests, and larger values of a, such as 0.8, 1, may be preferable.

Table 2 reports the means and mean square errors (MSE) of the MD estimates according to the values of a. The results are very promising. All means are very close to the true values, and as expected, the MSE's decrease with increasing sample size. Moreover, the choice of a does not affect the bias of these estimates, especially for the samples sizes of 200 and larger.

Table 1
Empirical levels and powers (Simulation 1).

	$a = 0.5$				$a = 0.8$				$a = 1.0$			
	100	200	300	500	100	200	300	500	100	200	300	500
M0	0.030	0.045	0.050	0.070	0.035	0.045	0.030	0.060	0.030	0.045	0.035	0.065
M1	0.050	0.090	0.160	0.275	0.055	0.110	0.200	0.390	0.050	0.145	0.210	0.455
M2	0.145	0.380	0.615	0.870	0.225	0.560	0.835	0.970	0.270	0.635	0.900	0.985
M3	0.295	0.595	0.820	0.985	0.390	0.830	0.965	1.000	0.440	0.900	0.985	1.000

Table 2
Mean and MSE of MD estimates in model M0 (Simulation 1).

n	a	β_1	β_2	ϑ_1	ϑ_2
	0.5	0.9978 (0.1271)	0.9689 (0.1231)	1.0110 (0.0968)	2.0162 (0.0385)
100	0.8	1.0322 (0.2363)	0.9579 (0.2026)	0.9972 (0.1238)	2.0152 (0.0385)
	1.0	1.0635 (0.3280)	0.9526 (0.2559)	0.9839 (0.1452)	2.0156 (0.0396)
	0.5	0.9854 (0.0888)	1.0263 (0.0664)	1.0088 (0.0515)	2.0073 (0.0217)
200	0.8	0.9810 (0.1581)	1.0261 (0.1063)	1.0103 (0.0700)	2.0108 (0.0236)
	1.0	0.9709 (0.2193)	1.0283 (0.1292)	1.0147 (0.0862)	2.0124 (0.0243)
	0.5	1.0261 (0.0574)	0.9890 (0.0522)	0.9738 (0.0351)	2.0100 (0.0169)
300	0.8	1.0315 (0.1177)	0.9856 (0.0786)	0.9699 (0.0483)	2.0108 (0.0179)
	1.0	1.0312 (0.1785)	0.9866 (0.0906)	0.9700 (0.0618)	2.0112 (0.0182)
	0.5	0.9999 (0.0434)	0.9804 (0.0353)	1.0037 (0.0202)	2.0050 (0.0075)
500	0.8	0.9887 (0.0911)	0.9829 (0.0530)	1.0074 (0.0312)	2.0035 (0.0077)
	1.0	0.9848 (0.1316)	0.9835 (0.0625)	1.0082 (0.0401)	2.0028 (0.0078)

For illustration, the histograms of 200 MD estimates with sample size $n = 300$ are also drawn in Fig. 1. The solid lines are the kernel density estimates for the distribution of the MD estimates using the default bandwidth in R function, and the dashed lines are normal density curves with means and variances calculated from the 200 estimates. Clearly, Fig. 1 indicates that the MD estimates do follow the asymptotic normal distribution.

Simulation 2: In this simulation study, we investigate the performance of the proposed test when optimal weights defined in Section 2.4 are used for testing the local models. To be specific, the following models are used to generate the data:

$$\text{M0:}\quad Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + e,$$
$$\text{M1:}\quad Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + 2d[\sin(T_1) + \sin(T_2 + 0.5)]/\sqrt{nh} + e,$$
$$\text{M2:}\quad Y = X_1 + X_2 + T_1 + 2(T_2 + 0.5)^2 + 2d(T_1 - 0.5)(T_2 + 0.5)^2/\sqrt{nh} + e,$$

where d is a positive number which will be specified later. Here the null hypothesis is $H_0 : \nu(t) = \gamma_1 t_1 + \gamma_2(t_2 + 0.5)^2$. The distributional setup is the same as in the first simulation except for now we generate (Z_1, Z_2) from a bivariate standard normal distribution, which is independent of (S_1, S_2). Data form model M0 were used to study the empirical levels, while the data from M1 and M2 were used to study the empirical powers of the tests. It is easy to see that for M1, $\delta(T) = 2d(T_1 - 0.5)(T_2 + 0.5)^2$, so $L(s) = E(\delta(T)|S_1 = s_1, S_2 = s_2) = 2d(s_1 - 0.5)(s_2^2 + 0.75)$. For M2, $\delta(T) = 2d[\sin(T_1) + \sin(T_2 + 0.5)]$, and $L(s) = E(\delta(T)|S_1 = s_1, S_2 = s_2) = 2d[\sin(s_1) + \sin(s_2 + 0.5)]E\cos(\eta_1)$. We can also obtain that $\tau^2(s) = \sigma_\varepsilon^2 + \beta^T\Sigma_\xi\beta + 4\text{Var}(\eta_2^2) + \text{Var}(\eta_1) + 4(4s_2^2 + 1)\text{Var}(\eta_2)$, therefore, a consistent estimate of $\tau^2(s)$ can be defined as $\hat{\tau}^2(s) = n^{-1}\sum_{i=1}^n \hat{e}_i^2 + 4\left(s_2^2 + n^{-1}\sum_{i=1}^n S_{2i}^2\right)$. Let $\hat{f}_w(s)$ be an estimate of the density function of (S_1, S_2) with bandwidth chosen according to (2.3). Thus the optimal weight functions for M1 and M2 can be defined according to the formula $g(s) = L^2(s)\hat{f}_w^2(s)/\hat{\tau}^4(s)$. For the sake of simplicity, in the simulation, the performance of the proposed test with the optimal weight will be compared with the one with weight function $g(s) = \hat{f}_w^2(s)$. d is set to be 1–10, the sample size n is chosen to be 200. The simulation result is summarized in Table 3 with $h = an^{-1/4.5}$ and $a = 0.5$ being selected as the bandwidth.

Other values for a and n are also checked in the simulation, and similar patterns are obtained, hence omitted here for brevity. It is seen from Table 3 that the test with optimal weight performs better than the one with non-optimal weight for M1 for all d values, while the simulation results for M2 are mixed for small d values. When d gets larger, however, the test with the optimal weight outperforms the one with non-optimal weight.

Simulation 3: In this simulation study, we compare the proposed test with the MD test constructed in Koul and Ni (2004) (KN), and the test constructed in Koul and Song (2010) (KS). Note that the KN test involves multiple numerical integration with number of folds being equal to the number of covariates from both linear part and nonlinear part in the model, so the computation is very complicated and time consuming. To expedite the simulation process, we consider the following simple

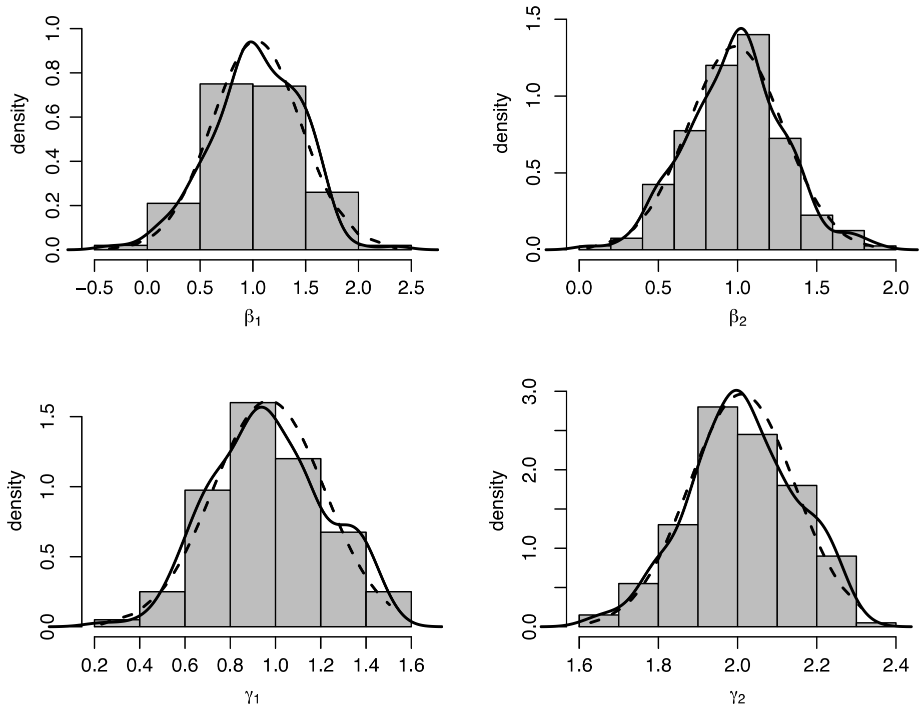

Fig. 1. Distributions of MD estimates in model M0 (Simulation 1).

partial linear regression models in which both X and T are one dimensional.

$$\text{M0:} \quad Y = X + 2T + e,$$
$$\text{M1:} \quad Y = X + 2T + T^2 + e,$$
$$\text{M2:} \quad Y = X + 2T + 2\sin(T) + e.$$

In the simulation, ε, Z, S are independently generated from $N(0, 1)$, ξ and η are independently generated from $N(0, 0.5^2)$. For the proposed method, the kernel function is chosen to be the Epanechnikov kernel $K(u) = 3(1 - u^2)I(|u| \le 1)/4$, $h = 0.5n^{-1/5}$, the compact set $\mathcal{I} = [-3, 3]$, and for the KN test, the kernel function is chosen to be $K(u)K(v)$, $h = 0.5n^{-1/4.5}$, the compact set $\mathcal{I} = [-3, 3] \times [-3, 3]$. The KS test is based on the supremum of a martingale transformation of a partial sum process of calibrated residuals which is only applicable to the cases where T is one-dimensional.

The simulation result is presented in Table 4. It shows that the KN test is more liberal than the proposed test, while the KS test keeps the nominal level very well except for the case of $n = 500$. It is clear that the proposed test outperforms both KN and KS tests for the selected alternative models and the KS test performs poorly for M2.

Remark. Bandwidth selection for estimation in nonparametric smoothing has been thoroughly discussed in the literature, but how to choose an "optimal" bandwidth in the context of hypothesis testing remains an open question. As evidenced in Simulation 1, the bandwidth h has little effect when estimating the regression coefficients, but does affect the finite sample performance of the tests, so does the choice of w. From Section 2.4, we know that the optimal weight function is given by $g(s) = L^2(s)f^2(s)/\tau^4(s)$. Hence, to avoid the complexity of selection of two bandwidths simultaneously, an ad-hoc method is using $\hat{f}_w(s)$ as the estimate of the density function of S in $g(s)$, as we did in Simulation 2.

4. Proofs of the main results

This section contains the proofs of the results stated in Section 2. The basic ideas of the proofs are similar to those appearing in Koul and Ni (2004) but details are necessarily different and more involved.

The proof of the consistency of the MD estimates is facilitated by the following lemma, which along with its proof appears as Theorem 2.2 part (2) in Bosq (1998).

Lemma 4.1. *Let $\hat{f}_w$ be the density kernel estimate with a kernel K satisfying a Lipschitz condition and bandwidth $w = w_n = a_n(\log n/n)^{1/(d+4)}$, where $a_n \to a_0 > 0$. Then under condition* (f2),

$$(\log_k n)^{-1}(n/\log n)^{2/(d+4)} \sup_{s \in I} |\hat{f}_w(s) - f(s)| \to 0, \quad a.s.$$

for all integers $k > 0$.

Table 3
Empirical levels and powers w/o optimal weight (Simulation 2).

d	$g(s)=\hat{f}_w(s)$			$g(s)=L^2(s)\hat{f}_w^2(s)/\hat{\tau}^4(s)$		
	M0	M1	M2	M0	M1	M2
1	0.03	0.025	0.035	0.035	0.030	0.030
2	0.03	0.025	0.075	0.035	0.040	0.095
3	0.03	0.030	0.145	0.035	0.045	0.140
4	0.03	0.055	0.265	0.035	0.065	0.260
5	0.03	0.070	0.350	0.035	0.110	0.385
6	0.03	0.090	0.430	0.035	0.160	0.470
7	0.03	0.160	0.485	0.035	0.220	0.565
8	0.03	0.245	0.550	0.035	0.305	0.615
9	0.03	0.280	0.570	0.035	0.360	0.635
10	0.03	0.325	0.585	0.035	0.470	0.655

Table 4
Comparison between the proposed, KN and KS tests (Simulation 3).

	The proposed test			KN test			KS test		
	M0	M1	M2	M0	M1	M2	M0	M1	M2
100	0.020	0.995	0.450	0.045	0.710	0.009	0.045	0.775	0.030
200	0.040	1.000	0.810	0.065	0.975	0.220	0.050	0.995	0.045
300	0.035	1.000	0.950	0.055	1.000	0.350	0.045	1.000	0.040
500	0.045	1.000	1.000	0.060	1.000	0.650	0.035	1.000	0.085

We begin with the proof of the consistency of the MD estimates.

Proof of the consistency of θ_n^*: Let $Y(s)=E(Y|S=s)$. Recall from Section 2 $Z(s)=E(Z|S=s)$, $\mathcal{Z}=\int Z(s)Z'(s)dG(s)$, and let

$$M^*(\theta)=\int[Y(s)-Z'(s)\beta-\mu(s;\vartheta)]^2dG(s),\qquad \theta^*=\mathrm{argmin}_{\theta\in\mathbb{R}\times\Theta}M^*(\theta).$$

By (m5), the unique minimizer of $M^*(\theta)$ is $\beta=\beta_0$ and $\vartheta=\vartheta_0$.

Setting the first derivatives of $M^*(\theta)$ w.r.t. β and ϑ equal to 0 yields

$$\frac{\partial M^*(\theta)}{\partial\beta}=\int[Y(s)-Z'(s)\beta-\mu(s;\vartheta)]Z(s)dG(s)=0, \tag{4.1}$$

$$\frac{\partial M^*(\theta)}{\partial\vartheta}=\int[Y(s)-Z'(s)\beta-\mu(s;\vartheta)]\dot{\mu}(s;\vartheta)dG(s)=0. \tag{4.2}$$

For any fixed ϑ, solve (4.1) for β to obtain

$$\beta^*(\vartheta)=\mathcal{Z}^{-1}\int[Y(s)-\mu(s;\vartheta)]Z(s)dG(s). \tag{4.3}$$

Plugging the solution into (4.2), we obtain

$$\int[Y(s)-Z'(s)\beta^*(\vartheta)-\mu(s;\vartheta)]\dot{\mu}(s;\vartheta)dG(s)=0. \tag{4.4}$$

Let $\tilde{\vartheta}$ denote the solution of (4.4) and $\tilde{\beta}=\beta^*(\tilde{\vartheta})$. In fact, $\tilde{\vartheta}=\mathrm{argmin}_{\vartheta\in\Theta}T^*(\vartheta)$, where $T^*(\vartheta)=\int[Y(s)-Z'(s)\beta^*(\vartheta)-\mu(s;\vartheta)]^2dG(s)$. Clearly, under the null hypothesis, the unique minimizer of $T^*(\vartheta)$ is $\tilde{\vartheta}=\vartheta_0$, and $\tilde{\beta}=\beta_0$.

To show the consistency of θ_n^*, we shall first show that ϑ_n^*, as the solution of (2.5) is consistent. Note that ϑ_n^* is the minimizer of $T_n^*(\vartheta)$, so to show the consistency, it suffices to show

$$\sup_{\vartheta\in\Theta}|T_n^*(\vartheta)-T^*(\vartheta)|=o_p(1). \tag{4.5}$$

In fact, (4.5) implies that $T_n^*(\vartheta_n^*)-T^*(\vartheta_n^*)=o_p(1)$, $T_n^*(\vartheta_0)-T^*(\vartheta_0)=o_p(1)$. Thus

$$T_n^*(\vartheta_n^*)-T_n^*(\vartheta_0)=T^*(\vartheta_n^*)-T^*(\vartheta_0)+o_p(1). \tag{4.6}$$

By the definition of ϑ_n^* and ϑ_0, the left-hand side of (4.6) is nonpositive, while the first term on the right-hand side of (4.6) is nonnegative. Hence $T^*(\vartheta_n^*)-T^*(\vartheta_0)=o_p(1)$. If ϑ_n^* does not converge to ϑ_0 in probability, then by the compactness of Θ and $\vartheta_n\in\Theta$, there must exist a subsequence $\vartheta_{n_k}^*$ of ϑ_n^* such that $\vartheta_{n_k}^*\to\vartheta_1$ in probability, $\vartheta_1\in\Theta$ and $\vartheta_1\neq\vartheta_0$. The continuity of $T^*(\vartheta)$ implies that $T^*(\vartheta_{n_k}^*)-T^*(\vartheta_1)=o_p(1)$, and therefore $T^*(\vartheta_1)=T^*(\vartheta_0)$. This contradicts the uniqueness of the minimizer of $T^*(\vartheta)$ over Θ.

To prove (4.5), note that $T_n^*(\vartheta) - T^*(\vartheta)$ can be rewritten as

$$\int \Big([Y_n(s) - Z_n'(s)\beta_n(\vartheta) - \mu(s;\vartheta)]^2 - [Y(s) - Z'(s)\beta(\vartheta) - \mu(s;\vartheta)]^2 \Big) dG(s)$$
$$= \int \Big\{ [Y_n(s) - Y(s)] - [Z_n'(s)\beta_n(\vartheta) - Z'(s)\beta(\vartheta)] \Big\}$$
$$\cdot \Big\{ Y_n(s) - Z_n'(s)\beta_n(\vartheta) - \mu(s;\vartheta) + Y(s) - Z'(s)\beta(\vartheta) - \mu(s;\vartheta) \Big\} dG(s).$$

By the Cauchy–Schwarz inequality, and the inequality $(\sum_{j=1}^m |a_j|)^r \le m^{r-1}\sum_{j=1}^m |a_j|^r$, for all $r \ge 1$, and all positive integers m, $|T_n^*(\vartheta) - T^*(\vartheta)| \le (2A_{1n} + 2A_{2n})^{1/2}(3B_{1n} + 3B_{2n} + 12B_3)^{1/2}$, where

$$A_{1n} = \int [Y_n(s) - Y(s)]^2 dG(s), \qquad A_{2n} = \int [Z_n'(s)\beta_n(\vartheta) - Z'(s)\beta(\vartheta)]^2 dG(s),$$
$$B_{1n} = \int [Y_n(s) + Y(s)]^2 dG(s), \qquad B_{2n} = \int [Z_n'(s)\beta_n(\vartheta) + Z'(s)\beta(\vartheta)]^2 dG(s),$$

and $B_3 = \int \mu^2(s;\vartheta)dG(s)$. Bound A_{1n} above by the sum of the following four terms:

$$A_{1n1} = \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i - Y(s)f(s) \Big]^2 \Big| \frac{1}{\hat f_w^2(s)} - \frac{1}{f^2(s)} \Big| dG(s),$$
$$A_{1n2} = \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i - Y(s)f(s) \Big]^2 d\phi(s),$$
$$A_{1n3} = \int \Big[(\hat f_w(s) - f(s))Y(s) \Big]^2 \Big| \frac{1}{\hat f_w^2(s)} - \frac{1}{f^2(s)} \Big| dG(s),$$
$$A_{1n4} = \int \Big[(\hat f_w(s) - f(s))Y(s) \Big]^2 d\phi(s).$$

Adding and subtracting $EK_{h1}(s)Y_1$ from $n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i - Y(s)f(s)$, we see that A_{1n2} is bounded above by

$$2\int \Big[n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i - EK_{h1}(s)Y_1 \Big]^2 d\phi(s) + 2\int \Big[EK_{h1}(s)Y_1 - Y(s)f(s) \Big]^2 d\phi(s).$$

Note that the expected value of the first term equals

$$2\int \mathrm{Var}\Big(n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i \Big) d\phi(s) \le \frac{C}{n}\int E[K_{h1}^2(s)Y_1^2]dG(s) = O\Big(\frac{1}{nh^q}\Big) = o(1).$$

Also, by the continuity assumption on $Y(s)$ and $f(s)$,

$$\int \big[EK_{h1}(s)Y_1 - Y(s)f(s) \big]^2 d\phi(s) = \int \big[E\{K_{h1}(s)E(Y_1|S=s)\} - Y(s)f(s) \big]^2 d\phi(s) = o(1).$$

Therefore, $A_{1n2} = o_p(1)$. By Lemma 4.1 we have

$$A_{1n1} \le \sup_{s\in I}\Big| \frac{1}{\hat f_w^2(s)} - \frac{1}{f^2(s)} \Big| \cdot \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(s)Y_i - Y(s)f(s) \Big]^2 dG(s) = o_p(1).$$

Similarly, one can show that $A_{1n3} = o_p(1) = A_{1n4}$. Therefore, $A_{1n} = o_p(1)$.

To deal with A_{2n}, consider first the difference $\beta_n^*(\vartheta) - \beta(\vartheta)$. With $\mathcal{Z}_n = \int Z_n(s)Z_n'(s)dG(s)$,

$$\beta_n^*(\vartheta) - \beta(\vartheta) = (\mathcal{Z}_n^{-1} - \mathcal{Z}^{-1})\int [Y_n(s) - \mu(s;\vartheta)]Z_n(s)dG(s) + \mathcal{Z}^{-1}\int [Y_n(s) - Y(s)]Z_n(s)dG(s).$$

Since $\mathcal{Z}_n \to_p \mathcal{Z}$ and $\mathcal{Z}$ is positive definite, we have $\mathcal{Z}_n^{-1} \to_p \mathcal{Z}^{-1}$. Also note that by the Cauchy–Schwarz inequality,

$$\Big| \int [Y_n(s) - Y(s)]Z_n(s)dG(s) \Big|^2 \le \int [Y_n(s) - Y(s)]^2 dG(s) \cdot \int Z_n^2(s)dG(s) = o_p(1).$$

The boundedness of $\mu(s;\vartheta)$ over $I \times \Gamma$ also implies that

$$\Big| \int [Y_n(s) - \mu(s;\vartheta)]Z_n(s)dG(s) \Big|^2 \le \int [Y_n(s) - \mu(s;\vartheta)]^2 dG(s) \cdot \int Z_n^2(s)dG(s) = O_p(1)$$

uniformly in $\vartheta \in \Theta$. Therefore, $\sup_{\vartheta\in\Theta} |\beta_n^*(\vartheta) - \beta(\vartheta)| = o_p(1)$. By this fact, we have

$$A_{2n} \le \int Z_n^2(s)dG(s)\|\beta_n^*(\vartheta) - \beta(\vartheta)\|^2 + \beta(\vartheta)\int [Z_n(s) - Z(s)]^2 dG(s) = o_p(1)$$

uniformly in $\vartheta \in \Theta$.

Similarly, one can show that $B_{1n} = O_p(1)$, $B_{2n} = O_p(1)$ and $B_3 = O(1)$. In summary, we finally obtain (4.5), hence the consistency of ϑ_n^*. Furthermore,

$$\beta_n^*(\vartheta_n^*) = \mathcal{Z}_n^{-1}\int (Y_n(s) - \mu(s;\vartheta_0))Z_n(s)dG(s) - \mathcal{Z}_n^{-1}\int (\mu(s;\vartheta_n^*) - \mu(s;\vartheta_0))Z_n(s)dG(s),$$

and the continuity and boundedness of $\mu(s;\vartheta)$ implies the second term converges to 0 in probability, and the first term converges to $\mathcal{Z}^{-1}\int (Y(s) - \mu(s;\vartheta_0))dG(s)$ in probability, which equals β_0.

To show the consistency of $\hat\theta_n$, we can first show the consistency of $\hat\vartheta_n$ using a similar argument as in the proof of Theorem 3.1 of Koul and Ni (2004) by showing $\sup_{\vartheta\in\Theta} |T_n(\vartheta) - T_n^*(\vartheta)| = o_p(1)$, then the consistency of $\hat\beta_n$ follows from the formula $\hat\beta_n = \hat\beta_n(\hat\vartheta_n)$. Due to the complex nature of the current setup, the argument is more tedious and omitted here for the sake of brevity. □

Proof of Theorem 2.2. Rewrite (2.7) as

$$\int Y_n(s)\dot\mu_n(s;\hat\vartheta_n)dG(s) = \int \Big[Z_n'(s)\hat\beta_n(\hat\vartheta_n) + \mu_n(s;\hat\vartheta_n)\Big]\dot\mu_n(s;\hat\vartheta_n)dG(s).$$

Subtracting and adding $\int [Z_n'(s)\hat\beta_n(\vartheta_0) + \mu_n(s;\vartheta_0)]\dot\mu_n(s;\hat\vartheta_n)dG(s)$ in the above expression, and after a slight arrangement of the terms, we obtain

$$\begin{aligned}&\int \Big[Y_n(s) - Z_n'(s)\hat\beta_n(\vartheta_0) - \mu_n(s;\vartheta_0)\Big]\cdot\dot\mu_n(s;\hat\vartheta_n)dG(s)\\ &\quad = \int Z_n'(s)[\hat\beta_n(\hat\vartheta_n) - \hat\beta_n(\vartheta_0)]\cdot\dot\mu_n(s;\hat\vartheta_n)dG(s) + \int [\mu_n(s;\hat\vartheta_n) - \mu_n(s;\vartheta_0)]\cdot\dot\mu_n(s;\hat\vartheta_n)dG(s). \end{aligned} \tag{4.7}$$

Rewrite the left hand side of (4.7) as the sum of the following two terms

$$S_{n1} = \int \Big[Y_n(s) - Z_n'(s)\beta_0 - \mu_n(s;\vartheta_0)\Big]\cdot\dot\mu_n(s;\hat\vartheta_n)dG(s);$$

$$S_{n2} = -\int \dot\mu_n(s;\hat\vartheta_n)Z_n'(s)dG(s)\cdot[\hat\beta_n(\vartheta_0) - \beta_0].$$

Next we shall show that $\sqrt{n}(S_{n1} + S_{n2})$ is asymptotically normally distributed. For this purpose, rewrite S_{n1} as

$$S_{n1} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)[Y_i - Z_i'\beta_0 - \mu(S_i;\vartheta_0)]\Big]\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu(S_i;\hat\vartheta_n)d\hat\phi(s).$$

Note that under H_0, $Y_i - Z_i'\beta_0 - \mu(S_i;\vartheta_0) = e_i$. Let

$$U_n(s) = \frac{1}{n}\sum_{i=1}^n K_{hi}(s)e_i, \qquad \eta_n(s;\hat\vartheta_n) = \frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu(S_i;\hat\vartheta_n),$$

$$\eta_n(s;\vartheta_0) = \frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu(S_i;\vartheta_0), \quad \eta_h(s) = EK_h(s-S)\dot\mu(S;\vartheta_0).$$

Then S_{n1} can be written as the sum of the following six terms

$$S_{n11} = \int U_n(s)\eta_h(s)d\phi(s), \quad S_{n12} = \int U_n(s)\eta_h(s)[\hat f_w^{-2}(s) - f^{-2}(s)]dG(s),$$

$$S_{n13} = \int U_n(s)[\eta_n(s;\hat\vartheta_n) - \eta_n(s;\vartheta_0)]d\phi(s),$$

$$S_{n14} = \int U_n(s)[\eta_n(s;\hat\vartheta_n) - \eta_n(s;\vartheta_0)][\hat f_w^{-2}(s) - f^{-2}(s)]d\phi(s),$$

$$S_{n15} = \int U_n(s)[\eta_n(s;\vartheta_0) - \eta_h(s)]d\phi(s),$$

$$S_{n16} = \int U_n(s)[\eta_n(s;\vartheta_0) - \eta_h(s)][\hat f_w^{-2}(s) - f^{-2}(s)]d\phi(s).$$

On the other hand, by definition (2.6), under the null hypothesis, $\hat{\beta}_n(\vartheta_0)$ can be written as

$$\hat{\beta}_n(\vartheta_0) = \beta_0 + \mathcal{Z}_n^{-1} \int U_n(s)\hat{f}_w^{-1}(s)\, Z_n(s)\, dG(s). \tag{4.8}$$

Then we verify that

$$\sqrt{n}S_{n2} = -\int \dot{\mu}(s;\vartheta_0)Z'(s)dG(s) \cdot \mathcal{Z}^{-1} \cdot \int U_n(s)Z(s)f^{-1}(s)dG(s) + o_p(1).$$

Recalling the notation $Q(s)$ from (2.9), we obtain $\sqrt{n}S_{n2} = -\sqrt{n}\int U_n(s)Q(s)f(s)d\phi(s) + o_p(1)$.

Upon combining the above derivations, we obtain $\sqrt{n}(S_{n11}+S_{n2}) = \sum_{i=1}^n \zeta_{ni}/\sqrt{n} + o_p(1)$, where $\zeta_{ni} = \int K_{hi}(s)e_i[\eta_h(s) - Q(s)f(s)]d\phi(s)$. For convenience, we shall prove the desired result only for $m = 1$. Hence $\eta_h(s)$ and $Q(s)$ are one dimensional. For $m > 1$, the result can be proved by Wald scheme in which the same arguments are applied to the linear combination of its components instead of $\eta_h(s)$ and $Q(s)$ themselves.

Since ζ_{ni}, $i = 1, 2, \ldots, n$, are i.i.d. r.v.'s with mean 0, so to apply Lindeberg–Feller central limit theorem, it suffices to show that for any $\lambda > 0$, $E\zeta_{n1}^2 \to \Sigma$, and $E\zeta_{n1}^2 I(|\zeta_{n1}| > \lambda\sqrt{n}) \to 0$.

By the Fubini theorem,

$$\begin{aligned} E\zeta_{n1}^2 &= \iint EK_h(u-S)K_h(v-S)\tau^2(S)[\eta_h(u) - Q(u)f(u)][\eta_h(v) - Q(v)f(v)]d\phi(u)d\phi(v) \\ &= \iiint K_h(u-s)K_h(v-s)\tau^2(s)[\eta_h(u) - Q(u)f(u)][\eta_h(v) - Q(v)f(v)]f(s)d\phi(u)d\phi(v)ds. \end{aligned}$$

By changing variables with $u-s = xh$, $v-s = yh$, $s = s$, and using the assumed continuity of τ^2, f, g and $\dot{\mu}(s;\vartheta_0)$, we obtain $\lim_{h\to 0} E\zeta_{n1}^2 = \Sigma$. By Hölder's inequality and the continuity of $E(e^{2+\delta}|S=s)$, $\dot{\mu}(s;\vartheta_0)$, for any $\delta > 0$, $E\zeta_{n1}^2 I(|\zeta_{n1}| > \lambda\sqrt{n})$ is bounded above by

$$\begin{aligned} n^{-\delta/2}\lambda^{-\delta}E\zeta_{n1}^{2+\delta} &= n^{-\delta/2}\lambda^{-\delta}E\Big[\int K_{h1}(s)e_1(\eta_h(s) - Q(s)f(s))d\phi(s)\Big]^{2+\delta} \\ &\le n^{-\delta/2}\lambda^{-\delta}E\Big[\Big(\int [K_h(s-S)(\eta_h(s) - Q(s)f(s))]^{1+\delta/2}d\phi(s)\Big)^2 E(e_1^{2+\delta}|S=s)\Big] \\ &= O((nh^{2q})^{-\delta/1}) = o(1). \end{aligned}$$

Therefore, we have

$$\sqrt{n}(S_{n11} + S_{n2}) \to_D N(0, \Sigma).$$

Next, we shall show that $\sqrt{n}S_{n1k} = o_p(1)$ for all $k = 2, \ldots, 6$. For $k = 2$, by Lemma 4.1, nS_{n12}^2 is bounded above by

$$n\int U_n^2(s)\eta_h^2(s)d\phi(s) \cdot \sup_{s\in I} |f^2(s)/\hat{f}_w^2(s) - 1|^2 = O_p\Big(\frac{1}{h^q}\Big) \cdot O_p\Big(\frac{(\log_k n)^2(\log n)^{4/(q+4)}}{n^{4/(q+4)}}\Big),$$

which is the order of $o_p(1)$ by assumption $h \sim n^{-a}$ and $a < \min(1/2q, 4/(q(q+4)))$.

By assumption (m7(b)), $\sqrt{n}|S_{n13}|$ is bounded above by

$$\begin{aligned} &\sqrt{n}\sup_{1\le i\le n} |\hat{\mu}(S_i;\hat{\vartheta}_n) - \hat{\mu}(S_i;\vartheta_0)| \cdot \Big[\int U_n^2(s)d\phi(s)\Big]^{1/2}\Big[\int \frac{1}{n}\sum_{i=1}^n K_{hi}^2(s)d\phi(s)\Big]^{1/2} \\ &= \sqrt{n}\cdot o_p(h^{q/2}) \cdot O_p\left(\frac{1}{\sqrt{nh^q}}\right) \cdot O_p(1) = o_p(1). \end{aligned}$$

This, together with Lemma 4.1, implies $\sqrt{n}S_{n14} = o_p(1)$.

By the Cauchy–Schwarz inequality,

$$n\|S_{n15}\|^2 \le n\int U_n^2(s)d\phi(s) \cdot \int (\eta_n(s;\vartheta_0) - \eta_h(s))^2 d\phi(s).$$

Direct calculations show that $E\int U_n^2(s)d\phi(s) = O(1/nh^q)$. For the second factor, first note that $\eta_n(s;\vartheta_0) - \eta_h(s)$ can be written as an average of centered i.i.d. r.v.'s. By the Fubini theorem, and the fact that the variance is bounded above by the second moment, we obtain that

$$E\int (\eta_n(s;\vartheta_0) - \eta_h(s))^2 d\phi(s) \le n^{-1}\int EK_h^2(s-S)\dot{\mu}^2(s;\vartheta_0)d\phi(s) = O(1/nh^q).$$

Hence $n\|S_{n15}\|^2 = n \cdot O_p(1/n^2h^{2q}) = o_p(1)$. This, together with Lemma 4.1, also implies that $\sqrt{n}S_{n16} = o_p(1)$. Thus, upon combining all of the above results, we obtain

$$\sqrt{n}(S_{n1} + S_{n2}) \to_D N(0, \Sigma). \tag{4.9}$$

We now continue the proof of Theorem 2.2. First, note that from (2.6),

$$\hat{\beta}_n(\hat{\vartheta}_n) - \hat{\beta}_n(\vartheta_0) = -\mathcal{Z}_n^{-1} \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s)\{\mu(S_i; \hat{\vartheta}_n) - \mu(S_i; \vartheta_0)\} \cdot \hat{f}_w(s) Z_n(s) d\hat{\phi}_w(s).$$

Then $\hat{\beta}_n(\hat{\vartheta}_n) - \hat{\beta}_n(\vartheta_0)$ can be written as the sum of $-\mathcal{Z}^{-1}\int Z(s)\dot{\mu}'(s; \vartheta_0)dG(s)(\hat{\vartheta}_n - \vartheta_0)$ and other 15 remaining terms. We show that all the terms are either of the order of $(\hat{\vartheta}_n - \vartheta_0) \cdot o_p(1)$ or $o_p(1/\sqrt{n})$. For example, with $d_{ni} = \mu(S_i; \hat{\vartheta}_n) - \mu(S_i; \vartheta_0) - \dot{\mu}'(S_i; \vartheta_0)(\hat{\vartheta}_n)$, a typical term is

$$\mathcal{Z}^{-1} \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) d_{ni} Z(s) f(s) d\phi(s) = \mathcal{Z}^{-1} \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \frac{d_{ni}}{\|\hat{\vartheta}_n - \vartheta_0\|} Z(s) f(s) d\phi(s) \frac{(\vartheta_n - \vartheta_0)'}{\|\hat{\vartheta}_n - \vartheta_0\|} (\hat{\vartheta}_n - \vartheta_0).$$

By assumption (m7), the coefficient of $(\hat{\vartheta}_n - \vartheta_0)$ in the above expression is bounded above by

$$\sup_{1\le i\le n} \frac{|d_{ni}|}{\|\hat{\vartheta}_n - \vartheta_0\|} \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \|\mathcal{Z}^{-1} Z(s)\| f(s) d\phi(s) = o_p(1).$$

Therefore, the first term on the right hand side of (4.7) multiplied by $\sqrt{n}$, can be written as

$$-\int \dot{\mu}_n(s; \hat{\vartheta}_n) Z_n'(s) dG(s) \cdot \mathcal{Z}^{-1} \int Z(s)\dot{\mu}'(s; \vartheta_0) dG(s) \cdot \sqrt{n}(\hat{\vartheta}_n - \vartheta_0)[1 + o_p(1)].$$

By the continuity of $\dot{\mu}(s; \vartheta)$, $Z(s)$, and the consistency of $\hat{\vartheta}_n$,

$$\int \dot{\mu}_n(s; \hat{\vartheta}_n) Z_n'(s) dG(s) = \int \dot{\mu}(s; \vartheta_0) Z'(s) dG(s) + o_p(1).$$

Therefore, the first term on the right hand side of (4.7) multiplied by $\sqrt{n}$, can be further written as $-\Sigma_1\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)[1 + o_p(1)]$.

The second term on the right of side of (4.7) can be written as the sum of the following four terms,

$$Q_{1n} = \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) d_{ni} \cdot \frac{1}{n} \sum_{i=1}^n K_{hi}(s)[\dot{\mu}(S_i; \hat{\vartheta}_n) - \dot{\mu}(S_i; \vartheta_0)] d\hat{\phi}_w(s);$$

$$Q_{2n} = \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) d_{ni} \cdot \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \dot{\mu}(S_i; \vartheta_0) d\hat{\phi}_w(s);$$

$$Q_{3n} = \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s)[\dot{\mu}(S_i; \hat{\vartheta}_n) - \dot{\mu}(S_i; \vartheta_0)] \cdot \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \dot{\mu}'(S_i; \vartheta_0) d\hat{\phi}_w(s)(\hat{\vartheta}_n - \vartheta_0);$$

$$Q_{4n} = \int \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \dot{\mu}(S_i; \vartheta_0) \cdot \frac{1}{n} \sum_{i=1}^n K_{hi}(s) \dot{\mu}'(S_i; \vartheta_0) d\hat{\phi}_w(s)(\hat{\vartheta}_n - \vartheta_0).$$

By conditions in (m7), we can show that $\sqrt{n}Q_{1n}$, $\sqrt{n}Q_{2n}$ and $\sqrt{n}Q_{3n}$ all equals $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \cdot o_p(1)$. By the continuity of $\dot{\mu}(s; \vartheta)$, we also have

$$\sqrt{n}Q_{4n} = \int \dot{\mu}(s; \vartheta_0)\dot{\mu}'(s; \vartheta_0) dG(s) \sqrt{n}(\hat{\vartheta}_n - \vartheta_0)[1 + o_p(1)].$$

Therefore, the second term on the right hand side of (4.7) multiplied by $\sqrt{n}$, can be further written as $\Sigma_2\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)[1 + o_p(1)]$.

Summarizing all of the above derivations and using the assumption that $\Sigma_1 + \Sigma_2$ is positive definite, we obtain that

$$\sqrt{n}(\vartheta_n - \vartheta_0) = (\Sigma_2 - \Sigma_1)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int K_{hi} e_i [\eta_h(s) + Q(s) f(s)] d\phi(s) + o_p(1),$$

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \to_D N(0, (\Sigma_2 - \Sigma_1)^{-1} \Sigma (\Sigma_2 - \Sigma_1)^{-1}).$$

Next, we prove the second part of the claim in Theorem 2.2. By (2.6), $\hat{\beta}_n - \beta_0$ can be written as

$$\hat{\beta}_n(\hat{\vartheta}_n) - \beta_0 = \mathcal{Z}_n^{-1} \int U_n(s)\hat{f}_w^{-1}(s)Z_n(s)dG(s) - \mathcal{Z}_n^{-1} \int \frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)]\hat{f}_w^{-1}(s)Z_n(s)dG(s).$$

By assumption (m7(b)) and Lemma 4.1, one can show that

$$\hat{\beta}_n(\hat{\vartheta}_n) - \beta_0 = \mathcal{Z}^{-1} \int U_n(s)Z(s)f^{-1}(s)dG(s) - \mathcal{Z}^{-1} \int Z(s)\dot{\mu}'(s;\vartheta_0)dG(s)(\hat{\vartheta}_n - \vartheta_0) + o_p(n^{-1/2}).$$

This fact and the definition of $M(s)$ in (2.9) yield

$$\sqrt{n}(\hat{\beta}_n(\hat{\vartheta}_n) - \beta_0) = n^{1/2} \int U_n(s)M(s)d\phi(s) + o_p(1).$$

This approximation and argument like the one used in the proof of the asymptotic normality of $\sqrt{n}(S_{n11} + S_{n2})$, together yield that $\sqrt{n}(\hat{\beta}_n(\hat{\vartheta}_n) - \beta_0) \to_D N(0, \Sigma_3)$. $\square$

Theorem 2.3 will be proved after we establish the following five lemmas. The proof of Lemma 4.3 is facilitated by Theorem 1 of Hall (1984) which is reproduced here for the sake of completeness.

Lemma 4.2. *Let X_i, $1 \le i \le n$ be i.i.d. random vectors, and let*

$$U_n = \sum_{1\le i<j\le n} H_n(X_i, X_j), \qquad G_n(x,y) = EH_n(X_1,x)H_n(X_1,y),$$

where H_n is a sequence of measurable functions symmetric under permutation, with

$$E[H_n(X_1,X_2)|X_1] = 0, \text{ a.s.} \quad \text{and} \quad EH_n^2(X_1,X_2) < \infty, \quad \text{for each } n \ge 1.$$

If $[EG_n^2(X_1,X_2)+n^{-1}EH_n^4(X_1,X_2)]/[EH_n^2(X_1,X_2)]^2 \to 0$, then U_n is asymptotically normally distributed with mean 0 and variance $n^2EH_n^2(X_1,X_2)/2$.

We begin with the statement and the proof of the first lemma.

Lemma 4.3. *Under H_0, suppose that* (z), (f1), (g), (k), (m1), (m3) *hold. Then*

$$nh^{q/2}[\tilde{M}_n(\theta_0) - \tilde{C}_n] \to_D N(0, \Gamma).$$

Proof of Lemma 4.3. By the definition of $\tilde{M}_n(\theta_0)$ and $\tilde{C}_n$, we have

$$nh^{q/2}[\tilde{M}_n(\theta_0) - \tilde{C}_n] = \frac{1}{n^2}\sum_{i\ne j} \int K_h(s-S_i)K_h(s-S_j)e_ie_jd\phi(s).$$

Denote $\tilde{S}_i = (S_i', e_i)'$ and

$$H_n(\tilde{S}_i, \tilde{S}_j) = \frac{h^{q/2}}{n} \int K_h(s-S_i)K_h(s-S_j)e_ie_jd\phi(s).$$

Then $nh^{q/2}[\tilde{M}_n(\theta_0) - \tilde{C}_n] = 2\sum_{1\le i<j\le n} H_n(\tilde{S}_i, \tilde{S}_j)$. Note that $H_n(\tilde{S}_1, \tilde{S}_2)$ is symmetric in its arguments, and $E(H_n(\tilde{S}_1, \tilde{S}_2)|\tilde{S}_1) = 0$ by the fact that $E(e_2|\tilde{S}_1) = 0$. To apply Lemma 4.2, we have to compute the quantities $EH_n^2(\tilde{S}_1, \tilde{S}_2)$, $EH_n^4(\tilde{S}_1, \tilde{S}_2)$, and $EG_n^2(\tilde{S}_1, \tilde{S}_2)$.

First, note that, changing variables,

$$\begin{aligned} EH_n^2(\tilde{S}_1, \tilde{S}_2) &= \frac{h^q}{n^2} \iint [EK_h(s-S_1)K_h(t-S_1)\tau^2(S_1)]^2 d\phi(s)d\phi(t) \\ &= \frac{h^q}{n^2} \iint \Big[\int K_h(s-u)K_h(t-u)\tau^2(u)f(u)du\Big]^2 d\phi(s)d\phi(t) \\ &= \frac{1}{n^2} \iint \Big[\int K(u)K(v+u)\tau^2(t-vh-uh)f(t-vh-uh)du\Big]^2 \cdot g(t-vh)f^{-2}(t-vh)dvd\phi(t). \end{aligned}$$

By the continuity of $\mathrm{Var}(g(T;\vartheta_0)|S=s)$, f and g, we obtain

$$EH_n^2(\tilde{S}_1, \tilde{S}_2) = \frac{1}{n^2} \int \Big[\int K(u)K(v+u)du\Big]^2 dv \cdot \int \tau^4(t)g(t)d\phi(t) + o\Big(\frac{1}{n^2}\Big). \tag{4.10}$$

Similarly, $EH_n^4(\tilde{S}_1, \tilde{S}_2)$ equals

$$\frac{h^{2q}}{n^4} \iiiint [EK_h(s-S_1)K_h(t-S_1)K_h(u-S_1)K_h(v-S_1)\tau^4(S_1)]^2 d\phi(s)d\phi(t)\phi(u)d\phi(v),$$

and after changing variables, by the continuity of $\mathrm{Var}(g(T;\vartheta_0)|S=s), f$ and g, we have

$$EH_n^4(\tilde{S}_1, \tilde{S}_2) = O\Big(\frac{1}{n^4 h^q}\Big). \tag{4.11}$$

Now let us consider $EG_n^2(\tilde{S}_1, \tilde{S}_2)$. Denote $d\phi_{xyuv} = d\phi(x)d\phi(y)d\phi(u)d\phi(v)$ for the sake of brevity, and

$$B_h(z-x) = h^{-d}\int K(u)K((z-x)/h+u)\tau^2(x-uh)f(x-uh)du.$$

Then we can show that

$$EG_n^2(\tilde{S}_1, \tilde{S}_2) = \frac{h^{2q}}{n^4}\iiiint B_h(x-u)B_h(y-x)B_h(y-v)B_h(v-u)d\phi_{xyuv} = O\Big(\frac{h^q}{n^4}\Big). \tag{4.12}$$

Therefore, from (4.10)–(4.12), we have

$$\frac{EG_n^2(\tilde{S}_1, \tilde{S}_2) + n^{-1}EH_n^4(\tilde{S}_1, \tilde{S}_2)}{[EH_n^2(\tilde{S}_1, \tilde{S}_2)]^2} = \frac{O(h^q/n^4) + n^{-1}\cdot O(1/n^4h^q)}{O(1/n^4)} = o(1)$$

because of $nh^q \to \infty$. Also, by (4.10), as $n \to \infty$,

$$n^2 EH_n^2(\tilde{S}_1, \tilde{S}_2)/2 \to \frac{1}{2}\int\Big[\int K(u)K(v+u)du\Big]^2 dv \cdot \int \tau^4(t)g(t)d\phi(t) + o(1). \tag{4.13}$$

Hence, by Lemma 4.2, we finally get

$$nh^{q/2}[\tilde{M}_n(\theta_0) - \tilde{C}_n] = 2\sum_{1\le i<j\le n} H_n(\tilde{S}_i, \tilde{S}_j) \to_D N(0, \Gamma),$$

as claimed. □

Lemma 4.4. *Under H_0, suppose* (z), (f1), (f2), (k), (m1), (m2), (m6)–(m7), (h2) *hold. Then* $nh^{q/2}[M_n(\theta_0) - \tilde{M}_n(\theta_0)] = o_p(1)$.

Proof of Lemma 4.4. In fact, by the definition of $M_n(\theta)$ and Lemma 4.1,

$$\begin{aligned} nh^{q/2}|M_n(\theta_0) - \tilde{M}_n(\theta_0)| &\le nh^{q/2}\int U_n^2(s)d\phi(s)\cdot \sup_{s\in I}\Big|\frac{f^2(s)}{\hat{f}_w^2(s)} - 1\Big| \\ &= nh^{q/2}\cdot O_p(1/nh^q)\cdot O_p\Big(\frac{(\log_k n)(\log n)^{2/(q+4)}}{n^{2/(q+4)}}\Big) = o_p(1). \end{aligned}$$

Lemma 4.5. *Under H_0, suppose* (z), (f1), (k), (m1), (m2), (m6)–(m7), (h2) *hold. Then* $nh^{q/2}[M_n(\hat{\theta}_n) - M_n(\theta_0)] = o_p(1)$.

Proof of Lemma 4.5. Adding and subtracting $Z_i'\beta_0 + \mu(S_i;\vartheta_0)$ from $\hat{e}_i = Y_i - Z_i'\hat{\beta}_n - \mu(S_i;\hat{\vartheta}_n)$, we can write $\hat{e}_i = e_i - Z_i'(\hat{\beta}_n - \beta_0) - (\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0))$. Then the difference $M_n(\hat{\theta}_n) - M_n(\theta_0)$ can be written as the sum of the following five terms

$$B_{n1} = \int\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i'(\hat{\beta}_n - \beta_0)\Big]^2 d\hat{\phi}_w(s),$$

$$B_{n2} = \int\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)]\Big]^2 d\hat{\phi}_w(s),$$

$$B_{n3} = -2\int \frac{1}{n}\sum_{i=1}^n K_{hi}(s)e_i \cdot \frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i' d\hat{\phi}_w(s)(\hat{\beta}_n - \beta_0),$$

$$B_{n4} = -2\int \frac{1}{n}\sum_{i=1}^n K_{hi}(s)e_i \cdot \frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)]d\hat{\phi}_w(s),$$

$$B_{n5} = 2\int \frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)]\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i' d\hat{\phi}_w(s)(\hat{\beta}_n - \beta_0).$$

For B_{n1}, by the $\sqrt{n}$-consistency of $\hat\beta_n$, we have

$$nh^{q/2}B_{n1} \le nh^{q/2}\|\hat\beta_n-\beta_0\|^2\int Z_n^2(s)dG(s)=O_p(h^{q/2})=o_p(1).$$

For B_{n2}, by (m7) and $\sqrt{n}$-consistency of $\hat\vartheta_n$, we have

$$\begin{aligned}B_{n2} &\le 2\int\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)d_{ni}\Big]^2 d\hat\phi_w(s)+2\int\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu'(S_i;\vartheta_0)(\hat\vartheta_n-\vartheta_0)\Big]^2 d\hat\phi_w(s)\\ &=\sup_{1\le i\le n}\frac{|d_{ni}^2|}{\|\hat\vartheta_n-\vartheta_0\|^2}\cdot O_p(1)\cdot\|\hat\vartheta_n-\vartheta_0\|^2+O_p(1)\cdot\|\hat\vartheta_n-\vartheta_0\|.\end{aligned}$$

Thus, $nh^{q/2}B_{b2}=o_p(h^{q/2})+O_p(h^{q/2})=o_p(1)$. Similarly, one can show that $nh^{q/2}B_{n5}=o_p(1)$.

Now let us consider B_{n3}. Denote $V_n(s)=n^{-1}\sum_{i=1}^n K_{hi}(s)Z_i$, and recall the definition of $U_n(s)$, we can rewrite B_{n3} as

$$\begin{aligned}B_{n3} &= -2\int U_n(s)[V_n(s)-EV_n(s)][1/\hat f_w^2(s)-1/f^2(s)]dG(s)(\hat\beta_n-\beta_0)-2\int U_n(s)[V_n(s)-EV_n(s)]d\phi(s)(\hat\beta_n-\beta_0)\\ &\quad -2\int U_n(s)EV_n(s)[1/\hat f_w^2(s)-1/f^2(s)]dG(s)(\hat\beta_n-\beta_0)-2\int U_n(s)EV_n(s)d\phi(s)(\hat\beta_n-\beta_0).\end{aligned}$$

We will only show that

$$nh^{q/2}\int U_n(s)EV_n(s)d\phi(s)(\hat\beta_n-\beta_0)=o_p(1). \tag{4.14}$$

Other three terms can be dealt with similarly. In fact, $\int U_n(s)EV_n(s)d\phi(s)(\hat\beta_n-\beta_0)$ can be written as

$$\int U_n(s)[EV_n'(s)-Z'(s)f(s)]d\phi(s)(\hat\beta_n-\beta_0)+\int U_n(s)Z'(s)f(s)d\phi(s)(\hat\beta_n-\beta_0).$$

Denote the first term as B_{n31}, the second term as B_{n32}, we have

$$nh^{q/2}|B_{n31}| \le nh^{q/2}\Big(\int U_n^2(s)d\phi(s)\cdot\int[EV_n'(s)-Z'(s)f(s)]^2d\phi(s)\Big)^{1/2}|\hat\beta_n-\beta_0|=o_p(1).$$

Similar to the argument as in showing the asymptotic normality of $\sqrt{n}(S_{n11}+S_{n2})$ in the previous section, we can show that $\int U_n(s)Z'(s)f(s)d\phi(s)=O_p(1/\sqrt{n})$, hence $nh^{q/2}B_{n32}=nh^{q/2}O_p(1/n)=O_p(h^{q/2})=o_p(1)$. Therefore, that (4.14) is justified, hence $nh^{q/2}B_{n3}=o_p(1)$.

Finally, let us consider B_{n4}. Recall the notation d_{ni}, we can rewrite B_{n4} as the sum of the following four terms

$$\begin{aligned}B_{n41} &= -2\int U_n(s)\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)d_{ni}\cdot[1/\hat f_w^2(s)-1/f^2(s)]dG(s),\\ B_{n42} &= -2\int U_n(s)\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)d_{ni}d\phi(s),\\ B_{n43} &= -2\int U_n(s)\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu'(S_i;\vartheta_0)\cdot[1/\hat f_w^2(s)-1/f^2(s)]dG(s)(\hat\vartheta_n-\vartheta_0),\\ B_{n44} &= -2\int U_n(s)\cdot\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\dot\mu'(S_i;\vartheta_0)d\phi(s)(\hat\vartheta_n-\vartheta_0).\end{aligned}$$

We shall only show $nh^{q/2}B_{n42}=o_p(1)$. Other cases can be dealt with similarly. In fact, by the Cauchy–Schwarz inequality, $nh^{q/2}B_{n42}|$ is bounded above by

$$2nh^{q/2}\sup_{1\le i\le n}\frac{|d_{ni}|}{\|\hat\vartheta_n-\vartheta_0\|}\Big(\int U_n^2(x)d\phi(s)\cdot\int\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Big]^2d\phi(s)\Big)^{1/2}\|\hat\vartheta_n-\vartheta_0\|.$$

By assumption (m7), the $\sqrt{n}$-consistency of $\hat\vartheta_n$, we have

$$nh^{q/2}B_{n42}=nh^{q/2}\cdot o_p(1)\cdot O_p(1/\sqrt{nh^q})\cdot O_p(1/\sqrt{n})=o_p(1).$$

Therefore, $nh^{q/2}B_{n4}=o_p(1)$ as claimed. This finishes the proof of the lemma. □

Lemma 4.6. *Under the same conditions as in Lemma 4.4,* $nh^{q/2}(\hat{C}_n - \tilde{C}_n) = o_p(1)$.

Proof of Lemma 4.6. Again using the fact that $\hat{e}_i = e_i - Z_i'(\hat{\beta}_n - \beta_0) - (\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_0))$, we can rewrite $\hat{C}_n$ as the sum of the following six terms:

$$C_{n1} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)e_i^2 d\hat{\phi}_w(s),$$

$$C_{n2} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)[\mu(S_i,\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)]^2 d\hat{\phi}_w(s),$$

$$C_{n3} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)[Z_i'(\hat{\beta}_n - \beta_0)]^2 d\hat{\phi}_w(s),$$

$$C_{n4} = -\frac{2}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)e_i Z_i' d\hat{\phi}_w(s)\cdot(\hat{\beta}_n - \beta_0),$$

$$C_{n5} = -\frac{2}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)e_i[\mu(S_i,\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)] d\hat{\phi}_w(s),$$

$$C_{n6} = -\frac{2}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)Z_i'[\mu(S_i,\hat{\vartheta}_n) - \mu(S_i;\vartheta_0)] d\hat{\phi}_w(s)\cdot(\hat{\beta}_n - \beta_0).$$

While C_{n1} can be further written as the sum

$$C_{n1} = \tilde{C}_n + \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)e_i^2[1/\hat{f}_w^2(s) - 1/f^2(s)]dG(s).$$

Note that by the continuity of $\tau^2(s), f(s)$ and $g(s)$, we have

$$E\Big[\frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)e_i^2 dG(s)\Big] = \frac{1}{n}\iint \frac{1}{h^{2q}}K^2\Big(\frac{s-u}{h}\Big)\tau^2(u)f(u)dudG(s) = O\Big(\frac{1}{nh^q}\Big),$$

so by Lemma 4.1,

$$nh^{q/2}|C_{n1} - \tilde{C}_n| \le nh^{q/2}\sup_{x\in I}\Big|\frac{1}{\hat{f}_w^2(s)} - \frac{1}{f^2(s)}\Big|\cdot O_p\Big(\frac{1}{nh^q}\Big) = o_p(1).$$

Using the notation d_{ni}, C_{n2} is bounded above by the sum $2C_{n21} + 2C_{n22}$, where

$$C_{n21} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)d_{ni}^2 d\hat{\phi}_w(s) \qquad C_{n22} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)[\dot{\mu}'(S_i;\vartheta_0)(\hat{\vartheta}_n - \vartheta_0)]^2 d\hat{\phi}_w(s).$$

By (m7), and the fact that for any continuous function $L(s)$,

$$\frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)L(S_i)d\hat{\phi}_w(s) = O_p\Big(\frac{1}{nh^q}\Big), \tag{4.15}$$

we have, by the $\sqrt{n}$-consistency of $\hat{\vartheta}_n$,

$$\begin{aligned} nh^{q/2}C_{n21} &\le nh^{q/2}\max_{1\le i\le n}\frac{d_{ni}^2}{\|\hat{\vartheta}_n - \vartheta_0\|^2}\cdot\frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)d\hat{\phi}_w(s)\cdot\|\hat{\vartheta}_n - \vartheta_0\|^2 \\ &= nh^{q/2}\cdot o_p(1)\cdot O_p\Big(\frac{1}{nh^q}\Big)\cdot O_p\Big(\frac{1}{n}\Big) = o_p(1), \end{aligned}$$

and

$$nh^{q/2}C_{n22} \le nh^{q/2}\cdot\frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)\|\dot{\mu}(S_i;\vartheta_0)\|^2 d\hat{\phi}_w(s)\cdot\|\hat{\vartheta}_n - \vartheta_0\|^2 = o_p(1).$$

Similarly, one can show $nh^{q/2}C_{n3} = o_p(1)$ by (4.15) and the $\sqrt{n}$-consistency of $\hat{\vartheta}_n$.

By the continuity of $E(|e|\,|S=s)$, we can show that $nh^{q/2}C_{n4}$ is bounded above by

$$nh^{q/2}\cdot\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(s)|e_i|\|Z_i\|d\hat{\phi}_w(s)\cdot\|\hat{\beta}_n-\beta_0\| = nh^{q/2}\cdot O_p\Big(\frac{1}{nh^q}\Big)\cdot O_p\Big(\frac{1}{\sqrt{n}}\Big)$$

which is the order of $o_p(1)$ by the assumption $nh^{2q}\to\infty$. We can also show that $nh^{q/2}C_{n5}=o_p(1)$. In fact, $nh^{q/2}|C_{n5}|$ is bounded by the sum

$$\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(s)|e_i|\,|d_{ni}|d\hat{\phi}_w(s)+\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(s)|e_i|\|\dot{\mu}(S_i;\vartheta_0)\|d\hat{\phi}_w(s)\|\hat{\vartheta}_n-\vartheta_0\|,$$

while both terms can be shown, by assumption (m7) and $\sqrt{n}$-consistency of $\hat{\vartheta}_n$, to be $O_p(1/nh^q)\cdot O_p(1/\sqrt{n})$, and therefore, $nh^{q/2}C_{n5}=O_p(1/\sqrt{nh^{2q}})$ which is $o_p(1)$ as claimed. Finally, $nh^{q/2}C_{n6}=o_p(1)$ can also be justified by the similar arguments under the assumption (m7) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ and $\hat{\vartheta}_n$.

Simply summarizing the above arguments concludes the proof of the lemma. □

Lemma 4.7. *Under the same conditions as in Lemma 4.5, $\hat{\Gamma}_n\to\Gamma$ in probability.*

Proof of Lemma 4.7. We shall prove the lemma by showing that $\hat{\Gamma}_n-\tilde{\Gamma}_n=o_p(1)$, $\tilde{\Gamma}_n-\Gamma_n=o_p(1)$ and $\Gamma_n-\Gamma=o_p(1)$. For the sake of brevity, denote $t_{ni}=Z_i'(\hat{\beta}_n-\beta_0)+\mu(S_i;\hat{\vartheta}_n)-\mu(S_i;\vartheta_0)$. Hence $\hat{e}_i=e_i-t_{ni}$, and $\hat{\Gamma}_n$ can be written as the sum $\hat{\Gamma}_n=\hat{\Gamma}_{n1}+R_n$, where

$$\hat{\Gamma}_{n1}=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_jd\hat{\phi}_h(s)\Big)^2,$$

and R_n can be further written as the sum of nine terms. In the following, we shall show that $\hat{\Gamma}_{n1}-\tilde{\Gamma}_n=o_p(1)$, and $R_n=o_p(1)$. We will not investigate all the nines terms involved in R_n, only the four terms in the following will be considered.

$$R_{n1}=\frac{4h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_jd\hat{\phi}_h(s)\Big)\Big(\int K_{hi}(s)K_{hj}(s)e_it_{nj}d\hat{\phi}_h(s)\Big),$$

$$R_{n2}=\frac{4h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_jd\hat{\phi}_h(s)\Big)\Big(\int K_{hi}(s)K_{hj}(s)t_{ni}t_{nj}d\hat{\phi}_h(s)\Big),$$

$$R_{n3}=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)t_{ni}t_{nj}d\hat{\phi}_h(s)\Big)^2,$$

$$R_{n4}=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_it_{nj}d\hat{\phi}_h(s)\Big)^2.$$

Note that $\hat{\Gamma}_{n1}-\tilde{\Gamma}_n$ can be further written as the sum of the following two terms

$$Q_{n1}=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_j\Big(\frac{f^2(s)}{\hat{f}_h^2(s)}-1\Big)d\phi(s)\Big)^2,$$

$$Q_{n2}=\frac{4h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_j\Big(\frac{f^2(s)}{\hat{f}_h^2(s)}-1\Big)d\phi(s)\Big)\Big(\int K_{hi}(s)K_{hj}(s)e_ie_jd\phi(s)\Big).$$

Also note that

$$\tilde{\Gamma}_n=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_ie_jdG(s)\Big)^2=2\sum_{i\neq j}H_n^2(\tilde{S}_i,\tilde{S}_j),$$

and by (4.10), $E\tilde{\Gamma}_n=E\Big[2\sum_{i\neq j}H_n^2(\tilde{S}_i,\tilde{S}_j)\Big]=2n(n-1)EH_n^2(\tilde{S}_1,\tilde{S}_2)=O(1)$. Hence $Q_{n1}\le\sup_{s\in I}\Big|f^2(s)/\hat{f}_h^2(s)-1\Big|\cdot\tilde{\Gamma}_n=o_p(1)$. Applying the Cauchy–Schwarz inequality to the double sum, we have $|Q_{n2}|\le 2Q_{n1}^{1/2}\cdot\tilde{\Gamma}_n^{1/2}=o_p(1)$. Therefore, $\hat{\Gamma}_{n1}-\tilde{\Gamma}_n=o_p(1)$ is proved.

Now, let us investigate the asymptotic behavior of the remainder terms $R_{nk},\ k=1,2,3,4$. Consider R_{n4} first. Recall the notation d_{ni}, we can write t_{ni} as

$$t_{ni}=Z_i'(\hat{\beta}_n-\beta_0)+d_{ni}+\dot{\mu}'(S_i;\vartheta_0)(\hat{\vartheta}_n-\vartheta_0). \tag{4.16}$$

Thus, R_{n4} can be bounded above by $6(R_{n41}+R_{n42}+R_{n43})$, where

$$R_{n41} = \frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_iZ_j' d\hat\phi_h(s)(\hat\beta_n-\beta_0)\Big)^2,$$
$$R_{n42} = \frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_i d_{nj} d\hat\phi_h(s)\Big)^2,$$
$$R_{n43} = \frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)e_i\dot\mu'(S_j,\vartheta_0)d\hat\phi_h(s)(\hat\vartheta_n-\vartheta_0)\Big)^2.$$

By taking the expected value, using Fubini Theorem, and usual calculation, we can obtain that

$$\frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)|e_i|\cdot[|e_j|+\|Z_j\|+\|\dot\mu(S_j;\vartheta_0)\|]d\hat\phi_h(s)\Big)^2 = O_p(1), \tag{4.17}$$

and if we denote $\|Z_i\|\cdot\|Z_j\|$, $\|\dot\mu(S_i;\vartheta_0)\|\cdot\|\dot\mu(S_j;\vartheta_0)\|$ and $\|Z_i\|\cdot\|\dot\mu(S_j;\vartheta_0)\|$ as $L_{ij}(Z,S)$, then

$$\frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)L_{ij}(Z,S)d\hat\phi_h(s)\Big)^2 = O_p(1). \tag{4.18}$$

Therefore, by (4.17), the $\sqrt{n}$-consistency of $\hat\beta_n$ and ϑ_n, we have

$$R_{n41} \le \frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)|e_i|\cdot\|Z_j\|d\hat\phi_h(s)\Big)^2\cdot\|\hat\beta_n-\beta_0\|^2 = o_p(1),$$
$$R_{n43} \le \frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)|e_i|\cdot\|\dot\mu'(S_j,\vartheta_0)\|d\hat\phi_h(s)\Big)^2\cdot\|\hat\vartheta_n-\vartheta_0\|^2 = o_p(1).$$

By (4.17), the $\sqrt{n}$-consistency of $\hat\vartheta_n$, and (m7), we have

$$R_{n42} \le \sup_{1\le j\le n}\Big|\frac{d_{nj}^2}{\|\hat\vartheta_n-\vartheta_0\|^2}\Big|\cdot\frac{h^q}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)|e_i|d\hat\phi_h(s)\Big)^2\cdot\|\hat\vartheta_n-\vartheta_0\|^2 = o_p(1).$$

For R_{n3}, using (4.16), expanding the squared terms, by (m7), (4.18), the $\sqrt{n}$-consistency of $\hat\beta_n$ and ϑ_n, and using similar arguments as in showing $R_{n4}=o_p(1)$, we can obtain $R_{n3}=o_p(1)$.

Finally, applying Cauchy–Schwarz inequality to the double sum, and using the facts that $R_{n3}=o_p(1)$, $R_{n4}=o_p(1)$, $\hat\Gamma_{n1}-\tilde\Gamma_n=o_p(1)$, $\tilde\Gamma_n=O_p(1)$, we can show that both R_{n1} and R_{n2} are the order of $o_p(1)$.

To show $\tilde\Gamma_n-\Gamma_n=o_p(1)$, note that $E\hat\Gamma_n=\Gamma_n$. Also, by (4.11), we have

$$\begin{aligned}E[\tilde\Gamma_n-\Gamma_n]^2 &= 4E\Big[\sum_{i\neq j}(H_n^2(\tilde S_i,\tilde S_j)-EH_n^2(\tilde S_i,\tilde S_j))\Big]^2\\ &\le 4\sum_{i\neq j}EH_n^4(\tilde S_i,\tilde S_j)+4\sum_{k\neq j\neq l}E(H_n^2(\tilde S_k,\tilde S_j)H_n^2(\tilde S_k,\tilde S_l))\\ &\le 4(n^2+n^3)EH_n^4(\tilde S_1,\tilde S_2)=O(h^q/n)=o(1).\end{aligned}$$

Cauchy–Schwarz inequality is used in the last equality.

To finish the proof, we note that as $n\to\infty$,

$$\Gamma_n = \frac{2h^q(n-1)}{n}\iint[EK_h(x-S)K_h(y-S)\tau^2(S)]^2d\phi(x)d\phi(y) = 2n(n-1)EH_n^2(\tilde S_1,\tilde S_2)\to\Gamma,$$

by (4.13) and the definition of Γ. □

Proof of Theorem 2.4. Let $Y_i^a = Z_i'\beta_a+\mu(S_i,\vartheta_a)+\xi_i$. It is easy to see that $\hat e_i$ now can be written as

$$\begin{aligned}\hat e_i &= Y_i-Y_i^a+Y_i^a-Z_i'\hat\beta_n-\mu(S_i;\hat\vartheta_n)=\hat e_i^a+Y_i-Y_i^a\\ &= \hat e_i^a+Z_i'(\beta_0-\beta_a)+[\mu(S_i)-\mu(S_i;\vartheta_a)]=\hat e_i^a+\Delta_i(\beta_0-\beta_a,\vartheta_a).\end{aligned} \tag{4.19}$$

Therefore, $M_n(\hat{\theta}_n)$ then can be written as $M_n(\hat{\theta}_n) = S_{n1} + 2S_{n2} + S_{n3}$, where

$$S_{n1} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\hat{e}_i^a\Big]^2 d\hat{\phi}_w(s),$$
$$S_{n2} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\hat{e}_i^a\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big] d\hat{\phi}_w(s),$$
$$S_{n3} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big]^2 d\hat{\phi}_w(s).$$

If we define $C_n^* = n^{-2}\sum_{i=1}^n \int K_h^2(s-S_i)\hat{e}_i^{a2} d\hat{\phi}_w(s)$, then we can show that $nh^{q/2}[M_n(\hat{\theta}_n) - C_n^*] \to_D N(0, \Gamma^*)$ with $\Gamma^* = 2\int \tau_a^4(s)g(s)d\phi(s) \cdot \int \Big[\int K(u)K(u+v)du\Big]^2 dv$ and $\tau_a^2(s) = E(\xi^2|S=s)$. Also, by a routine argument, we can show that

$$S_{n3} = \int [Z'(s)(\beta_0-\beta_a) + (\mu(s) - \mu(s;\vartheta_a))]^2 dG(s) + o_p(1).$$

Adding and subtracting $Z_i'\beta_\vartheta + \mu(S_i;\vartheta_a)$ from $\hat{e}_i^a$, S_{n2} can be written as the sum of the following three terms,

$$S_{n21} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)e_i^a\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big] d\hat{\phi}_w(s),$$
$$S_{n22} = (\hat{\beta}_n - \beta_a)' \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big] d\hat{\phi}_w(s),$$
$$S_{n23} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\hat{\vartheta}_n) - \mu(S_i;\vartheta_a)]\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big] d\hat{\phi}_w(s).$$

Cauchy–Schwarz inequality and the following facts

$$\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)e_i^a\Big]^2 d\hat{\phi}_w(s) = O_p\Big(\frac{1}{nh^q}\Big),$$
$$\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Big]^2 d\hat{\phi}_w(s) = O_p(1),$$

imply that $S_{n21} = o_p(1)$, and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ to β_a, and $\hat{\vartheta}_n$ to ϑ_a, implies that $S_{n22} = o_p(1) = S_{n23}$.

Using the relationship (4.19), we can write $\hat{C}_n$ as the sum of C_n^* and the following two terms

$$C_{n1} = \frac{2}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)\hat{e}_i^a \Delta_i(\beta_0-\beta_a, \vartheta_a)\hat{\phi}_w(s),$$
$$C_{n2} = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(s)\Delta_i^2(\beta_0-\beta_a, \vartheta_a)\hat{\phi}_w(s),$$

and it is easy to show that both terms are of the order of $o_p(1)$.

Next, we shall show that $\hat{\Gamma}_n \to \Gamma^*$ in probability under the fixed alternative H_a. For, again using the relationship (4.19), we can write $\hat{\Gamma}_n$ as the sum of the $\sum_{k=1}^3 \hat{\Gamma}_{nk} + R_n$, where

$$\Gamma_{n1} = 2h^q n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)\hat{e}_i^a\hat{e}_j^a\Big)^2 d\hat{\phi}_w(s),$$
$$\Gamma_{n2} = 2h^q n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)\Delta_i(\beta_0-\beta_a, \vartheta_a)\Delta_j(\beta_0-\beta_a, \vartheta_a)\Big)^2 d\hat{\phi}_w(s),$$
$$\Gamma_{n3} = 4h^q n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(s)K_{hj}(s)\hat{e}_i^a\Delta_j(\beta_0-\beta_a, \vartheta_a)\Big)^2 d\hat{\phi}_w(s).$$

It can be shown that the remainder term $R_n = o_p(1)$, and $\Gamma_{n1} \to \Gamma^*$,

$$\Gamma_{n2} = 2\int [Z'(s)(\beta_0 - \beta_a) + \mu(s) - \mu(s;\vartheta_a)]^4 g(s)d\phi(s) \cdot \int \Big[\int K(u)K(u+v)du\Big]^2 dv$$

$$\Gamma_{n3} = 4\int [Z'(s)(\beta_0 - \beta_a) + \mu(s) - \mu(s;\vartheta_a)]^2 \tau^2(s) g(s)d\phi(s) \cdot \int \Big[\int K(u)K(u+v)du\Big]^2 dv.$$

All of the above results imply that

$$T_n = nh^{q/2}\hat{\Gamma}_n^{-1/2}(S_{n1} - C_n^*) + nh^{q/2}\hat{\Gamma}_n^{-1/2}\int [Z'(s)(\beta_0 - \beta_a) + (\mu(s) - \mu(s;\vartheta_a))]^2 dG(s) + o_p(nh^{q/2}).$$

Hence the theorem. □

Proof of Theorem 2.5. Let $Y_i^L = Z_i'\beta_0 + \mu(S_i;\vartheta_0) + \xi_i$. Then $\hat{e}_i$ can be written as

$$\hat{e}_i = Y_i - Y_i^L + Y_i^L - Z_i'\hat{\beta}_n - \mu(S_i;\hat{\vartheta}_n) = \hat{e}_i^L + L(S_i)/\sqrt{nh^{q/2}}. \tag{4.20}$$

Therefore, $M_n(\hat{\theta}_n)$ then can be written as $M_n(\hat{\theta}_n) = S_{n1} + 2S_{n2} + S_{n3}$, where

$$S_{n1} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\hat{e}_i^L\Big]^2 d\hat{\phi}_w(s),$$

$$S_{n2} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\hat{e}_i^L\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s),$$

$$S_{n3} = \frac{1}{nh^{q/2}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big]^2 d\hat{\phi}_w(s).$$

If we define $C_n^L = n^{-2}\sum_{i=1}^n \int K_h^2(s - S_i)\hat{e}_i^{L2} d\hat{\phi}_w(s)$, then we can show that $nh^{q/2}[M_n(\hat{\theta}_n) - C_n^L] \to_D N(0, \Gamma)$ with $\Gamma = 2\int \tau^4(s)g(s)d\phi(s) \cdot \int \Big[\int K(u)K(u+v)du\Big]^2 dv$ and $\tau^2(s) = E(\xi^2|S = s)$. As for S_{n3}, we can easily show that

$$nh^{q/2}S_{n3} = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big]^2 d\hat{\phi}_w(s) = \int L^2(s)dG(s) + o_p(1).$$

Now, let us consider S_{n2}. Rewrite S_{n2} as the sum of the following three terms

$$S_{n21} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\xi_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s),$$

$$S_{n22} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i'(\beta_0 - \hat{\beta}_n)\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s),$$

$$S_{n23} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)[\mu(S_i;\vartheta_0) - \mu(S_i;\hat{\vartheta}_n)]\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s).$$

By the $\sqrt{n}$-consistency of $\hat{\beta}_n$, $\hat{\vartheta}_n$, the assumption (m7), and the fact that

$$\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)Z_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s) = O_p(1),$$

and

$$\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\cdot\mu(S_i;\vartheta_0)\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)L(S_i)\Big] d\hat{\phi}_w(s) = O_p(1),$$

one can show that $nh^{q/2}S_{n22} = O_p(h^{q/4}) = o_p(1)$, and $nh^{q/2}S_{n23} = O_p(h^{q/4}) = o_p(1)$.

Adding and subtracting $E[K_{h1}(s)L(S)]$ from $n^{-1}\sum_{i=1}^n K_{hi}(s)L(S_i)$ in S_{n41}, we can show that S_{n21} is the sum of the following two terms,

$$S_{n211} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\xi_i\Big] E[K_{h1}(s)L(S)] d\hat{\phi}_w(s),$$

$$S_{n212} = \frac{1}{\sqrt{nh^{q/2}}}\int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(s)\xi_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n (K_{hi}(s)L(S_i) - E[K_{h1}(s)L(S)])\Big] d\hat{\phi}_w(s).$$

By Cauchy–Schwarz inequality and a similar argument as before, S_{n212}^2 is bounded above by

$$\frac{1}{nh^{q/2}}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(s)\xi_i\Big]^2 d\hat{\phi}_w(s)\int\Big[\frac{1}{n}\sum_{i=1}^{n}(K_{hi}(s)L(S_i)-E[K_{h1}(s)L(S)])\Big]^2 d\hat{\phi}_w(s),$$

and it has the order of $O_p(1/n^3h^{5q/2})$. Therefore, $nh^{q/2}S_{n212}=O_p(1/\sqrt{nh^{3q/2}})=o_p(1)$ by the assumption $nh^{2q}\to\infty$. While using the similar argument as in showing $\sqrt{n}S_{n1}=O_p(1)$ in the proof of the asymptotic normality of $\hat{\theta}_n$ under H_0, we can also show that

$$\sqrt{n}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(s)\xi_i\Big]E[K_{h1}(s)L(S)]d\hat{\phi}_w(s)=O_p(1).$$

Hence, $nh^{q/2}S_{n211}=O_p(h^{q/4})=o_p(1)$.

By summarizing the above results, we proved that under the local alternative hypothesis H_{loc}, $nh^{q/2}(M_n(\hat{\theta}_n)-C_n^L)\to_D N\Big(\int L^2(s)dG(s),\Gamma\Big)$.

Now, let us consider the asymptotic behavior of C_n under H_{loc}. Using the notation $\hat{e}_i^L$, we can write C_n as the sum $C_n^L+2C_{n1}+2C_{n2}$, where

$$C_{n1}=\frac{1}{n^2\sqrt{nh^{q/2}}}\sum_{i=1}^{n}\int K_{hi}^2(s)\hat{e}_i^L L(S_i)d\hat{\phi}_w(s),$$

$$C_{n2}=\frac{1}{n^3h^{q/2}}\sum_{i=1}^{n}\int K_{hi}^2(s)L^2(S_i)d\hat{\phi}_w(s).$$

C_{n1} can be further written as the sum

$$C_{n11}=\frac{1}{n^2\sqrt{nh^{q/2}}}\sum_{i=1}^{n}\int K_{hi}^2(s)\xi_i L(S_i)d\hat{\phi}_w(s),$$

$$C_{n12}=\frac{1}{n^2\sqrt{nh^{q/2}}}\sum_{i=1}^{n}\int K_{hi}^2(s)[Z_i'(\beta_0-\hat{\beta}_n)+\mu(S_i;\vartheta_0)-\mu(S_i;\hat{\vartheta}_n)]d\hat{\phi}_w(s).$$

Then from the fact that

$$\frac{1}{n}\sum_{i=1}^{n}\int K_{hi}^2(s)[|\xi_i L(S_i)|+\|Z_i\|+|\dot{\mu}(S_i;\vartheta_0)|+L^2(S_i)]d\hat{\phi}_w(s)=O_p(1/h^q),$$

we can show that

$$nh^{q/2}C_{n11}=O_p\Big(\frac{1}{\sqrt{nh^{3q/2}}}\Big),\qquad nh^{q/2}C_{n12}=O_p\Big(\frac{1}{n\sqrt{h^{3q/2}}}\Big),\qquad nh^{q/2}C_{n2}=O_p\Big(\frac{1}{nh^q}\Big).$$

Therefore, $nh^{q/2}(\hat{C}_n-C_n^L)=o_p(1)$.

Finally, to see the asymptotic property of $\hat{\Gamma}_n$ under the local alternative, using the notation $\hat{e}_i^L$ again, we can write

$$\hat{\Gamma}_n=\frac{2h^q}{n^2}\sum_{i\neq j}\Big(\int K_h(s-S_i)K_h(s-S_j)\xi_i^L\xi_j^L d\hat{\phi}_h(s)\Big)^2+R_n.$$

The first term converges to Γ in probability. The remainder term $R_n=o_p(1)$ can be proven by using the Cauchy–Schwarz inequality on the double sum, consistency of $\hat{\beta}_n$, $\hat{\vartheta}_n$, and the facts (4.17), (4.18) with ξ_i replaced by ξ_i^L, ξ_j by ξ_j^L. □

References

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes, second ed. Springer, Berlin.

Cuzick, J., 1992. Semiparametric additive regression. J. Roy. Statist. Soc. Ser. B 54, 831–843.

Engle, R.F., Granger, C.W.J., Rice, J., Weiss, A., 1986. Semiparametric estimates of the relation between weather and electricity sales. J. Amer. Statist. Assoc. 81, 310–320.

Hall, P.J., 1984. Central limit theorem for integrated square error of multivariate nonparametric density esitmators. J. Multivariate Anal. 14, 1–16.

Härdle, W., Liang, H., Gao, J.T., 2000. Partially Linear Models. Springer Physica-Verlag, Heidelberg.

Hart, J.D., 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer-Verlag, New York.

Heckman, N.E., 1986. Spline smoothing in partly linear models. J. Roy. Statist. Soc. Ser. B 48, 244–248.

Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.

Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32, 995–1034.

Koul, H.L., Ni, P., 2004. Minimum distance regression model checking. J. Statist. Plann. Inference 119, 109–141.

Koul, H.L., Song, W.X., 2010. Model checking in partial linear regression models with Berkson measurment errors. Statist. Sinica 20, 1551–1579.
Liang, H., 2006. Checking linearity of non-parametric component in partially linear models with an application in systemic inflammatory response syndrome study. Stat. Methods Med. Res. 15, 273–284.
Mammen, E., van de Geer, S., 1997. Penalized quasi-likelihood estimation in partial linear models. Ann. Statist. 25, 1014–1035.
Rice, J.A., 1986. Convergence rates for partially splined models. Statist. Probab. Lett. 4, 203–208.
Robinson, P.M., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 931–954.
Rudemo, M., Ruppert, D., Streibig, J.C., 1989. Random effect models in nonlinear regression with applications to bioassay. Biometrics 45, 349–362.
Schennach, S.M., 2013. Regressions with Berkson errors in covariates - A nonparametric approach. Ann. Statist. 41, 1642–1668.
Speckman, P., 1988. Kernel smoothing in partial linear models. J. Roy. Statist. Soc. Ser. B 50, 413–436.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: an innovation process approach. Ann. Statist. 26, 1916–1934.
Stute, W., Zhu, L.X., 2002. Model checks for generalized linear models. Scand. J. Stat. 29, 535–545.
Stute, W., Zhu, L.X., 2005. Nonparametric checks for single-index models. Ann. Statist. 33, 1048–1083.
Wang, L., 2003. Estimation of nonlinear Berkson type measurement error models. Statist. Sinica 13, 1201–1210.
Wang, L., 2004. Estimation of nonliner models with Berkson measurement errors. Ann. Statist. 32, 2559–2579.
White, H., 1981. Consequences and detection of misspecified nonlinear regression models. J. Amer. Statist. Assoc. 76, 419–433.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.
Zhu, L.X., Ng, K.W., 2003. Checking the adequacy of a partial linear model. Statist. Sinica 13, 763–781.