Weighted Empirical Minimum Distance Estimators in Berkson Measurement Error Regression Models

Hira L. Koul and Pei Geng

Abstract We develop analogs of the two classes of weighted empirical minimum distance (m.d.) estimators of the underlying parameters in linear and nonlinear regression models when covariates are observed with Berkson measurement error. One class is based on the integral of the square of symmetrized weighted empirical of residuals while the other is based on a similar integral involving a weighted empirical of residual ranks. The former class requires the regression and measurement errors to be symmetric around zero while the latter class does not need any such assumption. The first class of estimators includes the analogs of the least absolute deviation and Hodges-Lehmann estimators while the second class includes an estimator that is asymptotically more efficient than these two estimators at some error distributions when there is no measurement error. In the case of linear model, no knowledge of the measurement error distribution is needed. We also develop these estimators for nonlinear models when the measurement error distribution is known and when it is unknown but validation data is available.

Keywords Analog of Hodges-Lehmann estimator · Validation data

1 Introduction

Statistical literature is replete with the various minimum distance estimation methods in the one and two sample location models. Beran [2, 3] and Donoho and Liu [7, 8] argue that the minimum distance estimators based on L_2 distances involving either density estimators or residual empirical distribution functions have some desirable

H. L. Koul (✉)
Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA
e-mail: koul@msu.edu

P. Geng
Department of Mathematics, College of Arts and Sciences, Illinois State University, Normal, IL, USA
e-mail: pgeng@ilstu.edu

M. Maciak et al. (eds.), *Analytical Methods in Statistics*, Springer Proceedings in Mathematics & Statistics 329, https://doi.org/10.1007/978-3-030-48814-7_3

finite sample properties, tend to be robust against some contaminated models and are also asymptotically efficient at some error distributions.

In the classical regression models without measurement error in the covariates, classes of minimum distance estimators of the underlying parameters based on Cramér-von Mises type distances between certain weighted residual empirical processes were developed in Koul [12–15]. These classes include some estimators that are robust against outliers in the regression errors and asymptotically efficient at some error distributions.

In practice there are numerous situations when covariates are not observable. Instead one observes their surrogate with some error. The regression models with such covariates are known as the measurement error regression models. Fuller [9], Cheng and Van Ness [6], Carroll et al. [5] and Yi [19] discuss numerous examples of practical importance of these models.

Given the desirable properties of the above minimum distance (m.d.) estimators and the importance of the measurement error regression models, it is desirable to develop their analogs for these models. The next section describes the m.d. estimators of interest and their asymptotic distributions in the classical linear regression model. Their analogs for the linear regression Berkson measurement error (ME) model are developed in Sect. 3. The two classes of m.d. estimators are developed. One assumes the symmetry of the regression model error and ME error distributions and then basis the m.d. estimators on the symmetrized weighted empirical of the residuals. This class includes an analog of the Hodges-Lehmann estimator of the one sample location parameter, see Hodges and Lehmann (1963), and the least absolute deviation (LAD) estimator. The second class is based on a weighted empirical of residual ranks. This class of estimators does not need the symmetry of the errors distributions. This class includes an estimator that is asymptotically more efficient than the analog of Hodges-Lehmann and LAD estimators at some error distributions. Neither classes need the knowledge of the measurement error or regression error distributions.

Section 4 discusses analogs of these estimators in the Berkson measurement error nonlinear regression models, where the measurement error distribution is assumed to be known. Section 5 develops their analogs when the ME distribution is unknown but validation data is available. In this case the consistency rate of these estimators is $\min(n, N)^{1/2}$, where n and N are the primary data and validation data sample sizes, respectively. Section 6 provides an application of the proposed estimators to a real data example. Several proofs are deferred to the last section.

2 Linear Regression Model

In this section we recall the definition of the m.d. estimators of interest here in the no measurement error linear regression model and their known asymptotic normality results.

Accordingly, consider the linear regression model where for some $\theta \in \mathbb{R}^p$, the response variable Y and the p dimensional observable predicting covariate vector X obey the relation

$$Y = X'\theta + \varepsilon, \tag{1}$$

where ε is independent of X and symmetrically distributed around $E(\varepsilon) = 0$. For an $x \in \mathbb{R}$, x' and $\|x\|$ denote its transpose and Euclidean norm, respectively. Let (X_i, Y_i), $1 \le i \le n$ be a random sample from this model. The two classes of m.d. estimators of θ based on weighted empirical processes of the residuals and residual ranks were developed in Koul [12–15]. To describe these estimators, let G be a nondecreasing right continuous function from $\mathbb{R}$ to $\mathbb{R}$ having left limits and define

$$V(x, \vartheta) := n^{-1/2} \sum_{i=1}^{n} X_i \big\{ I(Y_i - X_i'\vartheta \le x) - I(-Y_i + X_i'\vartheta < x) \big\},$$

$$M(\vartheta) := \int \big\| V(x, \vartheta) \big\|^2 dG(x), \qquad \hat{\theta} := \mathrm{argmin}_{\vartheta \in \mathbb{R}^p} M(\vartheta).$$

This class of estimators, one for each G, includes some well celebrated estimators. For example $\hat{\theta}$ corresponding to $G(x) \equiv x$ yields an analog of the one sample location parameter Hodges-Lehmann estimator in the linear regression model. Similarly, $G(x) \equiv \delta_0(x)$, the degenerate measure at zero, makes $\hat{\theta}$ equal to the least absolute deviation (LAD) estimator.

A class of m.d. estimators when the error distribution is not symmetric and unknown is obtained by using the weighted empirical of the residual ranks defined as follows. Write $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})'$, $i = 1, \ldots, n$. Let $\bar{X}_j := n^{-1} \sum_{i=1}^n X_{ij}$, $\bar{X} := (\bar{X}_1, \ldots, \bar{X}_p)'$ and $X_{ic} := X_i - \bar{X}$, $1 \le i \le n$. Let $R_{i\vartheta}$ denote the rank of the ith residual $Y_i - X_i'\vartheta$ among $Y_j - X_j'\vartheta$, $j = 1, \ldots, n$. Let Ψ be a distribution function on [0, 1] and define

$$V(u, \vartheta) := n^{-1/2} \sum_{i=1}^{n} X_{ic} I(R_{i\vartheta} \le nu), \quad K(\vartheta) := \int_0^1 \|\mathcal{V}(u, \vartheta)\|^2 d\Psi(u),$$

$$\hat{\theta}_R := \mathrm{argmin}_{\vartheta \in \mathbb{R}^p} K(\vartheta).$$

Yet another m.d. estimator, when error distribution is unknown and not symmetric, is

$$V_c(x, \vartheta) := n^{-1/2} \sum_{i=1}^{n} X_{ic} I(Y_i - X_i'\vartheta \le x),$$

$$M_c(\vartheta) := \int \big\| V_c(x, \vartheta) \big\|^2 dx, \qquad \hat{\theta}_c := \mathrm{argmin}_{\vartheta \in \mathbb{R}^p} M_c(\vartheta).$$

If one reduces the model (1) to the two sample location model, then $\hat{\theta}_c$ is the median of pairwise differences, the so called Hodges-Lehmann estimator of the two sample location parameter. Thus in general $\hat{\theta}_c$ is an analog of this estimator in the linear regression model.

The following asymptotic normality results can be deduced from Koul [15] and [16, Sect. 5.4].

Lemma 1 *Suppose the model (1) holds and $E\|X\|^2 < \infty$.*
(a). *In addition, suppose $\Sigma_X := E(XX')$ is positive definite and the error d.f. F is symmetric around zero and has density f. Further, suppose the following hold.*

$$G \text{ is a nondecreasing right continuous function on } \mathbb{R} \text{ to } \mathbb{R}, \quad (2)$$
$$\text{having left limits and } dG(x) = -dG(-x), \forall\, x \in \mathbb{R}.$$
$$0 < \int f^j dG < \infty, \quad \lim_{z\to 0} \int \big[f(x+z) - f(x)\big]^j dG(x) = 0, \; j = 1, 2, \quad (3)$$
$$\int_0^\infty (1-F)dG < \infty.$$

Then

$$n^{1/2}(\hat\theta - \theta) \to_D N\big(0, \sigma_G^2 \Sigma_X^{-1}\big), \quad \sigma_G^2 := \frac{Var\Big(\int_{-\infty}^{\varepsilon} f(x)dG(x)\Big)}{\big(\int f^2 dG\big)^2}.$$

(b). *In addition, suppose the error d.f. F has uniformly continuous bounded density f, $\Omega := E\{(X - EX)(X - EX)'\}$ is positive definite and Ψ is a d.f. on* $[0, 1]$ *such that $\int_0^1 f^2(F^{-1}(s))d\Psi(s) > 0$. Then*

$$n^{1/2}\big(\hat\theta_R - \theta\big) \to_D N(0, \gamma_\Psi^2 \Omega^{-1}), \quad \gamma_\Psi^2 := \frac{Var\Big(\int_0^{F(\varepsilon)} f(F^{-1}(s))d\Psi(s)\Big)}{\big(\int_0^1 f^2(F^{-1}(s))d\Psi(s)\big)^2}.$$

(c). *In addition, suppose Ω is positive definite, F has square integrable density f and $E|\varepsilon| < \infty$. Then $n^{1/2}(\hat\theta_c - \theta) \to_D N\big(0, \sigma_I^2 \Omega^{-1}\big)$, where $\sigma_I^2 := 1/12\big(\int f^2(x)dx\big)^2$.*

Before proceeding further we now describe some comparison of the above asymptotic variances. Let $\sigma_{LAD}^2 := 1/(4f^2(0))$ and $\sigma_{LSE}^2 := \mathrm{Var}(\varepsilon)$ denote the factors of the asymptotic covariance matrices of the LAD and the least squares estimators, respectively. Let γ_I^2 denote the γ_Ψ^2 when $\Psi(s) \equiv s$, i.e.,

$$\gamma_I^2 = \frac{\int\int \big[F(x \wedge y) - F(x)F(y)\big] f^2(x) f^2(y) dx dy}{\big(\int_0^1 f^3(x)dx\big)^2}.$$

Table 1, obtained from Koul [16], gives the values of these factors for some distributions F. From this table one sees that the estimator $\hat\theta_R$ corresponding to $\Psi(s) \equiv s$

Table 1 A comparison of asymptotic variances

F	γ_I^2	σ_I^2	σ_{LAD}^2	σ_{LSE}^2
Double Exp.	1.2	1.333	1	2
Logistic	3.0357	3	4	3.2899
Normal	1.0946	1.0472	1.5708	1
Cauchy	2.5739	3.2899	2.46	∞

is asymptotically more efficient than the LAD at logistic error distribution while it is asymptotically more efficient than the Hodges-Lehmann type estimator at the double exponential and Cauchy error distributions. For these reasons it is desirable to develop analogs of $\hat{\theta}_R$ also for the ME models.

As argued in Koul (Chap. 5, [16]), the estimators $\{\hat{\theta}_G,\ G$ a d.f.$\}$ are robust against heavy tails in the error distribution in the general linear regression model. The estimator $\hat{\theta}_I$, where $G(x) \equiv x$, not a d.f., is robust against heavy tails and also asymptotically efficient at the logistic errors.

3 Berkson ME Linear Regression Model

In this section we shall develop analogs of the above estimators in the Berkson ME linear regression model, where the response variable Y obeys the relation (1) and where, instead of observing X, one observes a surrogate Z obeying the relation

$$X = Z + \eta. \tag{4}$$

In (4), Z, η, ε are assumed to be mutually independent and $E(\eta) = 0$. Note that η is $p \times 1$ vector of errors and its distribution need not be known.

Analog of $\hat{\theta}$. We shall first develop and derive the asymptotic distribution of the analogs of the estimators $\hat{\theta}$ in the Berkson ME linear regression model (1) and (4). Rewrite the model as

$$Y = Z'\theta + \xi, \qquad \xi := \eta'\theta + \varepsilon, \quad E(\xi) = 0, \quad \exists\, \theta \in \mathbb{R}. \tag{5}$$

Because Z, η, ε are mutually independent, ξ is independent of Z in (5).

Let H denote the distribution functions (d.f.) of η. Assume that the d.f. F of ε is continuous and symmetric around zero and that H is also symmetric around zero, i.e., $-dH(v) = dH(-v)$, for all $v \in \mathbb{R}^p$. Then the d.f. of ξ

$$L(x) := P(\xi \le x) = P(\eta'\theta + \varepsilon \le x) = \int F(x - v'\theta)dH(v)$$

is also continuous and symmetric around zero. This symmetry in turn motivates the following definition of the class of m.d. estimators of θ in the model (5), which mimics the definition of $\hat{\theta}$ by simply replacing X_i by Z_i. Define

$$\widetilde{V}(x,t) := n^{-1/2}\sum_{i=1}^{n} Z_i\big\{I(Y_i - Z_i't \le x) - I(-Y_i + Z_i't < x)\big\},$$

$$\widetilde{M}(t) := \int \big\|\widetilde{V}(x,t)\big\|^2 dG(x), \qquad \widetilde{\theta} := \operatorname{argmin}_{t\in\mathbb{R}^p}\widetilde{M}(t).$$

Because L is continuous and symmetric around zero and ξ is independent of Z, $E\widetilde{V}(x,\theta) \equiv 0$.

The following assumptions are needed for the asymptotic normality of $\widetilde{\theta}$.

$$E\|Z\|^2 < \infty \text{ and } \Gamma := EZZ' \text{ is positive definite.} \tag{6}$$

$$H \text{ satisfies } dH(v) = -dH(-v),\ \forall\, v \in \mathbb{R}^p. \tag{7}$$

$$F \text{ has Lebesgue density } f, \text{ symmetric around zero, and} \tag{8}$$

such that $\ell(x) = \displaystyle\int f(x - v'\theta)dH(v)$ of L satisfies the following:

$$0 < \int \ell^j dG < \infty, \qquad \lim_{z\to 0}\int \big[\ell(y+z) - \ell(y)\big]^j dG(y) = 0, \quad j = 1, 2.$$

$$A := \int_0^\infty (1 - L)dG < \infty. \tag{9}$$

Under (6), $n^{-1}\sum_{i=1}^n Z_i Z_i' \to_p \Gamma$ and $n^{-1/2}\max_{1\le i\le n}\|Z_i\| \to_p 0$. Use these facts and argue as in Koul [15] to deduce that (2) and (6)–(9) imply

$$n^{1/2}\big(\widetilde{\theta} - \theta\big) \to_D \mathcal{N}(0, \tau_G^2\Gamma^{-1}), \qquad \tau_G^2 := \frac{\operatorname{Var}\big(\int_{-\infty}^{\xi}\ell dG\big)}{\big(\int \ell^2 dG\big)^2}. \tag{10}$$

Remark 1 We shall discuss some examples and some sufficient conditions for the above assumptions. The conditions (8) and (9) are satisfied by a large class of densities f, ME distributions H and integrating measure G. If G is a d.f., then f being uniformly continuous and bounded implies these conditions. In this case ℓ is also uniformly continuous, $\sup_x \ell(x) \le \sup_x f(x) < \infty$ so that $\int \ell^j dG \le \sup_x f^j(x) < \infty$ and $\int \big[\ell(y+z) - \ell(y)\big]^j dG(y) \le \sup_{|x-y|\le z}|\ell(y) - \ell(x)|^j \to 0$, as $z \to 0$. Moreover, here $A \le 1$. Thus these two assumptions reduce to assuming $\int \ell^j dG > 0$, $j = 1, 2$.

Given the importance of the two estimators corresponding to $G(x) \equiv x,\ G(x) \equiv \delta_0(x)$, it is of interest to provide some easy to verify sufficient conditions that imply conditions (8) and (9) for these two estimators.

Consider the case $G(x) \equiv x$. Assume f to be continuous and $\int f^2(x)dx < \infty$. Then because H is a d.f., ℓ is also continuous and symmetric around zero and $\int \ell(x+z)dx = \int \ell(x)dx = 1$. Moreover, by the Cauchy-Schwartz (C-S) inequality and Fubini's Theorem,

$$\begin{aligned} 0 < \int \ell^2(y)dy &= \int \Big(\int f(y - v'\theta)dH(v) \Big)^2 dy \\ &\le \int \int f^2(y - v'\theta)dy dH(v) = \int f^2(x)dx < \infty. \end{aligned}$$

Finally, because $\ell \in L_2$, by Theorem 9.5 in Rudin [18], it is shift continuous in L_2, i.e., (8) holds. Hence all conditions of (8) are satisfied.

Next, consider (9). The assumptions $E(\varepsilon) = 0$ and $E(\eta) = 0$ imply that $\int |x| f(x) dx < \infty$, $\int \|v\| dH(v) < \infty$ and hence

$$\int |y| dL(y) = \int |y| \int f(y - v'\theta)dH(v)dy = \int \int |x + v'\theta| f(x)dx dH(v) < \infty.$$

This in turn implies (9) in the case $G(x) \equiv x$.

To summarize, (6), (7), and F having continuous symmetric square integrable density f implies all of the above conditions needed for the asymptotic normality of the above analog of the Hodges-Lehmann estimator in the Berkson ME linear regression model. This fact is similar to the observation made in Berkson (1950) that the naive least square estimator, where one replace X_i's by Z_i's, continues to be consistent and asymptotically normal under the same conditions as when there is no ME. But, unlike in the no ME case, here the asymptotic variance

$$\tau_I^2 := \frac{\mathrm{Var}\big(L(\xi)\big)}{\big(\int \ell^2(y)dy\big)^2} = \frac{1}{12\big(\int \big(\int f(y - v'\theta)dH(v)\big)^2 dy\big)^2}$$

depends on θ. If H is degenerate at zero, i.e., if there is no ME, then $\tau_I^2 = \sigma_I^2$, the factor that appears in the asymptotic covariance matrix of the Hodges-Lehmann estimator in this case.

Next, consider the case $G(x) \equiv \delta_0(x)$—degenerate measure at 0. Assume f to be continuous and bounded from the above and

$$\ell(0) := \int f(v'\theta)dH(v) > 0. \tag{11}$$

Then the continuity and symmetry of f implies that as $z \to 0$,

$$\int \ell(y+z)dG(y) = \ell(z) = \int f(z - v'\theta)dH(v) \to \int f(-v'\theta)dH(v) = \ell(0),$$

$$\int \big[\ell(y+z) - \ell(y)\big]^2 dG(y) = \Big[\int \big\{f(z - v'\theta) - f(-v'\theta)\big\} dH(v)\Big]^2$$
$$\leq \int \big\{f(z - v'\theta) - f(-v'\theta)\big\}^2 dH(v) \to 0.$$

Moreover, here $\int_0^\infty (1-L)dG = 1 - L(0) = 1/2$ so that (9) is also satisfied.

To summarize, (6), (7), (11) and f being continuous, symmetric around zero and bounded from the above imply all the needed conditions for the asymptotic normality of the above analog of the LAD estimator in the Berkson ME linear regression model. Moreover, here

$$\int_{-\infty}^{\xi} \ell(x)dG(x) = \ell(0)I(\xi \geq 0), \quad \int \ell^2(x)dG(x) = \ell^2(0),$$
$$\mathrm{Var}\Big(\int_{-\infty}^{\xi} \ell(x)dG(x)\Big) = \ell^2(0)/4.$$

Consequently, here the asymptotic covariance matrix also depends on θ via

$$\tau_0^2 = 1\big/4\ell^2(0) = 1\Big/4\Big(\int f(v'\theta)dH(v)\Big)^2.$$

In the case of no ME, $\Gamma^{-1}\tau_0^2$ equals the asymptotic covariance matrix of the LAD estimator. Unlike in the case of the previous estimator, here the conditions needed for f are a bit more stringent than those required for the asymptotic normality of the LAD estimator when there is no ME.

Analog of $\hat{\theta}_R$. Here we shall describe the analogs of the class of estimators $\hat{\theta}_R$ based on the residual ranks obtained from the model (5). These estimators do not need the errors ξ_i's to be symmetrically distributed. Let $\widetilde{R}_{i\vartheta}$ denote the rank of $Y_i - Z_i'\vartheta$ among $Y_j - Z_j'\vartheta,\ j = 1, \ldots, n$, $\bar{Z} := n^{-1}\sum_{i=1}^n Z_i$, $Z_{ic} := Z_i - \bar{Z}$, $1 \leq i \leq n$ and define

$$\widetilde{\mathcal{V}}(u, \vartheta) := n^{-1/2}\sum_{i=1}^n Z_{ic} I(\widetilde{R}_{i\vartheta} \leq nu), \quad \widetilde{K}(\vartheta) := \int_0^1 \|\widetilde{\mathcal{V}}(u,\vartheta)\|^2 d\Psi(u),$$
$$\widetilde{\theta}_R := \mathrm{argmin}_{\vartheta \in \mathbb{R}^p} \widetilde{K}(\vartheta).$$

Use the facts $\sum_{i=1}^n Z_{ic} = 0$, $\Psi(\max(a,b)) = \max\{\Psi(a), \Psi(b)\}$ and $\max(a,b) = 2^{-1}[a + b + |a-b|]$, for any $a, b \in \mathbb{R}$, to obtain the computational formula

$$\widetilde{K}(t) = -\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n Z_{ic}' Z_{jc}\Big|\Psi\Big(\frac{R_{it}}{n} -\Big) - \Psi\Big(\frac{R_{jt}}{n} -\Big)\Big|.$$

The following result can be deduced from Koul [15]. Suppose $E\|Z\|^2 < \infty$, $\widetilde{\Gamma} := E(Z - EZ)(Z - EZ)'$ is positive definite, density ℓ of the r.v. ξ is uniformly continuous and bounded and $\int_0^1 \ell^2(L^{-1}(s))d\Psi(s) > 0$. Then $n^{-1/2}\max_{1\le i\le n}\|Z_i\| \to_p 0$, $n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})' \to_p \widetilde{\Gamma}$ and

$$n^{1/2}\big(\widetilde{\theta}_R - \theta) \to_D N\big(0, \widetilde{\tau}_\Psi^2 \widetilde{\Gamma}^{-1}\big), \quad \widetilde{\tau}_\Psi^2 := \frac{\mathrm{Var}\big(\int_0^{L(\xi)} \ell(L^{-1}(s))d\Psi(s)\big)}{\big(\int_0^1 \ell^2(L^{-1}(s))d\Psi(s)\big)^2}.$$

Density f of F being uniformly continuous and bounded implies the same for $\ell(x) = \int f(x - v'\theta)dH(v)$. It is also worth pointing out the assumptions on F, H and L needed here are relatively less stringent than those needed for the asymptotic normality of $\widetilde{\theta}$.

Of special interest is the case $\Psi(s) \equiv s$. Let $\widetilde{\tau}_I^2$ denote the corresponding $\widetilde{\tau}_\Psi^2$. Then by the change of variable formula,

$$\begin{aligned}\widetilde{\tau}_I^2 &= \frac{\mathrm{Var}\big(\int_0^{L(\xi)} \ell(L^{-1}(s))ds\big)}{\int_0^1 \ell^2(L^{-1}(s))ds} = \frac{\mathrm{Var}\big(\int_0^{\xi} \ell^2(x)dx\big)}{\big(\int_0^1 \ell^3(x)dx)^2} \\ &= \frac{\int\int \big[L(x\wedge y) - L(x)L(y)\big]\ell^2(x)\ell^2(y)dxdy}{\big(\int_0^1 \ell^3(x)dx\big)^2}.\end{aligned}$$

An analog of $\hat{\theta}_c$ here is $\widetilde{\theta}_c := \mathrm{argmin}_{\vartheta\in\mathbb{R}^p} \widetilde{M}_c(\vartheta)$, where

$$\widetilde{V}_c(x,\vartheta) := n^{-1/2}\sum_{i=1}^n Z_{ic} I(Y_i - Z_i'\vartheta \le x), \qquad \widetilde{M}_c(\vartheta) := \int \big\|\widetilde{V}_c(x,\vartheta)\big\|^2 dx.$$

Arguing as above one obtains that $n^{1/2}\big(\widetilde{\theta}_c - \theta\big) \to_D N\big(0, \tau_I^2\widetilde{\Gamma}^{-1}\big)$.

4 Nonlinear Regression with Berkson ME

In this section we shall investigate the analogs of the above m.d. estimators in nonlinear regression models with Berkson ME.

Let $q \ge 1$, $p \ge 1$ be known positive integers, $\Theta \subseteq \mathbb{R}^q$ be a subset of the q-dimensional Euclidean space $\mathbb{R}^q$ and consider the model where the unobservable p-dimensional covariate X, its observable surrogate Z and the response variable Y obey the relations

$$Y = m_\theta(X) + \varepsilon, \qquad X = Z + \eta, \tag{12}$$

for some $\theta \in \Theta$. Here $m_\vartheta(x)$ is a known parametric function, nonlinear in x, from $\Theta \times \mathbb{R}^p$ to $\mathbb{R}$ with $E|m_\vartheta(X)| < \infty$, for all $\vartheta \in \Theta$. The r.v.'s ε, Z, η are assumed to be mutually independent, $E\varepsilon = 0$ and $E\eta = 0$. Unlike in the linear case, here we need to assume that the d.f. H of η is known. See Sect. 5 for the unknown H case.

Fix a θ for which (12) holds. Let $\nu_\vartheta(z) := E(m_\vartheta(X)|Z = z)$, $\vartheta \in \mathbb{R}^q$, $z \in \mathbb{R}^p$. Under (12), $E(Y|Z = z) \equiv \nu_\theta(z)$. Moreover, because H is known,

$$\nu_\vartheta(z) = \int m_\vartheta(z + s)dH(s)$$

is a known parametric regression function. Thus, under (12), we have the regression model

$$Y = \nu_\theta(Z) + \zeta, \qquad E(\zeta|Z = z) = 0, \qquad z \in \mathbb{R}^p.$$

Unlike in the linear case, the error ζ is no longer independent of Z in general.

To proceed further we assume there is a vector of p functions $\dot{m}_\vartheta(x)$ such that, with $\dot{\nu}_\vartheta(z) := \int \dot{m}_\vartheta(z + s)dH(s)$, for every $0 < b < \infty$,

$$\max_{1\le i\le n, n^{1/2}\|\vartheta-\theta\|\le b} n^{1/2}\big|\nu_\vartheta(Z_i) - \nu_\theta(Z_i) - (\vartheta - \theta)'\dot{\nu}_\theta(Z_i)\big| = o_p(1), \tag{13}$$

$$E\|\dot{\nu}_\theta(Z)\|^2 < \infty. \tag{14}$$

Let

$$L_z(x) := P(\zeta \le x|Z = z), \qquad x \in \mathbb{R},\ z \in \mathbb{R}^p.$$

Assume the following. For every $z \in \mathbb{R}^p$,

$$L_z(\cdot) \text{ is continuous and } L_z(x) = 1 - L_z(-x),\ \forall x \in \mathbb{R}^p. \tag{15}$$

Let G be as before and define

$$U(x, \vartheta) := n^{-1/2}\sum_{i=1}^{n} \dot{\nu}_\vartheta(Z_i)\big\{I(Y_i - \nu_\vartheta(Z_i) \le x) - I(-Y_i + \nu_\vartheta(Z_i) < x)\big\}$$

$$D(\vartheta) := \int \big\|U(x, \vartheta)\big\|^2 dG(x), \qquad \widehat{\theta} := \text{argmin}_\vartheta D(\vartheta).$$

In the case $q = p$ and $m_\theta(x) = x'\theta$, $\widehat{\theta}$ agrees with $\widetilde{\theta}$. Thus the class of estimators $\widehat{\theta}$, one for each G, is an extension of the class of estimators $\widetilde{\theta}$ from the linear case to the above nonlinear case.

Next, consider the extension of $\hat{\theta}_R$ to the above nonlinear model (12). Let $S_{i\vartheta}$ denote the rank of $Y_i - \nu_\vartheta(Z_i)$ among $Y_j - \nu_\vartheta(Z_j)$, $j = 1, \ldots, n$ and define

$$\mathcal{U}_n(u,\vartheta) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\nu}_\vartheta(Z_i)\big\{I(S_{i\vartheta}\le nu) - u\big\},$$

$$\mathcal{K}(\vartheta) := \int \|\mathcal{U}_n(u,\vartheta)\|^2 d\Psi(u), \quad \widehat{\theta}_R := \mathrm{argmin}_\vartheta \mathcal{K}(\vartheta).$$

The estimator $\widehat{\theta}_R$ gives an analog of $\hat{\theta}_R$ in the present set up.

Our goal here is to prove the asymptotic normality of $\widehat{\theta}$, $\widehat{\theta}_R$. This will be done by following the general method of Sect. 5.4 of Koul [16]. This method requires the two steps. In the first step we need to show that the defining dispersions $D(\vartheta)$ and $\mathcal{K}(\vartheta)$ are AULQ (asymptotically uniformly locally quadratic) in $\vartheta-\theta$ for $\vartheta\in\mathcal{N}_n(b) := \{\vartheta\in\Theta, n^{1/2}\|\vartheta-\theta\|\le b\}$, for every $0<b<\infty$. The second step requires to show that $n^{1/2}\|\widehat{\theta}-\theta\| = O_p(1) = n^{1/2}\|\widehat{\theta}_R-\theta\|$.

4.1 Asymptotic Distribution of $\widehat{\theta}$

In this subsection we shall derive the asymptotic normality of $\widehat{\theta}$. To state the needed assumptions for achieving this goal we need some more notation. Let $\nu_{nt}(z) := \nu_{\theta+n^{-1/2}t}(z)$, $\xi_{it} := \nu_{nt}(Z_i) - \nu_\theta(Z_i)$, $1\le i\le n$, $\dot{\nu}_{nt}(z) := \dot{\nu}_{\theta+n^{-1/2}t}(z)$, and $\dot{\nu}_{ntj}(z)$ denote the jth coordinate of $\dot{\nu}_{nt}(z)$, $1\le j\le q$, $t\in\mathbb{R}^q$. For any real number a, let $a^\pm = \max(0,\pm a)$ so that $a = a^+ - a^-$. Also, let $\beta_i(x) := I(\zeta_i\le x) - L_{Z_i}(x)$ and $\alpha_i(x,t) := I(\zeta_i\le x+\xi_{it}) - I(\zeta_i\le x) - L_{Z_i}(x+\xi_{it}) + L_{Z_i}(x)$.

Because $dG(x)\equiv -dG(-x)$ and $U(x,\vartheta)\equiv U(-x,\vartheta)$, we have

$$D(\vartheta) \equiv 2\int_0^\infty \big\|U(x,\vartheta)\big\|^2 dG(x) \equiv 2\widetilde{D}(\vartheta), \qquad \text{say.} \tag{16}$$

We are now ready to state our assumptions.

$$\int_0^\infty E\Big(\|\dot{\nu}_\theta(Z)\|^2\big(1-L_Z(x)\big)\Big)dG(x) < \infty. \tag{17}$$

$$\int_0^\infty E\Big(\|\dot{\nu}_{nt}(Z)-\dot{\nu}_\theta(Z)\|^2 L_Z(x)(1-L_Z(x)\Big)dG(x) \to 0,\ \forall\, t\in\mathbb{R}^q.$$

$$\sup_{\|t\|\le b, 1\le i\le n}\big\|\dot{\nu}_{nt}(Z_i)-\dot{\nu}_\theta(Z_i)\big\| \to_p 0. \tag{18}$$

Density ℓ_z of L_z exists for all $z\in\mathbb{R}^p$ such that (19)

$$0<\int \ell_z(x)dG(x)<\infty,\ \forall\, z\in\mathbb{R}^p,\ \ 0<\int E(\ell_Z^2(x))dG(x)<\infty,$$

$$\int E\big(\|\dot{\nu}_\theta(Z)\|^2 \ell_Z^j(x)\big) dG(x) < \infty, \;\; j = 1, 2.$$

$$\lim_{u \to 0} \int_{-\infty}^{\infty} \big(\ell_z(x+u) - \ell_z(x)\big)^j dG(x) = 0, \;\; j = 1, 2, \forall\, z \in \mathbb{R}^p. \tag{20}$$

$$E\Big(\int_{-|\xi_t(Z)|}^{|\xi_t(Z)|} \|\dot{\nu}_{nt}(Z)\|^2 \int_{-\infty}^{\infty} \ell_Z(x+u) dG(x) du\Big) \to 0, \; \forall\, t \in \mathbb{R}^q, \tag{21}$$

where $\xi_t(z) := \nu_{nt}(z) - \nu_\theta(z)$.

With $\Gamma_\theta(x) := E\big(\dot{\nu}_\theta(Z)\dot{\nu}_\theta(Z)'\ell_Z(x)\big)$, the matrix (22)

$$\Omega_\theta := \int_{-\infty}^{\infty} \Gamma_\theta(x)\Gamma_\theta(x)' dG(x) \text{ is positive definite.}$$

For every $\epsilon > 0$ there is a $\delta > 0$ and $N_\epsilon < \infty$ such that $\forall\, \|s\| \le b, n > N_\epsilon$,

$$P\Big(\sup_{\|t-s\|<\delta} \Big(n^{-1/2} \int \sum_{i=1}^{n} \big[\dot{\nu}_{ntj}^{\pm}(Z_i) - \dot{\nu}_{nsj}^{\pm}(Z_i)\big] \alpha_i(x,t) dG(x)\Big)^2 > \epsilon\Big) < \epsilon, \tag{23}$$

$$P\Big(\sup_{\|t-s\|<\delta} n^{-1} \int_0^{\infty} \Big\| \sum_{i=1}^{n} \{\dot{\nu}_{nt}(Z_i) - \dot{\nu}_{ns}(Z_i)\} \beta_i(x)\Big\|^2 dG(x) > \epsilon\Big) < \epsilon. \tag{24}$$

For every $\epsilon > 0$, $\alpha > 0$ there exists $N \equiv N_{\alpha,\varepsilon}$ and $b \equiv b_{\alpha,\epsilon}$ such that

$$P\Big(\inf_{\|t\|>b} D(\theta + n^{-1/2} t) \ge \alpha\Big) \ge 1 - \epsilon, \quad \forall\, n > N. \tag{25}$$

From now onwards we shall write ν and $\dot{\nu}$ for ν_θ and $\dot{\nu}_\theta$, respectively.

Remark 2 We shall now discuss the above assumptions when $m_\vartheta(x) = \vartheta' h(x)$, where $h = (h_1, \ldots, h_q)'$ is a vector of q function on $\mathbb{R}^p$ with $E\|h(X)\|^2 < \infty$, first for general G and then for some special cases of G. An example of this is the polynomial regression model with Berkson ME, where $p = 1, h_j(x) = x^j$, $j = 1, \ldots, q$. Let $\beta(z) := E(h(X)|Z = z)$. Then $\nu_\vartheta(z) = \vartheta'\beta(z)$ and $\dot{\nu}_\vartheta(z) \equiv \beta(z)$, a constant in ϑ. Therefore (13), (14), (18), (23) and (24) are all vacuously satisfied. The condition (25) also holds here, in a similar way as in the linear regression model, cf., Koul [16, Proof of Lemma 5.5.4, pp. 183–185]. Direct calculations show that (26)–(29) below imply the remaining assumptions (17), (19), (21) and (22), respectively.

$$\int_0^\infty E\Big(\|\beta(Z)\|^2\big(1 - L_Z(x)\big)\Big)dG(x) < \infty. \tag{26}$$

$\forall\, z \in \mathbb{R}^p$, density ℓ_z of L_z exists and satisfies (27)

$$0 < \int \ell_z^j(x)dG(x) < \infty, \;\; j = 1, 2, \quad 0 < \int E(\ell_Z^2(x))dG(x) < \infty,$$

$$\int E\big(\|\beta(Z)\|^2 \ell_Z^j(x)\big)dG(x) < \infty, \;\; j = 1, 2, \quad \text{and (20) holds.}$$

$$E\Big(\int_{n^{-1/2}b\|\beta(Z)\|}^{|n^{-1/2}b\|\beta(Z)\|} \|\beta(Z)\|^2 \int_{-\infty}^{\infty} \ell_Z(x+u)dG(x)du\Big) \to 0, \tag{28}$$

for every $0 < b < \infty$.

With $\mathcal{B}(x) := E\big(\beta(Z)\beta(Z)'\ell_Z(x)\big)$, the matrix (29)

$$\int_{-\infty}^{\infty} \mathcal{B}(x)\mathcal{B}(x)'dG(x) \text{ is positive definite.}$$

Consider further the case $G(x) \equiv x$. Let $\sigma := (E\varepsilon^2)^{1/2}$. Assume

$$\text{(a)} \;\; E\|h(X)\|^3 < \infty, \quad E\zeta^2 < \infty. \quad \text{(b)} \;\; C := \sup_{x\in\mathbb{R}, z\in\mathbb{R}^p} \ell_z(x) < \infty. \tag{30}$$

Then $E\|\beta(Z)\|^j \le E\|h(X)\|^j < \infty$, $j = 1, 2, 3$. Let $\gamma(z) := 2\|\theta\|\|\beta(z)\| + \sigma$. Then

$$\begin{aligned} E\big(|\zeta|\big|Z = z\big) &= E\big(|Y - \theta'\beta(Z)|\big|Z = z\big) \\ &= E\big(|\theta'h(X) + \varepsilon - \theta'\beta(Z)|\big|Z = z\big) \le \gamma(z), \quad \forall\, z \in \mathbb{R}^p. \end{aligned} \tag{31}$$

Hence

$$\begin{aligned} &\int_0^\infty E\Big(\|\beta(Z)\|^2\big(1 - L_Z(x)\big)\Big)dx \\ &\le E\Big(\|\beta(Z)\|^2 E\big(|\zeta|\big|Z\big)\Big) \le E\Big(\|\beta(Z)\|^2\gamma(Z)\Big) \\ &\le 2\|\theta\|E\big(\|\beta(Z)\|^3\big) + \sigma E\big(\|\beta(Z)\|^2\big) < \infty, \end{aligned}$$

thereby showing that (26) is satisfied. The assumption (30)(b) and $\ell_z(x)$ being a density in x for each z and Theorem 9.5 of Rudin [18] readily imply (27) here. The left hand side of (28) equals $2n^{-1/2}bE\big(\|\beta(Z)\|^3\big) \to 0$, by (30)(a).

Next, consider the case $G(x) = \delta_0(x)$- measure degenerate at zero. Assume

$$\lim_{u\to 0} \ell_z(u) = \ell_z(0) > 0,\ \forall z \in \mathbb{R}^p, \quad 0 < E\ell_Z^2(0) < \infty, \tag{32}$$
$$E\big(\|\beta(Z)\|^2 \ell_Z^j(0)\big) < \infty, \quad j = 1, 2.$$

Then the left hand side of (26) equals $(1/2)E\|\beta(Z)\|^2 < E\|h(X)\|^2 < \infty$. Condition (27) is trivially satisfied and the left hand side of (28) equals

$$E\Big(\|\beta(Z)\|^2\big[L_Z(n^{-1/2}b\|\beta(Z)\|) - L_Z(-n^{-1/2}b\|\beta(Z)\|)\big]\Big) \to 0,$$

by the DCT and the continuity of $L_z(\cdot)$, for each z.

To summarize, in the case $m_\vartheta(x) = \vartheta' h(X)$ and $G(x) \equiv x$, assumptions (30)(a), (b) and $\int \mathcal{B}(x)\mathcal{B}(x)'dx$ being positive definite imply all of the above assumptions (13), (14) and (17)–(25). Similarly, in the case $m_\vartheta(x) = \vartheta' h(X)$ and $G(x) \equiv \delta_0(x)$, $E\|h(X)\|^2 < \infty$, (32) and $\mathcal{B}(0)\mathcal{B}(0)'$ being positive definite imply all these conditions.

Remark 3 Because of the importance of the estimators $\widehat{\theta}$ when $G(x) = x$, and $G(x) = \delta_0(x)$, it is of interest to give some simple sufficient conditions for a general m_ϑ that imply the given assumptions for these two estimators.

Suppose G satisfies $dG(x) \equiv g(x)dx$, where $g_\infty := \sup_{x\in\mathbb{R}} g(x) < \infty$. Note that $G(x) \equiv x$ corresponds to the case $g(x) \equiv 1$. Consider the following assumptions.

$$\text{(a)}\ \ E\big\|\dot m_\theta(X)\big\|^4 < \infty, \tag{33}$$
$$\text{(b)}\ \ E\big\|\dot m_{\theta+n^{-1/2}t}(X) - \dot m_\theta(X)\big\|^2 \to 0,\ \forall t \in \mathbb{R}^q.$$

Density ℓ_z of L_z exists for all $z \in \mathbb{R}^p$ and satisfies (34)

$$0 < \int \ell_z^2(x)dx < \infty, \forall z \in \mathbb{R}^p, \quad 0 < \int E(\ell_Z^2(x))dx < \infty,$$

$$0 < \int E\big(\|\dot\nu(Z)\|^2 \ell_Z^2(x)\big)dx < \infty.$$

$$E\big(\|\dot\nu_{nt}(Z)\|^2 |\nu_{nt}(Z) - \nu(Z)|\big) \to 0, \quad \forall t \in \mathbb{R}^q. \tag{35}$$

Because $\big\|\dot\nu(Z)\big\|^j \le E\big(\|\dot m_\theta(X)\big\|^j\big|Z\big)$, $E\big\|\dot\nu(Z)\big\|^j \le E\big\|\dot m_\theta(X)\big\|^j < \infty,\ j = 1,$ 2, 3, 4, by (33)(a). Similarly, for every $t \in \mathbb{R}^q$,

$$E\big\|\dot\nu_{nt}(Z) - \dot\nu(Z)\big\|^2 \le E\big\|\dot m_{nt}(X) - \dot m_\theta(X)\big\|^2 \to 0, \qquad \text{by (34)(b)}. \tag{36}$$

Next, similar to (31), $E\big(|\zeta|\big|Z\big) = E\big(|Y - \nu(Z)|\big|Z\big) \le 2|\nu(Z)| + \sigma$ implies that the left hand side of (17) is bounded from the above by

$$
\begin{aligned}
g_\infty E\Big(\|\dot\nu(Z)\|^2 E\big(|\zeta|\big|Z\big)\Big) &\le g_\infty E\Big(\|\dot\nu(Z)\|^2\big(2|\nu(Z)|+\sigma\big)\Big)\\
&\le g_\infty\big[2E^{1/2}(\|\dot\nu(Z)\|^4)E^{1/2}(m_\theta^2(X))+\sigma E\|\dot\nu(Z)\|^2\big]\\
&< \infty,
\end{aligned}
$$

by (33)(a), thereby verifying (17) here. Similarly, with C denoting the above upper bound, for every $t \in \mathbb{R}^q$, the left hand side of (18) is bounded from the above $CE\Big(\|\dot\nu_{nt}(Z)-\dot\nu_\theta(Z)\|^2\Big) \to 0$, by (36). The left hand side of (21) is bounded from the above by $2g_\infty E\big(\|\dot\nu_{nt}(Z)\|^2|\nu_{nt}(Z)-\nu(Z)|\big) \to 0$, by (35).

In other words, in the case G has bounded Lebesgue density, conditions (33)–(35) imply assumptions (14), (17), (18), (19), (20), and (21). Not much simplification occurs in the remaining assumptions (18) and (22)–(25). See Remark 2 for some special cases.

Next consider the case when $G(x) = \delta_0(x)$ and the following assumptions.

$$
\sup_{x\in\mathbb{R}} \ell_z(x) < \infty,\;\; 0 < \lim_{u\to 0} \ell_z(u) = \ell_z(0) < \infty,\;\; \forall\, z \in \mathbb{R}^p. \tag{37}
$$

$$
\Gamma_\theta(0) \text{ is positive definite.} \tag{38}
$$

In this case (33), (35), (37) and (38) together imply the assumptions (14), (17)–(22). Not much simplification occurs in the remaining three assumptions (23)–(25), except in some special cases as in Remark 2.

We now resume the discussion about the asymptotic normality of $\widehat\theta$. First, we show that $E(D(\theta)) < \infty$, so that by the Markov inequality, $D(\theta)$ is bounded in probability. To see this, by (15), $EU(x,\theta) \equiv 0$ and, for $x \ge 0$,

$$
\begin{aligned}
E\|U(x,\theta)\|^2 &= E\Big(\big\|\dot\nu(Z)\big\|\big\{I(\zeta\le x)-I(\zeta>-x)\big\}\Big)^2\\
&= 2E\Big(\big\|\dot\nu(Z)\|^2\big(1-L_Z(x)\big)\Big).
\end{aligned}
$$

By the Fubini Theorem, (16) and (17),

$$
E(D(\theta)) = 2E(\widetilde D(\theta)) = 4\int_0^\infty E\Big(\big\|\dot\nu(Z)\|^2\big(1-L_Z(x)\big)dG(x) < \infty. \tag{39}
$$

To state the AULQ result for D, we need some more notation. Let

$$
W(x,0) := n^{-1/2}\sum_{i=1}^n \dot\nu(Z_i)\big\{I\big(\zeta_i\le x\big)-L_{Z_i}(x)\big\}, \tag{40}
$$

$$
T_n := \int_{-\infty}^{\infty} \Gamma_\theta(x)\big\{W(x,0)+W(-x,0)\big\}dG(x),\quad \widehat t = -\Omega_\theta^{-1}T_n/2,
$$

where $\Gamma_\theta(x)$ and Ω_θ are as in (22). We are ready to state the following lemma.

Lemma 2 *Suppose the above set up and assumptions (17)–(24) hold. Then for every $b < \infty$,*

$$\sup_{\|t\|\le b} \big| D(\theta + n^{-1/2}t) - D(\theta) - 4T_n' t - 4t'\Omega_\theta t \big| \to_p 0. \tag{41}$$

If in addition (25) holds, then, with Σ_θ given at (45) below,

$$\begin{aligned} &\text{(a)} \quad \|n^{1/2}(\widehat{\theta} - \theta) - \widehat{t}\,\| \to_p 0. \\ &\text{(b)} \quad n^{1/2}(\widehat{\theta} - \theta) \to_D N\big(0, 4^{-1}\Omega_\theta^{-1}\Sigma_\theta\Omega_\theta^{-1}\big). \end{aligned} \tag{42}$$

Proof The proof of (41) appears in Sect. 6. The proof of the claim (42)(a), which uses (25), (39) and (41), is similar to that of Theorem 5.4.1 of Koul [16], where (25) and (39) are used to show that $n^{1/2}\|\widehat{\theta} - \theta\| = O_p(1)$.

Define, for $y \in \mathbb{R}$, $u \in \mathbb{R}^p$,

$$\psi_u(y) := \int_{-\infty}^{y} \ell_u(x)dG(x), \quad \varphi_u(y) := \psi_u(-y) - \psi_u(y). \tag{43}$$

By (19), $0 < \psi_u(y) \le \psi_u(\infty) = \int_{-\infty}^{\infty} \ell_u(x)dG(x) < \infty$, for all $u \in \mathbb{R}^p$. Thus for each u, $\psi_u(y)$ is an increasing continuous bounded function of y and $\psi_u(-y) \equiv \psi_u(\infty) - \psi_u(y)$, and $\varphi_u(y) = \psi_u(\infty) - 2\psi_u(y)$, for all $y \in \mathbb{R}$.

By (15), $E(\varphi_u(\zeta)|Z = z) = 0$, for all $u, z \in \mathbb{R}^p$. Let

$$\begin{aligned} C_z(u, v) &:= \mathrm{Cov}\big[\big(\varphi_u(\zeta), \varphi_v(\zeta)\big)\big|Z = z\big] = 4\mathrm{Cov}\big[\big(\psi_u(\zeta), \psi_v(\zeta)\big)\big|Z = z\big], \\ \mathcal{K}(u, v) &:= E\big(\dot\nu(Z)\dot\nu(Z)'C_Z(u, v)\big), \qquad u, v \in \mathbb{R}^p. \end{aligned}$$

Next let $\mu(z) := \dot\nu(z)\dot\nu(z)'$, Q denote the d.f. of Z and rewrite $\Gamma_\theta(x) = E\big(\dot\nu_\theta(Z)\,\dot\nu_\theta(Z)'\ell_Z(x)\big) = \int \mu(z)\ell_z(x)dQ(z)$. By the Fubini Theorem,

$$\begin{aligned} T_n &:= \int_{-\infty}^{\infty} \Gamma_\theta(x)\big\{W(x, 0) + W(-x, 0)\big\}dG(x) \\ &= \int \int_{-\infty}^{\infty} \mu(z)\big\{W(x, 0) + W(-x, 0)\big\}\ell_z(x)dG(x)dQ(z) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \mu(z)\dot\nu(Z_i)\varphi_z(\zeta_i)dQ(z). \end{aligned} \tag{44}$$

Clearly, $ET_n = 0$ and by the Fubini Theorem, the covariance matrix of T_n is

$$\begin{aligned}\Sigma_\theta &:= ET_nT_n' \\ &= E\Big\{\Big(\int \mu(z)\dot{\nu}(Z)\varphi_z(\zeta)dQ(z)\Big)\Big(\int \mu(v)\dot{\nu}(Z)\varphi_v(\zeta)dQ(v)\Big)'\Big\} \\ &= \int\int \mu(z)\mathcal{K}(z,v)\mu(v)'dQ(z)dQ(v).\end{aligned} \tag{45}$$

Thus T_n is a $p \times 1$ vector of independent centered finite variance r.v.'s. By the classical CLT, $T_n \to_D N(0, \Sigma_\theta)$. Hence, the minimizer $\widetilde{t}$ of the approximating quadratic form $D(\theta) + 4T_n t + 4t'\Omega_\theta t$ with respect to t satisfies $\tilde{t} = -\Omega_\theta^{-1}T_n/2 \to_D N\big(0, 4^{-1}\Omega_\theta^{-1}\Sigma_\theta\Omega_\theta^{-1}\big)$. The claim (42)(b) now follows from this result and (42)(a). □

4.2 Asymptotic Distribution of $\widehat{\theta}_R$

In this subsection we shall establish the asymptotic normality of $\widehat{\theta}_R$. For this we need the following assumptions, where $\mathcal{U}(b) := \{t \in \mathbb{R}^q;\ \|t\| \le b\}$, and $0 < b < \infty$.

$$\ell_z \text{ is uniformly continuous and bounded for every } z \in \mathbb{R}^p. \tag{46}$$

$$n^{-1}\sum_{i=1}^n E\|\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\|^2 \to 0, \quad \forall\, t \in \mathcal{U}(b). \tag{47}$$

$$n^{-1/2}\sum_{i=1}^n \|\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\| = O_p(1), \quad \forall\, t \in \mathcal{U}(b). \tag{48}$$

$\forall\, \epsilon > 0,\ \exists\, \delta > 0$ and $n_\epsilon < \infty$ such that for each $s \in \mathcal{U}(b)$, $\forall\, n > n_\epsilon$,

$$P\Big(\sup_{t\in\mathcal{U}(b);\,\|t-s\|\le\delta} n^{-1/2}\sum_{i=1}^n \|\dot{\nu}_{nt}(Z_i) - \dot{\nu}_{ns}(Z_i)\| \le \epsilon\Big) > 1 - \epsilon. \tag{49}$$

$\forall\, \epsilon > 0,\ 0 < \alpha < \infty,\ \exists\, N \equiv N_{\alpha,\epsilon}$ and $b \equiv b_{\epsilon,\alpha}$ such that

$$P\Big(\inf_{\|t\|>b} \mathcal{K}(\theta + n^{-1/2}t) \ge \alpha\Big) \ge 1 - \epsilon, \quad \forall\, n > N. \tag{50}$$

Let

$$\bar{\dot{\nu}} := n^{-1}\sum_{i=1}^n \dot{\nu}(Z_i), \quad \dot{\nu}^c(Z_i) := \dot{\nu}(Z_i) - \bar{\dot{\nu}},$$

$$\widehat{\Gamma}_\theta(u) := E\Big(\dot{\nu}^c(Z)\dot{\nu}^c(Z)'\ell_Z(L_Z^{-1}(u))\Big), \quad \widehat{\Omega}_\theta := \int_0^1 \widehat{\Gamma}_\theta(u)\widehat{\Gamma}_\theta(u)'d\Psi(u),$$

$$\widehat{\mathcal{U}}(u) := n^{-1/2}\sum_{i=1}^n \dot{\nu}^c(Z_i)\big\{I(L_{Z_i}(\zeta_i) \le u) - u\big\}, \quad 0 \le u \le 1,$$

$$\widehat{T}_n := \int_0^1 \widehat{\Gamma}_\theta(u)\widehat{\mathcal{U}}(u)d\Psi(u), \quad \widehat{\mathcal{K}}(t) := \int_0^1 \big\|\widehat{\mathcal{U}}(u)\big\|^2 d\Psi(u) + 2\widetilde{T}_n't + t'\widetilde{\Omega}_\theta t.$$

We need to have an alternate representation of the covariance matrix of $\widehat{T}_n$. Let, for $z \in \mathbb{R}^p, \quad 0 \le v \le 1$,

$$\kappa_z(v) := \int_0^v \ell_z(L_z^{-1}(u))d\Psi(u), \quad k_z^c(v) = k_z(v) - \int_0^1 k_z(u)du.$$

By (46), κ_z is a uniformly continuous increasing and bounded function on [0, 1], for all $z \in \mathbb{R}^p$. Let U denote a uniform [0, 1] r.v. Conditionally, given Z, $L_Z(\zeta) \sim_D U$. Hence, $E\big(k_z\big(L_Z(\zeta)\big)\big|Z\big) = Ek_z(U)$ so that $E\big(k_z^c(L_Z(\zeta))\big|Z\big) = Ek_z^c(U) = 0$, a.s. Let $\mu^c(z) := \dot{\nu}^c(z)\dot{\nu}^c(z)'$. Argue as for (44) and use the facts that $\sum_{i=1}^n \dot{\nu}_c(Z_i) \equiv 0$ and $\int_0^1 u dk_z(u) = k_z(1) - \int_0^1 k_z(u)du$ to obtain that

$$\widehat{T}_n = -n^{-1/2}\sum_{i=1}^n \int \mu^c(z)\dot{\nu}^c(Z_i)\kappa_z^c\big(L_{Z_i}(\zeta_i)\big)dQ(z).$$

Define

$$\widehat{C}_z(s,t) := E\big[\kappa_s^c(L_Z(\zeta))\kappa_t^c(L_Z(\zeta))\big|Z = z\big] = E\big[\kappa_s^c(U)\kappa_t^c(U)\big],$$
$$\widehat{K}(s,t) := E\big(\dot{\nu}^c(Z)\dot{\nu}^c(Z)'\widehat{C}_Z(s,t)\big).$$

Then argue as in (45) to obtain

$$\widehat{\Sigma}_\theta := E\widehat{T}_n\widehat{T}_n' = \int\int \mu^c(z)\widehat{K}(z,v)\mu^c(v)'dQ(z)dQ(v).$$

We are now ready to state the following asymptotic normality result for $\widehat{\theta}_R$.

Lemma 3 *Suppose the nonlinear Berkson measurement error model (12) and the assumptions (13), (14), (46)–(49) hold. Then the following holds.*

$$\sup_{\|t\|\le b} \left|\mathcal{K}(\theta + n^{-1/2}t) - \widehat{K}(t)\right| = o_p(1). \tag{51}$$

In addition, if (50) *holds and* $\widehat{\Omega}_\theta$ *is positive definite then* $n^{1/2}(\widehat{\theta}_R - \theta) \to_d N\left(0, \widehat{\Omega}_\theta^{-1}\widehat{\Sigma}_\theta\widehat{\Omega}_\theta^{-1}\right)$.

The proof of this lemma is similar to that of Theorem 1.2 of Koul [15], hence no details are given here. Assumption (50) is used to show that $n^{1/2}\|\widehat{\theta}_R - \theta\| = O_p(1)$.

Remark 4 As in Remark 1, let $m_\theta(x) = \theta' h(x)$. Then $\nu_\vartheta(z) = \vartheta'\beta(z)$, where $\beta(z) := E\big(h(X)|Z = z\big)$. Thus $\dot{\nu}_\vartheta(z) \equiv \beta(z)$ and the assumptions (47)–(49) are vacuously satisfied. The assumption (50) is shown to be satisfied by an argument similar to the one used in the proof of Lemma 5.4.4 of Koul [16, pp. 183–185]. This proof uses the monotonicity in t for every unit vector $e \in \mathbb{R}^p$ of simple linear rank statistics based on the ranks of $Y_i - te'h(X_i)$, $1 \le i \le n$, see Hájek [10, Theorem II.7E].

For the asymptotic normality of $\widehat{\widehat{\theta}}_R$ here, one only needs (46) and Ψ to be a d.f. such that $\widehat{\Omega}$ is positive definite. Note that here $\mu^c(z) = \beta^c(z) := \beta(z) - \bar{\beta}$, $\bar{\beta} := n^{-1}\sum_{i=1}^n \beta(Z_i)$ and

$$\widehat{K}(s,t) := E\big(\beta^c(Z)\beta^c(Z)'\widehat{C}_Z(s,t)\big),$$
$$\widehat{\Sigma} = \int\int \beta^c(z)\beta^c(z)'\widehat{K}(z,v)\beta^c(v)\beta^c(v)'dQ(z)dQ(v),$$
$$\widehat{\Gamma}(u) := E\Big(\beta^c(Z)\beta^c(Z)'\ell_Z(L_Z^{-1}(u))\Big), \quad \widehat{\Omega} = \int_0^1 \widehat{\Gamma}(u)\widehat{\Gamma}(u)'d\Psi(u),$$

do not depend on θ. Clearly, these assumptions are far less stringent than those needed for the asymptotic normality of $\widehat{\theta}$ corresponding to $G(x) \equiv x$.

5 M.D. Estimators with Validation Data

In this section we develop the m.d. estimators of Sect. 4 when the d.f. H of the Berkson ME η is unknown but a validation data set is available. Not knowing H renders ν_θ to be an unknown function. Validation data is used to estimate this function, which in turn is used to define m.d. estimators.

Let N be a known positive integer. A set of r.v.'s $\{(\tilde{X}_k, \tilde{Z}_k), k = 1, ..., N\}$ is said to be validation data if these r.v.'s are independent of the original sample and both $\tilde{Z}_k$ and $\tilde{X}_k$ are observable and obey the model (12). Besides having the primary data set $\{(Y_i, Z_i), 1 \le i \le n\}$, we assume that a validation data set of the covariate $\{(\tilde{X}_k, \tilde{Z}_k), 1 \le k \le N\}$ is available. Then $\tilde{\eta}_k := \tilde{X}_k - \tilde{Z}_k$, $1 \le k \le N$ are observable and their empirical d.f. $H_N(s) := N^{-1}\sum_{k=1}^N I(\tilde{\eta}_k \le s)$, $s \in \mathbb{R}$, provides an estimate of H.

Under (13)–(15), we have the following estimates of ν_θ and $\dot{\nu}_\theta$.

$$\hat{\nu}_\vartheta(z) := N^{-1}\sum_{k=1}^{N} m_\vartheta(z+\tilde{\eta}_k), \qquad \hat{\dot{\nu}}_\vartheta(z) := N^{-1}\sum_{k=1}^{N} \dot{m}_\vartheta(z+\tilde{\eta}_k).$$

An analog of $\widehat{\theta}$ in the current set up is defined as follows. Let

$$\widehat{U}(x,\vartheta) := n^{-1/2}\sum_{i=1}^{n} \hat{\dot{\nu}}_\vartheta(Z_i)\{I(Y_i - \hat{\nu}_\vartheta(Z_i) \le x) - I(-Y_i + \hat{\nu}_\vartheta(Z_i) < x)\},$$

$$D_1(\vartheta) := \int \|\widehat{U}(x,\vartheta)\|^2 dG(x), \qquad \widehat{\theta}_1 = \text{argmin}_\vartheta D_1(\vartheta).$$

To define the analog of $\hat{\theta}_R$ here, let $\tilde{S}_{i\vartheta}$ be the rank of $Y_i - \hat{\nu}_\vartheta(Z_i)$ among $Y_j - \hat{\nu}_\vartheta(Z_j)$, $1 \le j \le n$ and define

$$\tilde{\mathcal{U}}_n(u,\vartheta) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{\dot{\nu}}_\vartheta(Z_i)\{I(\tilde{S}_{i\vartheta} \le nu) - u\}, \quad 0 \le u \le 1,$$

$$\tilde{\mathcal{K}}(\vartheta) := \int_0^1 \|\tilde{\mathcal{U}}_n(u,\vartheta)\|^2 d\Psi(u), \quad \tilde{\theta}_R := \text{argmin}_\vartheta \tilde{\mathcal{K}}(\vartheta).$$

The asymptotic distributions of $\widehat{\theta}_1$ and $\tilde{\theta}_R$ as $n \wedge N \to \infty$ are described in the next two subsections. In their derivations, the $\lim(n/N)$ of the ratio n/N plays an important role. Some of the proofs are similar to those of $\widehat{\theta}$ and $\widehat{\theta}_R$. Some key steps of the proof can be found in the Appendix.

5.1 Asymptotic Distribution of $\widehat{\theta}_1$

In this subsection we derive the asymptotic distribution of $\widehat{\theta}_1$. In addition to (13)–(15) and (17)–(25), the following assumptions are needed, where $\Delta_\vartheta(z) := \hat{\nu}_\vartheta(z) - \nu_\vartheta(z)$ and θ is as in (12).

$$E\Big\| E\big\{\dot{m}_\theta(X)[\nu_\theta(Z) - m_\theta(X)]|Z\big\}\Big\|^2 < \infty, \tag{52}$$
$$E\Big\| E\big\{\dot{m}_\theta(X)[\nu_\theta(Z) - m_\theta(X)]|\eta\big\}\Big\|^2 < \infty.$$

The matrix (53)

$$\Sigma_1 := \text{Cov}\Big(E\big[\int\int \mu(z)\dot{\nu}_\theta(Z)\ell_z(x)\ell_Z(x)[m_\theta(Z+\eta) - \nu_\theta(Z)]dx\,dQ(z)\big|\eta\big]\Big)$$

is positive definite.

$$\lambda := \lim(n/N) \ge 0. \tag{54}$$

$$\max_{1\le i\le n}\Big|N^{-1}\sum_{k=1}^{N} m_\theta(Z_i+\tilde{\eta}_k)-\nu_\theta(Z_i)\Big| = o_p(1). \tag{55}$$

$$E\Big\{\|\dot{\nu}_\theta(Z)\|^2\big(m_\theta(X)-\nu_\theta(Z)\big)^2\Big\} < \infty. \tag{56}$$

$$\int_0^\infty E\Big(\|\hat{\dot{\nu}}_\theta(Z)\|^2\big(1-L_Z(x\pm\Delta_\theta(Z))\big)\Big)dG(x) < \infty. \tag{57}$$

$$\int_0^\infty E\Big(\|\hat{\dot{\nu}}_{nt}(Z)-\hat{\dot{\nu}}_\theta(Z)\|^2 L_Z(x\pm\Delta_\theta(Z)) \times(1-L_Z(x\pm\Delta_\theta(Z))\Big)dG(x)\to 0,\ \ \forall t\in\mathbb{R}^q. \tag{58}$$

We also assume that (18)–(24) and (25) hold with $\dot{\nu}_{nt}$, $\dot{\nu}_\theta$ and D replaced by $\hat{\dot{\nu}}_{nt}$, $\hat{\dot{\nu}}$ and D_1, respectively. We denote these assumptions as (18)*–(25)*.

Here we discuss some sufficient conditions for the above assumptions. By the C-S inequality, both the expressions of (52) are bounded from the above by $2E\big\|\dot{m}_\theta(X)\big\|^2 E\big|m_\theta(X)\big|^2$. Thus (52) is implied by (33)(a) and having $E\big|m_\theta(X)\big|^2 < \infty$.

Next, under (33)(a), (57) is trivially satisfied when $G(x)\equiv\delta_0(x)$. In the case $dG(x)=g(x)dx$ with $g_\infty := \sup_{y\in\mathbb{R}} g(y) < \infty$, (57) is implied by (33)(a) and the following conditions.

$$E\big(\|\dot{m}_\theta(X)\|^2|m_\theta(X)|\big) < \infty. \tag{59}$$

To see this, note that $E\big(|\Delta_\theta(Z)|\big|Z\big)\le 2E\big(|m_\theta(X)|\big|Z\big)$, and $\|\hat{\dot{\nu}}_\theta(Z)\|^2 \le N^{-1}\sum_{k=1}^N \|\dot{m}_\theta(Z+\eta_k)\|^2$ so that $E\|\hat{\dot{\nu}}_\theta(Z)\|^2\le E\|\dot{m}_\theta(X)\|^2$. Now argue as in Remark 4.2 and use these facts to obtain that the left hand side of (57) is bounded from the above by

$$\begin{aligned} &g_\infty E\Big(\|\hat{\dot{\nu}}_\theta(Z)\|^2 E\Big(|\zeta|+|\Delta_\theta(Z)|\big|Z\Big)\Big)\\ &\le g_\infty\Big[CE\big\|\dot{m}_\theta(X)\|^2+2E\big(\|\dot{m}_\theta(X)\|^2|m_\theta(X)|\big)\Big]<\infty,\end{aligned}$$

by (33)(a) and (59). Similarly, the left hand side of (58) is bounded from the above by a constant multiple of

$$E\big\|\hat{\dot{\nu}}_{nt}(Z)-\hat{\dot{\nu}}_\theta(Z)\big\|^2 \le E\big\|\dot{m}_{\theta+n^{-1/2}t}(X)-\dot{m}_\theta(X)\big\|^2 \to 0,\ \ \text{by (34)(b).}$$

We now turn to proving the asymptotic normality of $\widehat{\theta}_1$. Similar to Sect. 4.1, we first prove that $E(D_1(\theta)) < \infty$. Recall $\Delta_\vartheta(z) := \hat{\nu}_\vartheta(z) - \nu_\vartheta(z)$ and rewrite

$$\widehat{U}(x,\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\hat{\nu}}_\theta(Z_i) \Big\{ I(\zeta_i \le x + \Delta_\theta(Z_i)) - I(-\zeta_i < x - \Delta_\theta(Z_i)) \Big\}.$$

By the independence of the primary and validation data and a conditioning argument, for every $x > 0$,

$$\begin{aligned} E\|\widehat{U}(x,\theta)\|^2 &= E\Big(\|\hat{\hat{\nu}}_\theta(Z)\| \{ I(\zeta \le x + \Delta_\theta(Z)) - I(-\zeta < x - \Delta_\theta(Z)) \} \Big)^2 \\ &= E\Big(\|\hat{\hat{\nu}}_\theta(Z)\|^2 \{ 1 - L_Z(x + \Delta_\theta(Z)) + 1 - L_Z(x - \Delta_\theta(Z)) \} \Big). \end{aligned}$$

Hence by (56), $ED_1(\theta) < \infty$.

Next we sketch the proof of the AULQ property of $D_1(\vartheta)$. Define

$$\begin{aligned} \tilde{W}(x,0) &:= n^{-1/2} \sum_{i=1}^{n} \hat{\hat{\nu}}(Z_i) \{ I(\zeta_i \le x + \Delta_\theta(Z_i)) - L_{Z_i}(x) \}, \\ \tilde{T}_n &:= \int \Gamma_\theta(x) \{ \tilde{W}(x,0) + \tilde{W}(-x,0) \} dG(x). \end{aligned}$$

In the Appendix, we show that $\tilde{T}_n$ is approximated by a U-statistic based on the two independent samples. Theorem 6.1.4 in Lehmann [17] yields

$$\begin{aligned} &\text{(a)} \quad \tilde{T}_n \to N(0, \Sigma_\theta + 4\lambda\Sigma_1), \quad \lambda < \infty, \\ &\text{(b)} \quad \sqrt{N/n}\, \tilde{T}_n \to N(0, 4\Sigma_1), \quad \lambda = \infty. \end{aligned} \tag{60}$$

Next, the assumptions (54)–(58) and (18)*–(24)* ensure that the analog of Lemma 5 holds here also. Hence (41) with T_n and $D(\vartheta)$ replaced by $\widetilde{T}_n$ and $D_1(\vartheta)$, respectively, holds. Moreover, analog of (42) can be shown to hold in a similar manner as in Sect. 4 under (25)*. Consequently, the asymptotic distribution of $\widehat{\theta}_1$ based on data sets $\{(Y_i, Z_i), 1 \le i \le n\}$ and $\{(\tilde{X}_k, \tilde{Z}_k), 1 \le k \le N\}$ described in the following lemma.

Lemma 4 *Suppose model* (12) *with* H *unknown holds and an independent validation data* $\{(\tilde{X}_k, \tilde{Z}_k), 1 \le k \le N\}$ *obeying* (12) *is available. In addition assume that* (17)–(25) *and* (52)–(57) *hold. Then*

$$\sqrt{n}(\widehat{\theta}_1 - \theta) \to N(0, 4^{-1}\Omega_\theta^{-1}(\Sigma_\theta + 4\lambda\Sigma_1)\Omega_\theta^{-1}), \ \text{for } 0 \le \lambda < \infty;$$

$$\sqrt{N}(\widehat{\theta}_1 - \theta) \to N(0, 16^{-1}\Omega_\theta^{-1}\Sigma_1\Omega_\theta^{-1}), \ \text{for } \lambda = \infty.$$

The above result shows that the estimation step of regression function $\nu_\theta(z)$ due to the unknown distribution H introduces more variation in the asymptotic distribution of the m.d. estimators. Moreover, the limiting ratio λ of the sample sizes plays a role in the additional variation. When $\lambda = \lim n/N = 0$, the additional covariance term vanishes, therefore it reduces to the case when the ME distribution is known. In other words, when the validation sample size N is sufficiently large, compared to the primary sample size n, both $\hat{\theta}$ and $\widehat{\theta}_1$ achieve the same asymptotic efficiency. On the other hand, when $\lambda = \infty$, i.e., when the validation data size is very limited compared to the primary data size, the estimation consistency rate is restricted to $\sqrt{N}$ instead of $\sqrt{n}$.

5.2 Asymptotic Distribution of $\tilde{\theta}_R$

In this subsection we present the asymptotic distribution of the class of estimators $\tilde{\theta}_R$. First, we provide the additional assumptions. Let $\hat{\dot{\nu}}_{nt}(z) = N^{-1}\sum_{k=1}^{N} \dot{\nu}_{\theta+n^{-1/2}t}(z + \tilde{\eta}_k)$. Consider the following assumptions.

$$n^{-1}\sum_{i=1}^{n} E\|\hat{\dot{\nu}}_{nt}(Z_i) - \hat{\dot{\nu}}(Z_i)\|^2 \to 0, \quad \forall\, t \in \mathcal{U}(b). \tag{61}$$

$$n^{-1/2}\sum_{i=1}^{n} \|\hat{\dot{\nu}}_{nt}(Z_i) - \hat{\dot{\nu}}(Z_i)\| = O_p(1), \quad \forall\, t \in \mathcal{U}(b). \tag{62}$$

$\forall\, \epsilon > 0,\ \exists\, \delta > 0$ and $n_\epsilon < \infty$ such that for each $s \in \mathcal{U}(b)$, $\forall\, n > n_\epsilon$,

$$P\Big(\sup_{t\in\mathcal{U}(b);\,\|t-s\|\le\delta} n^{-1/2}\sum_{i=1}^{n} \|\hat{\dot{\nu}}_{nt}(Z_i) - \hat{\dot{\nu}}_{ns}(Z_i)\| \le \epsilon\Big) > 1 - \epsilon. \tag{63}$$

For every $\epsilon > 0,\ 0 < \alpha < \infty$, there exist an N_ϵ and $b \equiv b_{\epsilon,\alpha}$ such that

$$P\Big(\inf_{\|t\|>b} \tilde{\mathcal{K}}(\theta + n^{-1/2}t) \ge \alpha\Big) \ge 1 - \epsilon, \quad \forall\, n > N_\epsilon. \tag{64}$$

The matrix (65)

$$\Sigma_2 := \mathrm{Cov}\Big(E\Big[\int\int \mu^c(z)\{\dot{\nu}_\theta(Z) - E(\dot{\nu}_\theta(Z))\}\ell_z(x)\ell_Z(x) \times\{m_\theta(Z+\eta) - \nu_\theta(Z)\}dx\,dQ(z)\Big|\eta\Big]\Big)$$

is positive definite.

Next, define $\tilde{\dot{\nu}} := n^{-1}\sum_{i=1}^{n} \hat{\dot{\nu}}(Z_i)$, $\quad \tilde{\dot{\nu}}^c(Z_i) := \hat{\dot{\nu}}(Z_i) - \tilde{\dot{\nu}}$,

$$\tilde{T}_{n,R} := \int_0^1 \widehat{\Gamma}_\theta(u)\tilde{\mathcal{U}}_R(u)d\Psi(u),$$

$$\tilde{\mathcal{K}}_R(t) := \int_0^1 \|\tilde{\mathcal{U}}_R(u)\|^2 d\Psi(u) + 2\tilde{T}'_{n,R}t + t'\widehat{\Omega}_\theta t,$$

where $\widehat{\Gamma}_\theta$ and $\widehat{\Omega}_\theta$ are defined in Sect. 4.2. Similar to Lemma 4. we have the following lemma.

Lemma 5 *Suppose model* (12) *with H unknown holds and an independent validation data* $\{(\tilde{X}_k, \tilde{Z}_k), 1 \le k \le N\}$ *obeying* (12) *is available. In addition assume that* (54)*,* (55)*,* (61)*–*(65) *hold. Then, for* $0 \le \lambda < \infty$*,*

$$\text{(a)} \quad \tilde{T}_{n,R} \to N(0, \widehat{\Sigma}_\theta + \lambda\Sigma_2), \tag{66}$$
$$\text{(b)} \quad n^{1/2}(\tilde{\theta}_R - \theta) \to N(0, \widehat{\Omega}_\theta^{-1}(\widehat{\Sigma}_\theta + \lambda\Sigma_2)\widehat{\Omega}_\theta^{-1}).$$

Moreover, $N^{1/2}(\tilde{\theta}_R - \theta) \to N(0, \widehat{\Omega}_\theta^{-1}\Sigma_2\widehat{\Omega}_\theta^{-1})$*, for* $\lambda = \infty$*.*

See Appendix for some details of the proof.

6 Data Analysis

Example. We shall now compute the above estimators based on some real data. The data pertains to the study of the relationship between the enzyme reaction speed (Y) and the basal density (X) of the UDP-galactose, see Bates and Watts [1], p. 70. A suitable model commonly used to analyze this data is the Michaelis-Menten model

$$m_\theta(x) = \frac{\alpha x}{\beta + x}, \quad \theta := (\alpha, \beta)', \ \alpha > 0, \ \beta > 0, \ x > 0.$$

In the primary data, consisting of $n = 30$ observations, the basal density variable was measured using a simple chemical method. It was believed that this method caused measurement error in the observation. Hence, in the validation data, consisting of $N = 10$ observations, an expensive procedure with a precision machine tool was used to produce precise observations of the basal density. Let Z denote the basal-density obtained by the chemical method parts per millions (ppm), $\widetilde{Z}$ denote the basal-density obtained by the exact measure (ppm) and Y, the reaction speed (counts/min^2). The primary and validation data are as follows. Table 2 gives the m.d. estimators $\hat{\theta}_1$ with $G(x) = x$ and $\tilde{\theta}_R$ with $\Psi(u) = u$, based on the above primary and validation data, and the naive least squares estimators $\hat{\theta}_{\text{nLS}}$ obtained by ignoring measurement errors. The MSEs are calculated by using the following formulas, where $\widetilde{\eta}_k = \widetilde{X}_k - \widetilde{Z}_k$, $MSE(\hat{\theta}_1) = \frac{1}{n}\sum_{i=1}^n \Big[Y_i -$

Table 2 M.D. and naive estimators and their MSE

Estimators	$\hat{\theta}_1$	$\tilde{\theta}_R$	$\hat{\theta}_{\mathrm{nLS}}$
α	217.30	217.53	212.7
γ	0.069	0.063	0.064
MSE	48.96	57.85	49.87

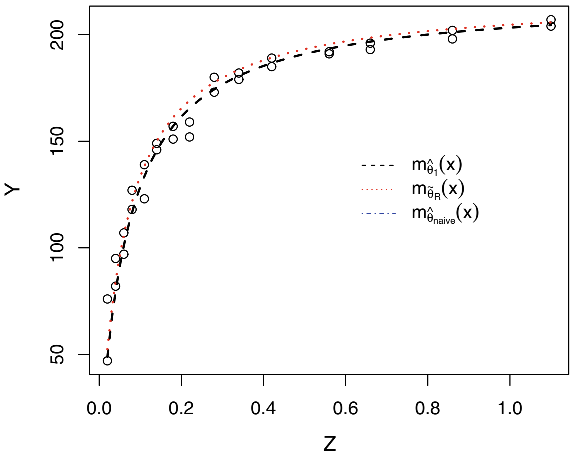

Fig. 1 Fitted regressions based on the three estimators

$\frac{1}{N}\sum_{k=1}^{N} m_{\hat{\theta}_1}(Z_i+\tilde{\eta}_k)\Big]^2$, $MSE(\tilde{\theta}_R)=\frac{1}{n}\sum_{i=1}^{n}\Big[Y_i-\frac{1}{N}\sum_{k=1}^{N} m_{\tilde{\theta}_R}(Z_i+\tilde{\eta}_k)\Big]^2$ and $MSE(\hat{\theta}_{\mathrm{nLS}})=\frac{1}{n}\sum_{i=1}^{n}\Big[Y_i-m_{\hat{\theta}_{\mathrm{nLS}}}(Z_i)\Big]^2$. Figure 1 presents the fitted regression curves using the three estimators.

Z	0.02	0.02	0.04	0.04	0.06	0.06	0.08	0.08	0.11	0.11	0.14	0.14	0.18	0.18	0.22
Y	76	47	82	95	97	107	118	127	123	139	146	149	157	151	159
Z	0.42	0.42	0.56	0.56	0.66	0.66	0.86	0.86	1.10	1.10	0.22	0.28	0.28	0.34	0.34
Y	185	189	191	192	193	196	198	202	207	2.04	152	173	180	179	182

$\widetilde{Z}$	0.04	0.07	0.20	0.30	0.38	0.48	0.60	0.76	0.95	1.110
$\widetilde{X}$	0.035	0.076	0.207	0.295	0.388	0.486	0.601	0.754	0.952	1.112

Appendix

This section contains some details of the proofs of the various results.

Proof of (41). Let $\widetilde{M}(t)=\widetilde{D}(\theta+n^{-1/2}t)$, where $\widetilde{D}(\vartheta)$ is as in (16). Define

$$\nu_{nt}(z) := \nu_{\theta+n^{-1/2}t}(z), \qquad \xi_{it} := \nu_{nt}(Z_i) - \nu_\theta(Z_i),$$

$$V_s(x,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\nu}_{ns}(Z_i) I(Y_i - \nu_{nt}(Z_i) \le x)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\nu}_{ns}(Z_i) I\big(\zeta_i \le x + \xi_{it}\big),$$

$$V(x,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\nu}(Z_i) I\big(\zeta_i \le x + \xi_{it}\big),$$

$$J(x,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\nu}(Z_i) L_{Z_i}(x + \xi_{it}),$$

$$J_s(x,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\nu}_{ns}(Z_i) L_{Z_i}(x + \xi_{it}), \quad W_s(x,t) := V_s(x,t) - J_s(x,t),$$

$$W(x,t) := V(x,t) - J(x,t), \quad s, t \in \mathbb{R}^q, x \in \mathbb{R}.$$

Note that $EV_s(x,t) \equiv EJ_s(x,t),\ EW_s(x,t) \equiv 0$. By (15), $\forall\, s \in \mathbb{R}^q, x \in \mathbb{R}$,

$$n^{-1/2} \sum_{i=1}^n \dot{\nu}_{ns}(Z_i) \big\{L_{Z_i}(x) + L_{Z_i}(-x)\big\} = n^{-1/2} \sum_{i=1}^n \dot{\nu}_{ns}(Z_i).$$

Define

$$\gamma_{nt}(x) := n^{-1/2} \sum_{i=1}^n \dot{\nu}_{nt}(Z_i) \xi_{it} \ell_{Z_i}(x), \quad g_n(x) := n^{-1} \sum_{i=1}^n \dot{\nu}(Z_i) \dot{\nu}(Z_i)' \ell_{Z_i}(x).$$

Because of (15), $\gamma_{nt}(x) \equiv \gamma_{nt}(-x),\ g_n(x) \equiv g_n(-x)$ and we rewrite

$$\widetilde{M}(t) = \int_0^\infty \Big\| V_t(x,t) + V_t(-x,t) - n^{-1/2} \sum_{i=1}^n \dot{\nu}_{nt}(Z_i) \Big\|^2 dG(x)$$

$$\begin{aligned} = \int_0^\infty \Big\| &\big\{W_t(x,t) - W_t(x,0)\big\} + \big\{W_t(x,0) - W(x,0)\big\} \\ &+ \big\{W_t(-x,t) - W_t(-x,0)\big\} + \big\{W_t(-x,0) - W(-x,0)\big\} \\ &+ \big\{J_t(x,t) - J_t(x,0) - \gamma_{nt}(x)\big\} \\ &+ \big\{J_t(-x,t) - J_t(-x,0) - \gamma_{nt}(-x)\big\} + 2\big\{\gamma_{nt}(x) - g_n(x)t\big\} \\ &+ \big\{W(x,0) + W(-x,0) + 2g_n(x)t\big\} \Big\|^2 dG(x). \end{aligned}$$

Expand the quadratic of the six summands in the integrand to obtain

$$\widetilde{M}(t) = M_1(t) + M_2(t) + \cdots + M_8(t) + 28 \text{ cross product terms,}$$

where

$$M_1(t) := \int_0^\infty \left\| W_t(x,t) - W_t(x,0) \right\|^2 dG(x),$$

$$M_2(t) := \int_0^\infty \left\| W_t(x,0) - W(x,0) \right\|^2 dG(x),$$

$$M_3(t) := \int_0^\infty \left\| W_t(-x,t) - W_t(-x,0) \right\|^2 dG(x),$$

$$M_4(t) := \int_0^\infty \left\| W_t(-x,0) - W(-x,0) \right\|^2 dG(x),$$

$$M_5(t) := \int_0^\infty \left\| J_t(x,t) - J_t(x,0) - \gamma_{nt}(x) \right\|^2 dG(x),$$

$$M_6(t) := \int_0^\infty \left\| J_t(-x,t) - J_t(-x,0) - \gamma_{nt}(-x) \right\|^2 dG(x),$$

$$M_7(t) := 4 \int_0^\infty \left\| \gamma_{nt}(x) - g_n(x)t \right\|^2 dG(x),$$

$$M_8(t) := \int_0^\infty \left\| W(x,0) + W(-x,0) + 2g_n(x)t \right\|^2 dG(x).$$

Recall $\mathcal{U}(b) := \{t \in \mathbb{R}^q; \|t\| \le b\}$, $b > 0$. We shall prove the following lemma shortly.

Lemma 6 *Under the assumptions* (13) *to (18),* $\forall\, 0 < b < \infty$,

$$\sup_{t \in \mathcal{U}(b)} M_j(t) \to_p 0, \quad j = 1, 2, \ldots, 7, \tag{67}$$

$$\sup_{t \in \mathcal{U}(b)} M_8(t) \;=\; O_p(1). \tag{68}$$

Unless mentioned otherwise, all the supremum below are taken over $t \in \mathcal{U}(b)$. Lemma 6 together with the C-S inequality implies that the supremum over t of all the cross product terms tends to zero, in probability. For example, by the C-S inequality,

$$\sup_t \Big| \int_0^\infty \big\{W_t(x,t) - W_t(x,0)\big\}\big\{J_t(x,t) - J_t(x,0) - \gamma_{nt}(x)\big\} dG(x)\Big|^2 \\ \leq \sup_t M_1(t) \sup_t M_5(t) = o_p(1),$$

by (67) used with $j = 1, 5$. Similarly, by (67) with $j = 1$ and (68),

$$\sup_t \Big| \int_0^\infty \big\{W_t(x,t) - W_t(x,0)\big\}\big\{W(x,0) + W(-x,0) + 2g_n(x)t\big\} dG(x)\Big|^2 \\ \leq \sup_t M_1(t) \sup_t M_8(t) = o_p(1) \times O_p(1) = o_p(1).$$

Consequently, we obtain

$$\sup_t \big|\widetilde{M}(t) - M_8(t)\big| = o_p(1). \tag{69}$$

Expand the quadratic in M_8 to write

$$\begin{aligned} M_8(t) := & \int_0^\infty \big\|W(x,0) + W(-x,0)\big\|^2 dG(x) + \\ & 4t' \int_0^\infty g_n(x)\big\{W(x,0) + W(-x,0)\big\} dG(x) + 4\int_0^\infty \big(t' g_n(x)\big)^2 dG(x) \\ = & \widetilde{M}(0) + 4t'\widetilde{T}_n + 4\int_0^\infty \big(t' g_n(x)\big)^2 dG(x), \end{aligned} \tag{70}$$

where $\widetilde{T}_n := \int_0^\infty g_n(x)\big\{W(x,0) + W(-x,0)\big\} dG(x)$. Let

$$T_n^* := \int_0^\infty \Gamma_\theta(x)\big\{W(x,0) + W(-x,0)\big\} dG(x).$$

By the LLNs and an Extended Dominated Convergence Theorem

$$\sup_t \left\| t'(g_n(x) - \Gamma_\theta(x)) \right\| \to_p 0, \quad \forall\, x \in \mathbb{R};$$

$$\sup_t \int_0^\infty \left\| t'(g_n(x) - \Gamma_\theta(x)) \right\|^2 dG(x) \to_p 0.$$

Moreover, recall $\widetilde{M}(0) = \widetilde{D}(\theta)$, so that by (39), $\widetilde{M}(0) = O_p(1)$. These facts together with the C-S inequality imply that

$$\begin{aligned}\left\| \widetilde{T}_n - T_n^* \right\|^2 &= \left\| \int_0^\infty \left\{ g_n(x) - \Gamma_\theta(x) \right\} \left\{ W(x,0) + W(-x,0) \right\} dG(x) \right\|^2 \\ &\le \widetilde{M}(0) \int_0^\infty \left\| g_n(x) - \Gamma_\theta(x) \right\|^2 dG(x) \to_p 0.\end{aligned}$$

These facts combined with (22), (69), (70) yield that

$$\sup_t \left| \widetilde{M}(t) - \widetilde{M}(0) - 4T_n^* t - 4t' \int_0^\infty \Gamma_\theta(x)\Gamma_\theta(x) dG(x)\, t \right| = o_p(1).$$

Now recall that $D(\vartheta) = 2\widetilde{D}(\vartheta)$, $\widetilde{M}(t) = \widetilde{D}(\theta + n^{-1/2}t)$, $\Omega_\theta = 2\int_0^\infty \Gamma_\theta \Gamma_\theta dG$ and $T_n = 2T_n^*$, see (40). Hence the above expansion is equivalent to

$$\sup_t \left| \widetilde{D}(\theta + n^{-1/2}t) - \widetilde{D}(\theta) - 2T_n t - 2t'\Omega_\theta t \right| = o_p(1),$$
$$\sup_t \left| D(\theta + n^{-1/2}t) - D(\theta) - 4T_n t - 4t'\Omega_\theta t \right| = o_p(1),$$

which is precisely the claim (41). □

Proof of Lemma 6. Let $\delta_{it} := \xi_{it} - n^{-1/2} t' \dot{\nu}(Z_i)$. By (13) and (14),

$$\max_{1 \le i \le n,\, t} n^{1/2} \left| \delta_{it} \right| = o_p(1), \qquad \max_{1 \le i \le n} n^{-1/2} \|\dot{\nu}(Z_i)\| = o_p(1). \tag{71}$$

Hence,

$$\begin{aligned}\max_{1 \le i \le n,\, t} \left| \xi_{it} \right| &\le \max_{1 \le i \le n,\, \|t\| \le b} \left| \delta_{it} \right| + \max_{1 \le i \le n,\, t} n^{-1/2} \left\| t \right\| \|\dot{\nu}(Z_i)\| \qquad (72) \\ &\le o_p(n^{-1/2}) + b \max_{1 \le i \le n} n^{-1/2} \|\dot{\nu}(Z_i)\| = o_p(1),\end{aligned}$$

$$\sum_{i=1}^n \xi_{it}^2 = \sum_{i=1}^n (\nu_{nt}(Z_i) - \nu(Z_i))^2 = \sum_{i=1}^n \delta_{it}^2 + n^{-1} \sum_{i=1}^n (t'\dot{\nu}(Z_i))^2,$$

$$\sup_t \sum_{i=1}^n \xi_{it}^2 \le n \max_{1\le i\le n,\, \|t\|\le b} \left|\delta_{it}\right|^2 + b^2 n^{-1} \sum_{i=1}^n \|\dot{\nu}(Z_i)\|^2 = O_p(1), \tag{73}$$

by (14). Moreover, by (14) and the Law of Large Numbers,

$$\begin{aligned} &\sup_t \left\| n^{-1/2} \sum_{i=1}^n \dot{\nu}_\theta(Z_i)\xi_{it} \right\| \qquad (74)\\ &\le \max_{1\le i\le n,\, \|t\|\le b} n^{1/2}|\delta_{it}| n^{-1} \sum_{i=1}^n \|\dot{\nu}_\theta(Z_i)\| + bn^{-1} \left\| \sum_{i=1}^n \dot{\nu}_\theta(Z_i)\dot{\nu}_\theta(Z_i)' \right\| \\ &= o_p(1) + O_p(1) = O_p(1). \end{aligned}$$

These facts will be use in the sequel.

Consider the term M_7. Write

$$\begin{aligned} &\gamma_{nt}(x) - g_n(x)t \\ &= n^{-1/2} \sum_{i=1}^n \dot{\nu}_{nt}(Z_i)\xi_{it}\ell_{Z_i}(x) - n^{-1} \sum_{i=1}^n \dot{\nu}(Z_i)\dot{\nu}(Z_i)'\ell_{Z_i}(x)t \\ &= n^{-1/2} \sum_{i=1}^n \left[\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\right]\xi_{it}\ell_{Z_i}(x) + n^{-1/2} \sum_{i=1}^n \dot{\nu}(Z_i)\delta_{it}\ell_{Z_i}(x) \\ &= n^{-1/2} \sum_{i=1}^n \left[\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\right]\delta_{it}\ell_{Z_i}(x) \\ &\quad + n^{-1} \sum_{i=1}^n \left[\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\right]\dot{\nu}(Z_i)'\ell_{Z_i}(x)t + n^{-1/2} \sum_{i=1}^n \dot{\nu}(Z_i)\delta_{it}\ell_{Z_i}(x). \end{aligned}$$

Hence

$$M_7 = \int_0^\infty \left\|\gamma_{nt}(x) - g_n(x)t\right\|^2 dG(x) \le 4\{M_{71}(t) + M_{72}(t) + M_{73}(t)\},$$

where

$$\begin{aligned} M_{71}(t) &= n^{-1} \int_0^\infty \left\| \sum_{i=1}^n \left[\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\right]\delta_{it}\ell_{Z_i}(x) \right\|^2 dG(x), \\ M_{72}(t) &= n^{-2} \int_0^\infty \left\| \sum_{i=1}^n \left[\dot{\nu}_{nt}(Z_i) - \dot{\nu}(Z_i)\right]\dot{\nu}(Z_i)'\ell_{Z_i}(x)t \right\|^2 dG(x), \end{aligned}$$

$$M_{73}(t) = n^{-1} \int_0^\infty \Big\| \sum_{i=1}^n \dot\nu(Z_i)\delta_{it}\ell_{Z_i}(x) \Big\|^2 dG(x).$$

But, by (18) and (71),

$$\begin{aligned} &\sup_t M_{71}(t) \\ &\le n \sup_{t,1\le i\le n} \delta_{it}^2 \sup_{t,1\le i\le n} \big\|\dot\nu_{nt}(Z_i) - \dot\nu(Z_i)\big\|^2 \int_0^\infty n^{-1} \sum_{i=1}^n \ell_{Z_i}^2(x) dG(x) = o_p(1). \end{aligned}$$

Similarly, by the C-S inequality,

$$\begin{aligned} \sup_t M_{72}(t) &\le b^2 \sup_{t,1\le i\le n} \big\|\dot\nu_{nt}(Z_i) - \dot\nu(Z_i)\big\|^2 n^{-1} \int_0^\infty \sum_{i=1}^n \|\dot\nu(Z_i)\|^2 \ell_{Z_i}^2(x) dG(x) \\ &= o_p(1) O_p(1) = o_p(1), \end{aligned}$$

by (18) and (19). Again, by (19) and (71),

$$\sup_t M_{73}(t) \le \sup_{t,1\le i\le n} n|\delta_{it}|^2 n^{-1} \int_0^\infty \sum_{i=1}^n \|\dot\nu(Z_i)\|^2 \ell_{Z_i}^2(x) dG(x) = o_p(1).$$

These facts prove (67) for $j = 7$.

Next consider M_5. Let $D_{it}(x) := L_{Z_i}(x + \xi_{it}) - L_{Z_i}(x) - \xi_{it}\ell_{Z_i}(x)$. Then

$$\begin{aligned} M_5(t) &:= \frac{1}{n} \int_0^\infty \Big\| \sum_{i=1}^n \dot\nu_{nt}(Z_i) D_{it}(x) \Big\|^2 dG(x) \\ &\le \frac{1}{n} \sum_{i=1}^n \Big\|\dot\nu_{nt}(Z_i)\Big\|^2 \int_0^\infty \sum_{i=1}^n D_{it}^2(x) dG(x). \end{aligned} \tag{75}$$

By (14) and (18),

$$\begin{aligned} \sup_t n^{-1} \sum_{i=1}^n \Big\|\dot\nu_{nt}(Z_i)\Big\|^2 &\le \sup_t \; n^{-1} \sum_{i=1}^n \Big\|\dot\nu_{nt}(Z_i) - \dot\nu(Z_i)\Big\|^2 \\ &\quad + \sup_t n^{-1} \sum_{i=1}^n \Big\|\dot\nu(Z_i)\Big\|^2 = o_p(1). \end{aligned}$$

By the C-S inequality, Fubini Theorem, (20) and (73),

$$\begin{aligned}
&\int_0^\infty \sum_{i=1}^n D_{it}^2(x)dG(x)\\
&\le \int_0^\infty \sum_{i=1}^n \Big(\int_{-|\xi_{it}|}^{|\xi_{it}|} \big(\ell_{Z_i}(x+u)-\ell_{Z_i}(x)\big)du\Big)^2 dG(x)\\
&\le \int_0^\infty \sum_{i=1}^n |\xi_{it}| \int_{-|\xi_{it}|}^{|\xi_{it}|} \big(\ell_{Z_i}(x+u)-\ell_{Z_i}(x)\big)^2 du dG(x)\\
&\le \max_{1\le i\le n,t} |\xi_{it}|^{-1} \int_{-|\xi_{it}|}^{|\xi_{it}|} \int_0^\infty \big(\ell_{Z_i}(x+u)-\ell_{Z_i}(x)\big)^2 dG(x)du \sum_{i=1}^n |\xi_{it}|^2\\
&= o_p(1).
\end{aligned}$$

Upon combining these facts with (75) we obtain $\sup_t M_5(t) = o_p(1)$, thereby proving (67) for $j = 5$. The proof for $j = 6$ is exactly similar.

Now consider M_1. Let $\xi_t(Z) := \nu_{nt}(Z) - \nu(Z)$. Then

$$\begin{aligned}
EM_1(t) :\le& \int_{-\infty}^{\infty} E\big\|W_t(x,t)-W_t(x,0)\big\|^2 dG(x)\\
\le&\ n^{-1}\sum_{i=1}^n E\Big(\|\dot{\nu}_{nt}(Z_i)\|^2 \int_{-\infty}^{\infty} \Big|L_{Z_i}(x+\xi_{it})-L_{Z_i}(x)\Big| dG(x)\Big)\\
\le&\ n^{-1}\sum_{i=1}^n E\Big(\|\dot{\nu}_{nt}(Z_i)\|^2 \int_{-\infty}^{\infty}\int_{-|\xi_{it}|}^{|\xi_{it}|} \ell_{Z_i}(x+u)dudG(x)\\
=&\ E\Big(\int_{-|\xi_t(Z)|}^{|\xi_t(Z)|} \|\dot{\nu}_{nt}(Z)\|^2 \int_{-\infty}^{\infty} \ell_Z(x+u)dG(x)\,du\Big) \to 0,
\end{aligned}$$

by (21). Thus

$$M_1(t) = o_p(1), \qquad \forall\, t \in \mathcal{U}(b). \tag{76}$$

To prove that this holds uniformly in $t \in \mathcal{U}(b)$, because of the compactness of the ball $\mathcal{U}(b)$, it suffices to show that for every $\epsilon > 0$ there is a $\delta > 0$ and an N_ϵ such that for every $s \in \mathcal{U}(b)$,

$$P\Big(\sup_{\|t-s\|<\delta} \|M_1(t) - M_1(s)| \ge \epsilon\Big) \le \epsilon, \quad \forall n > N_\epsilon. \tag{77}$$

Let $\dot{\nu}_{ntj}(z)$ denote the jth coordinate of $\dot{\nu}_{nt}(z)$, $j = 1, \dots, q$ and let

$$\alpha_i(x,t) := I(\zeta_i \le x + \xi_{it}) - I(\zeta_i \le x) - L_{Z_i}(x + \xi_{it}) + L_{Z_i}(x).$$

Then

$$M_1(t) = \int_0^\infty \big\|W_t(x,t) - W_t(x,0)\big\|^2 dG(x)$$
$$= \sum_{j=1}^q \int_0^\infty \Big(n^{-1/2} \sum_{i=1}^n \dot{\nu}_{ntj}(Z_i)\alpha_i(x,t)\Big)^2 dG(x) = \sum_{j=1}^q M_{1j}(t), \quad \text{say.}$$

Thus it suffices to prove (77) with M_1 replaced by M_{1j} for each $j = 1, \dots, q$.

Any real number a can be written as $a = a^+ - a^-$, where $a^+ = \max(0, a)$, $a^- = \max(0, -a)$. Note that $a^\pm \ge 0$. Fix a $j = 1, \dots, q$, write $\dot{\nu}_{ntj}(Z_i) = \dot{\nu}^+_{ntj}(Z_i) - \dot{\nu}^-_{ntj}(Z_i)$ and define

$$W_j^\pm(x,t) := n^{-1/2} \sum_{i=1}^n \dot{\nu}^\pm_{ntj}(Z_i)\alpha_i(x,t),$$

$$D_j^\pm(x,s,t) := W_j^\pm(x,t) - W_j^\pm(x,s), \quad R_j^\pm(s,t) := \int_0^\infty \big(D_j^\pm(x,s,t)\big)^2 dG(x).$$

Then

$$\begin{aligned}
&\big|M_{1j}(t) - M_{1j}(s)\big| \qquad (78)\\
&= \Big| \int_0^\infty \big(W_j^+(x,t) - W_j^-(x,t)\big)^2 dG(x) \\
&\qquad - \int_0^\infty \big(W_j^+(x,s) - W_j^-(x,s)\big)^2 dG(x) \Big| \\
&\le \int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x) + \int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x) \\
&\quad + 2\Big\{ \int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x) \int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x) \Big\}^{1/2}
\end{aligned}$$

$$+2\Big[\Big\{\int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x)\Big\}^{1/2} + \Big\{\int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x)\Big\}^{1/2}\Big] M_{1j}^{1/2}(s)$$

$$= R_j^+(s,t) + R_j^-(s,t) + 2\big(R_j^+(s,t)R_j^-(s,t)\big)^{1/2} + \big\{(R_j^+(s,t))^{1/2} + \big(R_j^-(s,t)\big)^{1/2}\big\} M_{1j}^{1/2}(s).$$

Write

$$\begin{aligned} D_j^+(x,s,t) &= n^{-1/2}\sum_{i=1}^n \dot\nu_{ntj}^+(Z_i)\alpha_i(x,t) - n^{-1/2}\sum_{i=1}^n \dot\nu_{nsj}^+(Z_i)\alpha_i(x,s) \\ &= n^{-1/2}\sum_{i=1}^n \big[\dot\nu_{ntj}^+(Z_i) - \dot\nu_{nsj}^+(Z_i)\big]\alpha_i(x,t) \\ &\quad + n^{-1/2}\sum_{i=1}^n \dot\nu_{nsj}^+(Z_i)\big[\alpha_i(x,t) - \alpha_i(x,s)\big] \\ &= D_{j1}^+(x,s,t) + D_{j2}^+(x,s,t), \quad \text{say.} \end{aligned}$$

Hence

$$R_j^+(s,t) \le 2\int_0^\infty \big(D_{j1}^+(x,s,t)\big)^2 dG(x) + 2\int_0^\infty \big(D_{j2}^+(x,s,t)\big)^2 dG(x). \tag{79}$$

By (23), the first term here satisfies (77). We proceed to verify it for the second term. Fix an $s \in \mathcal{U}_b$, $\epsilon > 0$ and $\delta > 0$. Let

$$\Delta_{ni} := n^{-1/2}\big(\delta\|\dot\nu(Z_i)\| + 2\epsilon\big), \quad B_n := \Big\{\sup_{t\in\mathcal{N}_b, \|t-s\|\le\delta} \big|\xi_{it} - \xi_{is}\big| \le \Delta_{ni}\Big\}.$$

By (18), there exists an N_ϵ such that $P(B_n) > 1-\epsilon$, for all $n > N_\epsilon$. On B_n, $\xi_{is} - \Delta_{ni} \le \xi_{it} \le \xi_{is} + \Delta_{ni}$ and, by the nondecreasing property of the indicator function and d.f., we obtain

$$\begin{aligned} &I(\zeta_i \le x + \xi_{is} - \Delta_{ni}) - I(\zeta_i \le x) - L_{Z_i}(x - \xi_{is} + \Delta_{ni}) + L_{Z_i}(x) \\ &\qquad - L_{Z_i}(x + \xi_{is} + \Delta_{ni}) + L_{Z_i}(x + \xi_{is} - \Delta_{ni}) \\ &\le \alpha_i(x,t) = I(\zeta_i \le x + \xi_{it}) - I(\zeta_i \le x) - L_{Z_i}(x + \xi_{it}) + L_{Z_i}(x) \\ &\le I(\zeta_i \le x + \xi_{is} + \Delta_{ni}) - I(\zeta_i \le x) - L_{Z_i}(x + \xi_{is} + \Delta_{ni}) + L_{Z_i}(x) \\ &\qquad + L_{Z_i}(x + \xi_{is} + \Delta_{ni}) - L_{Z_i}(x + \xi_{is} - \Delta_{ni}). \end{aligned}$$

Let

$$\mathcal{D}^{\pm}_{j2}(x,s,a) := n^{-1/2}\sum_{i=1}^{n}\dot{\nu}^{\pm}_{njs}\Big\{I(\zeta_i \le x+\xi_{is}+a\Delta_{ni}) - I(\zeta_i \le x) \\ -L_{Z_i}(x+\xi_{is}+a\Delta_{ni}) + L_{Z_i}(x)\Big\}.$$

The above inequalities and $\dot{\nu}^{+}_{njs}(Z_i)$ being nonnegative yield that on B_n,

$$\begin{aligned}
&\int_0^\infty \big(D^{+}_{j2}(x,s,t)\big)^2 dG(x) \\
&\le \int_0^\infty \big(\mathcal{D}^{+}_{j2}(x,s,1) - \mathcal{D}^{+}_{j2}(x,s,0)\big)^2 dG(x) \\
&\quad + \int_0^\infty \big(\mathcal{D}^{+}_{j2}(x,s,-1) - \mathcal{D}^{+}_{j2}(x,s,0)\big)^2 dG(x) \\
&\quad + \int_0^\infty \Big(n^{-1/2}\sum_{i=1}^{n}\dot{\nu}^{+}_{njs}(Z_i)\big\{L_{Z_i}(x+\xi_{is}+\Delta_{ni}) \\
&\qquad\qquad -L_{Z_i}(x+\xi_{is}-\Delta_{ni})\big\}dG(x)\Big)^2.
\end{aligned}$$

Note that $\max_{1\le i\le n}(|\xi_{is}|+\Delta_{ni}) = o_p(1)$. Argue as for (76) to see that the first two terms in the above bound are $o_p(1)$, while the last term is bounded from the above by

$$\begin{aligned}
&\int_0^\infty \Big(n^{-1/2}\sum_{i=1}^{n}\dot{\nu}^{+}_{njs}(Z_i)\int_{\xi_{is}-\Delta_{ni}}^{\xi_{is}+\Delta_{ni}} \ell_{Z_i}(x+u)du\, dG(x)\Big)^2 \qquad (80)\\
&\le 2n^{-1}\sum_{i=1}^{n}(\dot{\nu}^{+}_{njs}(Z_i))^2\sum_{i=1}^{n}\Delta_{ni}\int_{\xi_{is}-\Delta_{ni}}^{\xi_{is}+\Delta_{ni}}\int_0^\infty \big[\ell^2_{Z_i}(x+u) - \ell^2_{Z_i}(x)\big]\, dG(x)\, du \\
&\quad + 4n^{-1}\sum_{i=1}^{n}(\dot{\nu}^{+}_{njs}(Z_i))^2\sum_{i=1}^{n}\Delta^2_{ni}\int_0^\infty \ell^2_{Z_i}(x)dG(x).
\end{aligned}$$

The first summand in the above bound is bounded above by

$$2 \max_{1\le i\le n} (2\Delta_{ni})^{-1} \int_{\xi_{is}-\Delta_{ni}}^{\xi_{is}+\Delta_{ni}} \int_0^\infty \big[\ell^2_{Z_i}(x+u) - \ell^2_{Z_i}(x)\big] du dG(x)$$
$$\times n^{-1} \sum_{i=1}^n (\dot\nu^+_{njs}(Z_i))^2 \sum_{i=1}^n \Delta^2_{ni} = o_p(1),$$

because the first factor tends to zero in probability by (20) and the second factor satisfies

$$n^{-1}\sum_{i=1}^n (\dot\nu^+_{njs}(Z_i))^2 \sum_{i=1}^n \Delta^2_{ni} \le n^{-1}\sum_{i=1}^n \|\dot\nu_{ns}\|^2\big(2n^{-1}\delta^2 \sum_{i=1}^n \|\dot\nu(Z_i)\|^2 + 4\epsilon^2\big).$$

The second term in the upper bound of (80) is bounded from the above by

$$4n^{-1}\sum_{i=1}^n \|\dot\nu_{ns}(Z_i)\|^2\, n^{-1}\sum_{i=1}^n \big(\delta^2\|\dot\nu(Z_i)\|^2 + 4\epsilon^2\big) \int_0^\infty \ell^2_{Z_i}(x) dG(x)$$
$$\to_p E\|\dot\nu(Z)\|^2\Big[\delta^2 \int_0^\infty E(\|\dot\nu(Z)\|^2\ell^2_Z(x))dG(x) + 4\epsilon^2 \int_0^\infty E(\ell^2_Z(x))dG(x)\Big].$$

Since the factor multiplying δ^2 is positive, the above term can be made smaller than ϵ by the choice of δ. Hence (77) is satisfied by the second term in the upper bound of (79). This then completes the proof of R^+_j satisfying (77). The details of the proof for verifying (77) for R^-_j are exactly similar. These facts together with the upper bound of (78) show that (77) is satisfied by M_{1j} for each $j = 1, \ldots, q$. This also completes the proof of $\sup_t M_1(t) = o_p(1)$, thereby proving (67) for $j = 1$. The proof for $j = 3$ is similar.

Next, consider M_2. Recall $\beta_i(x) := I(\zeta_i \le x) - L_{Z_i}(x)$. Then

$$M_2(t) := n^{-1} \int_0^\infty \Big\| \sum_{i=1}^n \{\dot\nu_{nt}(Z_i) - \dot\nu(Z_i)\}\beta_i(x)\Big\|^2 dG(x).$$

Because $E(\beta_i(x)|Z_i) \equiv 0$, a.s., we have

$$EM_2(t) = \int_0^\infty E\Big(\big\|\dot\nu_{nt}(Z) - \dot\nu(Z)\|^2 L_Z(x)(1 - L_Z(x)\Big)dG(x) \to 0,$$

by (18). Thus

$$M_2(t) = o_p(1), \qquad \forall\, t \in \mathbb{R}^q. \tag{81}$$

To prove this holds uniformly in $t \in \mathcal{U}(b)$, we shall verify (77) for M_2. Accordingly, let $\delta > 0$, $s \in \mathcal{U}(b)$ be fixed. Then forall $t \in \mathcal{U}(b)$ such that $\|t - s\| < \delta$,

$$\begin{aligned}
&\left|M_2(t) - M_2(s)\right| \\
&\le n^{-1} \int_0^\infty \Big\| \sum_{i=1}^n \{\dot{\nu}_{nt}(Z_i) - \dot{\nu}_{ns}(Z_i)\}\beta_i(x)\Big\|^2 dG(x) \\
&\quad + 2\Big(n^{-1} \int_0^\infty \Big\| \sum_{i=1}^n \{\dot{\nu}_{nt}(Z_i) - \dot{\nu}_{ns}(Z_i)\}\beta_i(x)\Big\|^2 dG(x)\Big)^{1/2} M_2(s)^{1/2}
\end{aligned}$$

This bound, (24) and (81) now readily verifies (77) for M_2, which also completes the proof of (67) for $j = 2$. The proof of (67) for $j = 4$ is precisely similar. This in turn completes the proof of Lemma 6. □

Proof of (60). Recall (43). Let $D(z, x) := m_\theta(z + x) - \nu_\theta(z)$. Use the fact $\widehat{U}(x, \theta) = \tilde{W}(x, 0) + \tilde{W}(-x, 0)$, to rewrite

$$\begin{aligned}
\tilde{T}_n &= \frac{1}{N^2\sqrt{n}} \sum_{i=1}^n \sum_{j,k=1}^N \int \mu(z)\dot{m}_\theta(Z_i + \tilde{\eta}_j)\{\varphi_z(\zeta_i) - 2\ell_z(\zeta_i)D(Z_i, \tilde{\eta}_k)\}dQ(z) \\
&\quad + \text{ higher order terms},
\end{aligned}$$

where $\varphi_z(x)$ is defined as in Sect. 4.1.

For further analysis of $\tilde{T}_n$, with the two independent samples $\{(Z_i, \zeta_i), 1 \le i \le n\}$ and $\{\tilde{\eta}_k, 1 \le k \le N\}$, define the symmetric kernel function ϕ and its projections as follows.

$$\begin{aligned}
&\phi(Z_1, \zeta_1, \tilde{\eta}_1, \tilde{\eta}_2) \\
&\quad := \int \mu(z)\dot{m}_\theta(Z_1 + \tilde{\eta}_1)\{\varphi_z(\zeta_1) - 2\ell_z(\zeta_1)D(Z_1, \tilde{\eta}_2)\}dQ(z) \\
&\qquad + \int \mu(z)\dot{m}_\theta(Z_1 + \tilde{\eta}_2)\{\varphi_z(\zeta_1) - 2\ell_z(\zeta_1)D(Z_1, \tilde{\eta}_1)\}dQ(z) \\
&E(\phi|Z_1, \zeta_1) = 2\int \mu(z)\dot{\nu}_\theta(Z_1)\varphi_z(\zeta_1)dQ(z), \quad E\phi(Z_1, \zeta_1, \tilde{\eta}_1, \tilde{\eta}_2) = 0, \\
&E(\phi|\tilde{\eta}_1) = -2\int \mu(z)E\{\dot{\nu}(Z)\ell_z(\zeta)D(Z, \tilde{\eta}_1)|\tilde{\eta}_1\}dQ(z).
\end{aligned}$$

Let $\tilde{T}_{n1}$ denote the first term in the right hand side of $\tilde{T}_n$. Then

$$\begin{aligned}\tilde{T}_{n1} &= \frac{1}{N^2\sqrt{n}}\sum_{i=1}^{n}\sum_{1\le j<k\le N}\phi(Z_i,\zeta_i,\tilde{\eta}_j,\tilde{\eta}_k)\\ &\quad+\frac{1}{N^2\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{N}\int\mu(z)\dot{m}_\theta(Z_i+\tilde{\eta}_j)\{\varphi_z(\zeta_i)-2\ell_z(\zeta_i)D(Z_i,\tilde{\eta}_j)\}dQ(z)\\ &=: \tilde{T}_{n11}+\tilde{T}_{n12}.\end{aligned}$$

Note that that $\tilde{T}_{n11}$ is a U-statistic with permutation degree 1 in the primary sample $\{(Z_i,\zeta_i), 1\le i\le n\}$ and permutation degree 2 in the validation sample $\{\tilde{\eta}_k, 1\le k\le N\}$. Theorem 6.1.4 in Lehmann [17] and (52) yield that, for $0\le\lambda<\infty$,

$$\begin{aligned}\tilde{T}_{n11} &= \sqrt{n}\times\frac{N(N-1)}{2N^2}\times\frac{1}{\binom{n}{1}\binom{N}{2}}\sum_{i=1}^{n}\sum_{1\le j<k\le N}\phi(Z_i,\zeta_i,\tilde{\eta}_j,\tilde{\eta}_k)\\ &\to_D \frac{1}{2}N\Big(0,\operatorname{Var}(E(\phi|Z_1,\zeta_1))+4\,\lambda\operatorname{Var}(E(\phi|\tilde{\eta}_1))\Big)=N(0,\Sigma_\theta+4\lambda\Sigma_1).\end{aligned}$$

Moreover, for $\lambda=\infty$, Theorem 6.1.4 in Lehmann [17] also yields that $\sqrt{N/n}\,\tilde{T}_{n11}\to_D N(0,4\Sigma_1)$.

Similarly, $\tilde{T}_{n12}$ is a U-statistic with permutation degree 1 for both samples. Since (52) implies that $E\{\|\dot{m}_\theta(X)[m_\theta(X)-\nu_\theta(Z)]\|\}<\infty$, therefore $E\tilde{T}_{n12}=O(n^{-1/2})$. Moreover, Theorem 6.1.3 of U-statistics in Lehmann [17] implies that $\operatorname{Var}(\tilde{T}_{n12})=O(n^{-1})$ and hence $\tilde{T}_{n12}=o_p(1)$. Hence the claim (60).

Proof of (66). Let $\dot{D}_{ijk}:=\dot{m}_\theta(Z_i+\tilde{\eta}_k)-\dot{m}_\theta(Z_j+\tilde{\eta}_k)$. Based on the definitions of $\widehat{\Gamma}_\theta(u)$ and $\kappa_z(v)$, $T_{n,R}$ can be rewritten as

$$\begin{aligned}\tilde{T}_{n,R} &= \int\int_0^1\mu^c(z)\tilde{\mathcal{U}}_R(u)\ell_z(L_z^{-1}(u))d\Psi(u)dQ(z)\\ &= -n^{-1/2}\sum_{i=1}^{n}\int\mu^c(z)\hat{\dot{\nu}}^c(Z_i)\kappa_z(L_{Z_i}(\zeta_i-\Delta(Z_i)))dQ(z)\\ &= -\frac{n^{-1/2}}{nN}\sum_{i=1}^{n}\sum_{j=1,j\ne i}^{n}\sum_{k=1}^{N}\int\mu^c(z)\dot{D}_{ijk}\kappa_z(L_{Z_i}(\zeta_i))dQ(z)\\ &\quad-\frac{n^{-1/2}}{nN}\sum_{i=1}^{n}\sum_{j=1,j\ne i}^{n}\sum_{k=1}^{N}\int\mu^c(z)\dot{D}_{ijk}\ell_z(\zeta_i)\Delta(Z_i)dQ(z)\\ &\quad+\text{ higher order} := T_{n,R1}+T_{n,R2}+\text{ higher order}.\end{aligned}$$

First, we study the asymptotic distribution of $T_{n,R1}$. Define for $1\le i,j\le n$, $i\ne j$, and $1\le k\le N$,

$$\psi_1(Z_i, \zeta_i, Z_j, \zeta_j, \tilde{\eta}_k) := \int \mu^c(z)\dot{D}_{ijk}\Big\{\kappa_z(L_{Z_i}(\zeta_i)) + \kappa_z(L_{Z_j}(\zeta_j))\Big\}dQ(z).$$

Then $T_{n,R1}$ can be rewritten as

$$T_{n,R1} = -\frac{n(n-1)}{2n^2} \times \frac{\sqrt{n}}{\binom{n}{2}\binom{N}{1}} \sum_{1\le i<j\le n} \sum_{k=1}^{N} \psi_1(Z_i, \zeta_i, Z_j, \zeta_j, \tilde{\eta}_k).$$

By the definition of U-statistics in Lehmann [17], $T_{n,R1}$ is a two sample U-statistic based on function ψ_1 with permutation degree of 2 on the sample $\{(Z_i, \zeta_i), 1 \le i \le n\}$ and permutation degree of 1 on the sample $\{\tilde{\eta}_k, 1 \le k \le N\}$. Because conditionally, $L_{Z_i}(\zeta_i)$, given Z_i, is a uniformly distributed r.v., we have $E(L_{Z_i}(\zeta_i)|Z_i) = \int_0^1 \kappa_z(u)du := K(z)$. Then the conditional expectations of ψ can be calculated as follows.

$$\begin{aligned}
E(\psi_1|Z_1, \zeta_1) &= \int \mu^c(z)[\dot{\nu}_\theta(Z_1) - E(\dot{\nu}_\theta(Z))]\kappa_z^c(L_{Z_1}(\zeta_1))dQ(z),\\
E(\psi_1|Z_1, Z_2, \tilde{\eta}_1) &= \int \mu^c(z)\big\{\dot{D}_{121} + \dot{D}_{211}\big\}K(z)dQ(z) = 0,\\
E(\psi_1|\tilde{\eta}_1) &= 0.
\end{aligned}$$

It can be seen that $\mathrm{Cov}(E(\psi_1|Z_1, \zeta_1)) = \widehat{\Sigma}_\theta$ as defined in Sect. 4.2. Then Theorem 6.1.4 in Lehmann [17] yields that

$$T_{n,R1} \to_D \frac{1}{2}N(0, 4\mathrm{Cov}(E(\psi|Z_1, \zeta_1))) = N(0, \widehat{\Sigma}_\theta).$$

Next, in order to study $T_{n,R2}$, define

$$\begin{aligned}
&\psi_2(Z_i, \zeta_i, Z_j, \zeta_j, \tilde{\eta}_k, \tilde{\eta}_l)\\
&= \int \mu^c(z)[\dot{m}_\theta(Z_i + \tilde{\eta}_k) - \dot{m}_\theta(Z_j + \eta_k)]\ell_z(\zeta_i)D(Z_i, \tilde{\eta}_l)dQ(z)\\
&\quad + \int \mu^c(z)[\dot{m}_\theta(Z_j + \tilde{\eta}_k) - \dot{m}_\theta(Z_i + \eta_k)]\ell_z(\zeta_j)D(Z_j, \tilde{\eta}_l)dQ(z)\\
&\quad + \int \mu^c(z)[\dot{m}_\theta(Z_i + \tilde{\eta}_l) - \dot{m}_\theta(Z_j + \eta_l)]\ell_z(\zeta_i)D(Z_i, \tilde{\eta}_k)dQ(z)\\
&\quad + \int \mu^c(z)[\dot{m}_\theta(Z_j + \tilde{\eta}_l) - \dot{m}_\theta(Z_i + \eta_l)]\ell_z(\zeta_j)D(Z_j, \tilde{\eta}_k)dQ(z).
\end{aligned}$$

Then $T_{n,R2}$ can be rewritten as a two sample U-statistic with permutation degree 2 for both primary sample and validation sample.

$$T_{n,R2} = -\frac{n(n-1)}{2n^2}\frac{N(N-1)}{2N^2}\frac{\sqrt{n}}{\binom{n}{2}\binom{N}{2}}\sum_{1\le i<j\le n}\sum_{1\le k<l\le N}\psi_2(Z_i,\zeta_i,Z_j,\zeta_j,\tilde\eta_k,\tilde\eta_l).$$

The conditional expectations of ψ_2 are calculated as

$$E(\psi_2|\tilde\eta_1) = 2\int \mu^c(z)E\big\{[\dot\nu_\theta(Z) - E(\dot\nu_\theta(Z))]\ell_z(\zeta)D(Z,\tilde\eta_1)dQ(z)$$
$$E(\psi_2|Z_1,\zeta_1) = 0.$$

Then Theorem 6.1.4 in Lehmann [17] shows that, for $0 \le \lambda < \infty$,

$$T_{n,R2} \to_D \frac{1}{4}N(0, 4\lambda \mathrm{Cov}(E(\psi_2|\tilde\eta_1))) = N(0,\lambda\Sigma_2).$$

The two terms $T_{n,R1}$ and $T_{n,R2}$ are asymptotically independent becuase of the independence between the primary sample and validation sample. In fact, $T_{n,R1}$ is based on $E(\psi_1|Z_1,\zeta_1)$ and $T_{n,R2}$ is based on $E(\psi_2|\tilde\eta_1)$. Therefore, (66)(a) holds. An argument similar to one used for (51) yields that $\sup_{\|t\|\le b}|\tilde{\mathcal{K}}(\theta + n^{-1/2}t) - \tilde{\mathcal{K}}_R(t)| = o_p(1)$, which in turn yields the claim (66)(b) about $\tilde\theta_R$.

When $\lambda = \infty$, by Theorem 6.1.4 in Lehmann [17], $\sqrt{N/n}\,T_{n,R2} \to_D N(0,\Sigma_2)$. Then $\sqrt{N/n}\,\tilde T_{n,R} = \sqrt{N/n}\,\tilde T_{n,R1} + \sqrt{N/n}\,\tilde T_{n,R2} \to_D N(0,\Sigma_2)$. Therefore, we obtain that $\sqrt{N}(\tilde\theta_R - \theta) \to_D N(0, \widehat{\Omega}_\theta^{-1}\Sigma_2\widehat{\Omega}_\theta^{-1})$ for $\lambda = \infty$. □

References

1. Bates, D.M., Watts, D.G.: Nonlinear Regression Analysis and Its Applications. Wiley, New York (1998)
2. Beran, R.J.: Minimum Helinger distance estimates for parametric models. Ann. Statist. **5**, 445–463 (1977)
3. Beran, R.J.: An efficient and robust adaptive estimator of location. Ann. Statist. **6**, 292–313 (1978)
4. Berkson, J.: Are these two regressions? J. Amer. Statist. Assoc. **5**, 164–180 (1950)
5. Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.P.: Measurement Error in Nonlinear Models: A Modern Perspective, 2nd edn. Chapman & Hall/CRC, Bota Raton, FL (2006)
6. Cheng, C.L., Van Ness, J.W.: Statistical Regression with Measurement Error. Wiley, New York (1999)
7. Donoho, D.L., Liu, R.C.: Pathologies of some minimum distance estimators. Ann. Statist. **16**, 587–608 (1988a)
8. Donoho, D.L., Liu, R.C.: The "automatic" robustness of minimum distance functionals. Ann. Statist. **16**, 552–586 (1988b)
9. Fuller, W.A.: Measurement Error Models. Wiley, New York (1987)
10. Hájek, J.: Nonparametric Statistics. Holden Day, San Francisco, USA (1969)
11. Hodges Jr., J.L., Lehmann, E.L.: Estimates of location based on rank tests. Ann. Math. Statist. **34**, 598–611 (1963)

12. Koul, Hira L.: Weighted empirical processes and the regression model. J. Indian Statist. Assoc. **17**, 83–91 (1979)
13. Koul, Hira L.: Minimum distance estimation in multiple linear regression. Sankhya Ser. A., 47. Part **1**, 57–74 (1985a)
14. Koul, Hira L.: Minimum distance estimation in linear regression with unknown errors. Statist. Prob. Lett. **3**, 1–8 (1985b)
15. Koul, Hira L.: Asymptotics of some estimators and sequential residual empiricals in non-linear time series. Ann. Statist. **24**, 380–404 (1996)
16. Koul, H.L.: Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture Notes Series in Statistics, 2nd edn., vol. 166. Springer, New York (2002)
17. Lehmann, E.L.: Elements of Large-Sample Theory. Springer, New York, N.Y., USA (1999)
18. Rudin, W.: Real and Complex Analysis, 2nd edn. McGraw-Hill, New York (1974)
19. Yi, G.: Statistical analysis with measurement error or misclassification. Strategy, method and application. With a foreword by Raymond J. Carroll. In: Springer Series in Statistics. Springer, New York (2017)