# Weighted empirical minimum distance estimators in Berkson measurement error regression models

Hira L Koul
Michigan State University

**Abstract**

We develop analogs of the two classes of weighted empirical 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. In the case of linear model, no knowledge of the measurement error distribution is required while for the non-linear models we need to assume that such distribution is known. 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.

# 1 Introduction

Statistical literature is replete with the various minimum distance estimation methods in the one and two sample location parameter models. Beran (1977, 1978) and Donoho and Liu (1988a,b) argue that the minimum distance estimators based on $L_2$ distances involving either density estimators or residual empirical distribution functions have some desirable finite sample properties, tend to be robust against some contaminated models and are also asymptotically efficient at some models.

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 (1979, 1985a,b, 1996). 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 (1987), Cheng and Van Ness (1999) and Carroll et al. (2006) discuss numerous examples of the measurement error regression models of practical importance.

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 Section 3. The two classes of m.d. estimators are developed. One assumes the symmetry of the errors 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 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 nonlinear Berkson measurement error regression models, where now the knowledge of the measurement error distribution is needed. Again, the estimators based on residual ranks do neither need the symmetry nor the knowledge of the regression error distribution. Some 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

$$(2.1) \qquad Y = X'\theta + \varepsilon, \quad \varepsilon \text{ independent of } X \text{ and symmetrically distributed around } 0.$$

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 $\theta$ based on weighted empirical processes of the residuals and residual ranks were developed in Koul (1979, 1985a,b, 1996). To describe these estimators, let $G$ be a nondecreasing right continuous function from $\mathbb{R}$ to $\mathbb{R}$ having left limits and define

$$\begin{aligned} V(x, \vartheta) &:= n^{-1/2} \sum_{i=1}^{n} X_i \big\{ I(Y_i - X_i'\vartheta \le x) - I(-Y_i + X_i'\vartheta < x) \big\}, \\ M(\vartheta) &:= \int \big\| V(x, \vartheta) \big\|^2 dG(x), \qquad \hat{\theta} := \operatorname{argmin}_{\vartheta \in \mathbb{R}^p} M(\vartheta). \end{aligned}$$

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.

Another class of 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 $X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})'$, $i = 1, \cdots, n$. Let $\bar{X}_j := n^{-1} \sum_{i=1}^{n} X_{ij}$, and $\bar{X} := (\bar{X}_1, \cdots, \bar{X}_p)'$. Let $R_{i\vartheta}$ denote the rank of the $i$th residual $Y_i - X_i'\vartheta$ among $Y_j - X_j'\vartheta$, $j = 1, \cdots, n$. Let $\Psi$ be a distribution function on $[0, 1]$ and define

$$\mathcal{V}(u, \vartheta) := n^{-1/2} \sum_{i=1}^{n} \big( X_i - \bar{X} \big) I(R_{i\vartheta} \le nu), \qquad K(\vartheta) := \int_0^1 \big\| \mathcal{V}(u, \vartheta) \big\|^2 d\Psi(u),$$
$$\hat{\theta}_R := \operatorname{argmin}_{\vartheta \in \mathbb{R}^p} K(\vartheta).$$

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

$$\begin{aligned} V_c(x,\vartheta) &:= n^{-1/2}\sum_{i=1}^n \big(X_i-\bar X\big) I(Y_i - X_i'\vartheta \le x), \quad M_c(\vartheta) := \int \|V_c(x,\vartheta)\|^2 dx, \\ \hat\theta_c &:= \text{argmin}_{\vartheta\in\mathbb{R}^p} M_c(\vartheta). \end{aligned}$$

If one reduces the model (2.1) 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 (1996) and Koul (2002, Sec. 5.4).

**Lemma 2.1** *Suppose the model (2.1) 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.*

(2.2) $\quad G$ *is a nondecreasing right continuous function on* $\mathbb{R}$ *to* $\mathbb{R}$,
$\quad dG(x) = -dG(-x)$, $\forall\, x\in\mathbb{R}$.

(2.3) $$0 < \int f^j dG < \infty, \quad \lim_{z\to 0}\int f^j(x+z)dG(x) = \int f^j(x) dG(x), \quad j=1,2.$$
$$\int_0^\infty (1-F)dG < \infty.$$

*Then* $n^{1/2}(\hat\theta-\theta) \to_D N\big(0,\sigma_G^2\Sigma_X^{-1}\big)$, *where*

$$\sigma_G^2 := \frac{\text{Var}\big(\int_{-\infty}^{\varepsilon} f(x)dG(x)\big)}{\big(\int f^2 dG\big)^2}.$$

(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* $\Psi$ *is a d.f. on* $[0,1]$. *Then* $n^{1/2}\big(\hat\theta_R - \theta\big) \to_D N(0,\gamma_\Psi^2\Omega^{-1})$, *where*

$$\gamma_\Psi^2 := \frac{\text{Var}\big(\int_0^{F(\varepsilon)} f(F^{-1}(s))d\Psi(s)\big)}{\big(\int_0^1 f^2(F^{-1}(s))d\Psi(s)\big)^2}.$$

(c). *In addition, suppose* $\Omega$ *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*

$$\sigma_I^2 := \frac{1}{12\big(\int f^2(x)dx\big)^2}.$$

Before proceeding further we now describe some comparison of the above asymptotic variances. Let $\sigma^2_{LAD}$ and $\sigma^2_{LSE}$ denote the factors of the asymptotic covariance matrices of the LAD and the least squares estimators, respectively. That is

$$\sigma^2_{LAD} := \frac{1}{4f^2(0)}, \quad \sigma^2_{LSE} := \text{Var}(\varepsilon).$$

Let $\gamma_I^2$ denote the $\gamma_\Psi^2$ when $\Psi(s) \equiv s$. Then

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

Table 1 below, obtained from Koul (1992, 2002), gives the values of these factors for some distributions $F$. From this table one sees that the estimator $\hat{\theta}_R$ corresponding to $\Psi(s) \equiv s$ 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.

**Table 1**

| F | $\gamma_I^2$ | $\sigma_I^2$ | $\sigma_{LAD}^2$ | $\sigma_{LSE}^2$ |
|---|---|---|---|---|
| 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 | $\infty$ |

As argued in Koul (Ch. 5, 2002), the m.d. estimators $\hat{\theta}_G$, when $G$ is 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 (2.1) and where, instead of observing $X$, one observes a surrogate $Z$ obeying the relation

$$X = Z + \eta. \tag{3.1}$$

In (3.1), $Z, \eta, \varepsilon$ are assumed to be mutually independent and $E(\eta) = 0$. Note that $\eta$ 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 analogs of the estimators $\hat{\theta}$ in the model (2.1) and (3.1). Rewrite the Berkson ME linear regression model (2.1) and (3.1) as

$$Y = Z'\theta + \xi, \qquad \xi := \eta'\theta + \varepsilon, \qquad \exists\, \theta \in \mathbb{R}. \tag{3.2}$$

Because $Z, \eta, \varepsilon$ are mutually independent, $\xi$ is independent of $Z$ in (3.2).

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

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

is also continuous and symmetric around zero. To see the symmetry, note that by the symmetry of $F$ and a change of variable from $v$ to $-u$ and using $dH(-u) = -dH(u)$, for all $u \in \mathbb{R}^p$, we obtain

$$\begin{aligned} L(-x) &= \int F(-x - v'\theta)dH(v) = 1 - \int F(x + v'\theta)dH(v) = 1 - \int F(x - u'\theta)dH(u) \\ &= 1 - L(x), \qquad \forall\, x \in \mathbb{R}. \end{aligned}$$

This symmetry in turn motivates the following definition of the class of m.d. estimators of $\theta$ in the model (3.2), which mimics the definition of $\hat{\theta}$ by simply replacing $X_i$ by $Z_i$. Define

$$\begin{aligned} \widetilde{V}(x,t) &:= n^{-1/2} \sum_{i=1}^{n} Z_i \big\{ I(Y_i - Z_i' t \le x) - I(-Y_i + Z_i' t < x) \big\}, \\ \widetilde{M}(t) &:= \int \big\| \widetilde{V}(x,t) \big\|^2 dG(x), \qquad \widetilde{\theta} := \text{argmin}_{t \in \mathbb{R}^p} \widetilde{M}(t). \end{aligned}$$

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

To describe the asymptotic normality of $\widetilde{\theta}$, we make the following assumptions.

(3.3) $E\|Z\|^2 < \infty$ and $\Gamma := EZZ'$ is positive definite.

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

(3.5) $F$ has Lebesgue density $f$, symmetric around zero, and such that the density $\ell(x) = \int f(x - v'\theta)dH(v)$ of $L$ satisfies the following:

(a) $\lim_{z \to 0} \int \ell(y+z)dG(y) = \int \ell(y)dG(y) < \infty, \quad 0 < \int \ell^2 dG < \infty,$

(b) $\lim_{z \to 0} \int \big[\ell(y+z) - \ell(y)\big]^2 dG(y) = 0.$

(3.6) $A := \int_0^\infty (1 - L)dG < \infty.$

Under (3.3), $n^{-1}\sum_{i=1}^n Z_i Z_i' \to_p \Gamma$ and $n^{-1/2}\max_{1 \le i \le n} \|Z_i\| \to_p 0$. Using these facts and arguing as in Koul (1996), one deduces that under (2.2) and the above conditions,

$$n^{1/2}\big(\widetilde{\theta} - \theta\big) \to_D \mathcal{N}(0, \tau_G^2 \Gamma^{-1}), \qquad \tau_G^2 := \frac{\text{Var}\big(\int_{-\infty}^{\xi} \ell dG\big)}{\big(\int \ell^2 dG\big)^2}. \tag{3.7}$$

Because in the no measurement error case the estimator corresponding to $G(y) \equiv y$ is an analog of the Hogdes-Lehmann estimator while the one corresponding to the $G(y) \equiv \delta_0(y)$- the degenerate measure at zero, is the LAD estimator, it is of interest to investigate some sufficient conditions that imply conditions (3.5) and (3.6) for these two important and interesting estimators.

Consider the case $G(y) \equiv y$. Assume $f$ to be continuous and $\int f^2(y)dy < \infty$. Then because $H$ is a d.f., $\ell$ is also continuous and symmetric around zero and $\int \ell(y+z)dy =$

$\int \ell(y)dy = 1$. Moreover, by the C-S inequality and Fubini's Theorem,

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

Finally, because $\ell \in L_2$, by Theorem 9.5 in Rudin (1974), it is shift continuous in $L_2$, i.e., (3.5)(b) holds. Hence all conditions of (3.5) are satisfied.

Next, consider (3.6). Note that $E(\varepsilon) = 0$ and $E(\eta) = 0$ imply that $\int |x| f(x)dx < \infty$, $\int \|v\| dH(v) < \infty$ and hence

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

This in turn implies (3.6) in the case $G(y) \equiv y$.

To summarise, in the case $G(y) \equiv y$, (3.3), (3.4), and $F$ having continuous symmetric square integrable density $f$ implies all of the above condition needed for the asymptotic normality of the above analog of the Hodges-Lehmann estimator in the Berkson measurement error 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 measurement error. But, unlike in the no measurement error 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 $\theta$. If $H$ is degenerate at zero, i.e., if there is no measurement error, then $\tau_I^2 = \sigma_I^2$, the asymptotic variance of the Hodges-Lehman estimator of the one sample location parameter not depending on $\theta$.

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

$$
\ell(0) := \int f(v'\theta)dH(v) > 0. \tag{3.8}
$$

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

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

Moreover, here $\int_0^\infty (1 - L)dG = 1 - L(0) = 1/2$ so that (3.6) is also satisfied.

To summarize, in the case $G(y) = \delta_0(y)$, (3.3), (3.4), (3.8) and $f$ being continuous, symmetric around zero and bounded from the above implies 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

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

Consequently, here the asymptotic covariance matrix also depends on $\theta$ via

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

Again, in the case of no measurement error, $\Gamma^{-1}\tau_0^2$ 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 measurement error.

**Analog of $\hat{\theta}_R$.** Here we shall now describe the analogs of the class of estimators $\hat{\theta}_R$ based on the residual ranks obtained from the model (3.2). These estimators do not need the errors $\xi_i$'s to be symmetrically distributed. Let $\widetilde{R}_{i\vartheta}$ denote the rank of $Y_i - Z_i'\vartheta$ among $Y_j - Z_j'\vartheta$, $j = 1, \cdots, n$, and define

$$\begin{aligned} \widetilde{\mathcal{V}}(u,\vartheta) &:= n^{-1/2}\sum_{i=1}^{n}\big(Z_i - \bar{Z}\big)I(\widetilde{R}_{i\vartheta} \leq nu), \qquad \widetilde{K}(\vartheta) := \int_0^1 \|\widetilde{\mathcal{V}}(u,\vartheta)\|^2 d\Psi(u), \\ \widetilde{\theta}_R &:= \mathrm{argmin}_{\vartheta\in\mathbb{R}^p}\widetilde{K}(\vartheta). \end{aligned}$$

Let $Z_{ic} := Z_i - \bar{Z}$. Using the fact the $Z_{ic}$ are centered and for any real numbers $a, b$, $\Psi(\max(a,b)) = \max\{\Psi(a), \Psi(b)\}$ and that $\max(a,b) = 2^{-1}[a + b + |a - b|]$, one obtains a computational form of $\widetilde{K}(t)$ given as follows.

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

The following result can be deduced from Koul (1996). Assume that the density $\ell$ of the r.v. $\xi$ is uniformly continuous and bounded, $E\|Z\|^2 < \infty$, $\widetilde{\Gamma} := E(Z - EZ)(Z - EZ)'$ is positive definite and $\int_0^1 \ell^2(L^{-1}(s))d\Psi(s) > 0$. Then $n^{-1/2}\max_{1\leq i\leq n}\|Z_i\| \to_p 0$, and $n^{-1}\sum_{i=1}^{n}(Z_i - \bar{Z})(Z_i - \bar{Z})' \to_p \widetilde{\Gamma}$. Moreover,

$$n^{1/2}\big(\widetilde{\theta}_R - \theta\big) \to_D N\big(0, \widetilde{\tau}_\Psi^2\widetilde{\Gamma}^{-1}\big), \qquad \widetilde{\tau}_\Psi^2 := \frac{\mathrm{Var}\big(\int_0^{L(\xi)}\ell(L^{-1}(s))d\Psi(s)\big)}{\big(\int_0^1 \ell^2(L^{-1}(s))d\Psi(s)\big)^2}. \tag{3.9}$$

Note that density $f$ of $F$ being uniformly continuous and bounded implies the same for the density $\ell(x) = \int f(x - v'\theta)dH(v)$. 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,

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

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

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

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. Accordingly, 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 response variable $Y$, $p$-dimensional covariate $X$ and its surrogate $Z$ obey the relations

$$(4.1) \qquad Y = m_\theta(X) + \varepsilon, \qquad X = Z + \eta,$$

for some $\theta \in \Theta$. Here $\varepsilon, Z, \eta$ are assumed to be mutually independent, $E\varepsilon = 0$ and $E\eta = 0$. Moreover, $m_\vartheta(x)$ 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 $\vartheta \in \Theta$. Unlike in the linear case, in the absence of any other additional information, here we need to assume that the d.f. $H$ of $\eta$ is known.

Fix a $\theta$ for which (4.1) holds. Let $\nu_\vartheta(z) := E(m_\vartheta(X)|Z = z)$, $\vartheta \in \mathbb{R}^q, z \in \mathbb{R}^p$. Under (4.1), $E(Y|Z = z) \equiv \nu_\theta(z)$. Moreover, because $H$ is known,

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

is a known parametric regression function. Thus, under (4.1), we have the regression model

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

Unlike in the linear case, the error $\zeta$ is no longer independent of $Z$.

To proceed further we need to assume the following. 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 $0 < b < \infty$,

$$(4.2) \qquad \max_{1\le i\le n, n^{1/2}\|\vartheta-\theta\|\le b} n^{1/2}\big|\nu_\vartheta(Z_i) - \nu_\theta(Z_i) - (\vartheta-\theta)'\dot{\nu}_\theta(Z_i)\big| = o_p(1),$$

$$(4.3) \qquad E\|\dot{\nu}_\theta(Z)\|^2 < \infty.$$

Let

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

Assume that

(4.4) For every $z \in \mathbb{R}^p$, $L_z(\cdot)$ is continuous and $L_z(x) = 1 - L_z(-x)$, $\forall\, x \in \mathbb{R}^p$.

We are now ready to define analogs of $\hat{\theta}$ here. Let $G$ be as before and define

$$\begin{aligned} U(x, \vartheta) &:= n^{-1/2} \sum_{i=1}^n \dot{\nu}_\vartheta(Z_i) \big\{ I(Y_i - \nu_\vartheta(Z_i) \le x) - I(-Y_i + \nu_\vartheta(Z_i) < x) \big\} \\ D(\vartheta) &:= \int \big\| U(x, \vartheta) \big\|^2 dG(x), \qquad \widehat{\theta} := \text{argmin}_\vartheta D(\vartheta). \end{aligned}$$

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 (4.1). Let $S_{i\vartheta}$ denote the rank of $Y_i - \nu_\vartheta(Z_i)$ among $Y_j - \nu_\vartheta(Z_j)$, $j = 1, \cdots, n$ and define

$$\begin{aligned} \mathcal{U}_n(u, \vartheta) &:= n^{-1/2} \sum_{i=1}^n \dot{\nu}_\vartheta(Z_i) \big\{ I(S_{i\vartheta} \le nu) - u \big\}, \qquad \mathcal{K}(\vartheta) := \int_0^1 \|\mathcal{U}_n(u, \vartheta)\|^2 d\Psi(u), \\ \widehat{\theta}_R &:= \text{argmin}_\vartheta \mathcal{M}(\vartheta). \end{aligned}$$

The estimator $\widehat{\theta}_R$ gives an analog of the estimator $\hat{\theta}_R$ 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 Section 5.4 of Koul (2002). This method requires the two steps. In the first step we need to show that the defining dispersions $D(\vartheta)$ and $\mathcal{M}(\vartheta)$ 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 $\nu_{nt}(z) := \nu_{\theta + n^{-1/2}t}(z)$, $\dot{\nu}_{nt}(z) := \dot{\nu}_{\theta + n^{-1/2}t}(z)$, and $\dot{\nu}_{ntj}(z)$ denote the $j$th coordinate of $\dot{\nu}_{nt}(z)$. For any real number $a$, let $a^\pm = \max(0, \pm a)$ so that $a = a^+ - a^-$. Also, let

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

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

$$D(\vartheta) \equiv 2 \int_0^\infty \big\| U(x, \vartheta) \big\|^2 dG(x) \equiv 2\widetilde{D}(\vartheta), \qquad \text{say.} \tag{4.5}$$

We are now ready to state our assumptions.

(4.6) $\int_0^\infty E\Big(\|\dot\nu_\theta(Z)\|^2\big(1-L_Z(x)\big)\Big)dG(x)<\infty.$

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

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

(4.9) Density $\ell_z$ of $L_z$ exists for all $z\in\mathbb{R}^p$ and satisfies $0<\int\ell_z(x)dG(x)<\infty$, $\forall\, z\in\mathbb{R}^p$,
$\int E\big(\|\dot\nu_\theta(Z)\|^2\ell_Z^j(x)\big)dG(x)<\infty,\ j=1,2,$ and $\int E(\ell_Z^2(x))dG(x)<\infty.$

(4.10) For all $z\in\mathbb{R}^p$, $\lim_{u\to 0}\int_{-\infty}^\infty\big(\ell_z(x+u)-\ell_z(x)\big)^2dG(x)=0.$

(4.11) $\lim_{u\to 0}\int_{-\infty}^\infty E\Big(\|\dot\nu_{nt}(Z)\|^2\ell_Z(x+u)\Big)dG(x)=\int_{-\infty}^\infty E\big(\|\dot\nu_{nt}(Z)\|^2\ell_Z(x)\big)dG(x)$, $\forall\, t\in\mathbb{R}^q$.

(4.12) With $\xi_t(z):=\nu_{nt}(z)-\nu_\theta(z)$, $E\Big(\int_{-|\xi_t(Z)|}^{|\xi_t(Z)|}\|\dot\nu_{nt}(Z)\|^2\int_{-\infty}^\infty\ell_Z(x+u)dG(x)du\Big)\to 0,$
$\forall\, t\in\mathbb{R}^q, z\in\mathbb{R}^p$.

(4.13) With $\Gamma_\theta(x):=E\big(\dot\nu_\theta(Z)\dot\nu_\theta(Z)'\ell_Z(x)\big)$, the matrix $\Omega_\theta:=\int_{-\infty}^\infty\Gamma_\theta(x)\Gamma_\theta(x)'dG(x)$
is positive definite.

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

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

(4.15) $P\Big(\sup_{\|t-s\|<\delta}n^{-1}\int_0^\infty\big\|\sum_{i=1}^n\{\dot\nu_{nt}(Z_i)-\dot\nu_{ns}(Z_i)\}\beta_i(x)\big\|^2dG(x)>\epsilon\Big)<\epsilon,\quad\forall\, n>N_\epsilon.$

(4.16) For every $\epsilon>0,\alpha>0$ there exists $N_\epsilon$ and $b=b_{\alpha,\epsilon}$ such that
$P(\inf_{\|t\|>b}D(\theta+n^{-1/2}t)\ge\alpha)\ge 1-\epsilon,\quad\forall\, n>N.$

From now onwards we shall write $\nu$ and $\dot\nu$ for $\nu_\theta$ and $\dot\nu_\theta$, respectively.

First, we show that $E(D(\theta))<\infty$, so that by the Markov inequality, $D(\theta)$ is bounded in probability. To see this, by (4.4), $EU(x,\theta)\equiv 0$ and, for $x\ge 0$,

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

By Fubini Theorem and (4.6),

$$(4.17)\qquad E(D(\theta))=2E(\widetilde D(\theta))=4\int_0^\infty E\Big(\big\|\dot\nu(Z)\|^2\big(1-L_Z(x)\big)\Big)dG(x)<\infty.$$

To state the AULQ result for $D$, we need some more notation. With $\Gamma_\theta(x)$ and $\Omega_\theta$ as in (4.13), define

$$(4.18)\qquad W(x,0):=n^{-1/2}\sum_{i=1}^n\dot\nu(Z_i)\big\{I\big(\zeta_i\le x)-L_{Z_i}(x)\big\},$$

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

Note that for any function $K(\vartheta)$, $\sup_{\vartheta\in\mathcal{N}_n(b)} K(\vartheta) = \sup_{\|t\|\le b} K(\theta+n^{-1/2}t)$. We are ready to state the following lemma.

**Lemma 4.1** *Suppose the above set up and assumptions (4.6)– (4.15) hold. Then for every $b<\infty$,*

$$\sup_{\|t\|\le b}\left|D(\theta+n^{-1/2}t)-D(\theta)-4T_n't-4t'\Omega_\theta t\right| \to_p 0. \tag{4.19}$$

*If in addition (4.16) holds, then*

$$\|n^{1/2}(\widetilde{\theta}-\theta)-\widetilde{t}\| \to_p 0. \tag{4.20}$$

$$n^{1/2}(\widetilde{\theta}-\theta) \to_D N\big(0, 4^{-1}\Omega_\theta^{-1}\Sigma_\theta\Omega_\theta^{-1}\big). \tag{4.21}$$

**Proof.** The proof of (4.19) appears in the last section. The proof of the claim (4.20), which uses (4.16), (4.17) and (4.19), is similar to that of Theorem 5.4.1 of Koul (2002).

Define

$$\psi_u(y) \;:=\; \int_{-\infty}^{y}\ell_u(x)dG(x), \quad y\in\mathbb{R},\, u\in\mathbb{R}^p.$$

By (4.9), $0<\psi_u(y)\le\psi_u(\infty)=\int_{-\infty}^{\infty}\ell_u(x)dG(x)<\infty$, for all $u\in\mathbb{R}^p$. Thus for each $u$, $\psi_u(y)$ is an increasing continuous bounded function of $y$ and $\psi_u(-y)\equiv\psi_u(\infty)-\psi_u(y)$, for all $y\in\mathbb{R}$. Let $\varphi_u(y):=\psi_u(-y)-\psi_u(y)=\psi_u(\infty)-2\psi_u(y)$. By (4.4), $E(\varphi_u(\zeta)|Z=z)=0$, 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 $\Gamma_\theta(x)=\int\mu(z)\ell_z(x)dQ(z)$. By the Fubini Theorem,

$$\begin{aligned} T_n &:= \int_{-\infty}^{\infty}\Gamma_\theta(x)\big\{W(x,0)+W(-x,0)\big\}dG(x) \\ &= \int\int_{-\infty}^{\infty}\mu(z)\big\{W(x,0)+W(-x,0)\big\}\ell_z(x)dG(x)dQ(z)\\ &= \int\int_{-\infty}^{\infty}\mu(z)\sum_{i=1}^n\dot\nu(Z_i)\big\{I(\zeta_i\le x)-I(-\zeta_i<x)\big\}d\psi_z(x)dQ(z)\\ &= n^{-1/2}\int\mu(z)\sum_{i=1}^n\dot\nu(Z_i)\big\{\psi_z(-\zeta_i)-\psi_z(\zeta_i)\big\}dQ(z)\\ &= n^{-1/2}\sum_{i=1}^n\int\mu(z)\dot\nu(Z_i)\varphi_z(\zeta_i)dQ(z). \end{aligned} \tag{4.22}$$

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

$$\begin{aligned} \Sigma_\theta := ET_nT_n' &= E\Big\{\Big(\int\mu(z)\dot\nu(Z)\varphi_z(\zeta)dQ(z)\Big)\Big(\int\mu(v)\dot\nu(Z)\varphi_v(\zeta)dQ(v)\Big)'\Big\}\\ &= \int\int\mu(z)\mathcal{K}(z,v)\mu(v)'dQ(z)dQ(v). \end{aligned} \tag{4.23}$$

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 $D(\theta)+4T_nt+4t'\Omega_\theta t$ 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 (4.21) now follows from this result and (4.20).

## 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 additional assumptions, where $\mathcal{U}(b) := \{t\in\mathbb{R}^q; \|t\|\le b\}$, and $0<b<\infty$.

$$\text{(4.24)}\qquad \ell_z \text{ is uniformly continuous and bounded for every } z\in\mathbb{R}^p.$$

$$\text{(4.25)}\qquad n^{-1}\sum_{i=1}^n E\|\dot\nu_{nt}(Z_i)-\dot\nu(Z_i)\|^2 \to 0, \quad \forall t\in\mathcal{U}(b).$$

$$\text{(4.26)}\qquad n^{-1/2}\sum_{i=1}^n \|\dot\nu_{nt}(Z_i)-\dot\nu(Z_i)\| = O_p(1), \quad \forall t\in\mathcal{U}(b).$$

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

$$\text{(4.27)}\qquad P\Big(\sup_{t\in\mathcal{U}(b);\|t-s\|\le\delta} n^{-1/2}\sum_{i=1}^n \|\dot\nu_{nt}(Z_i)-\dot\nu_{ns}(Z_i)\| \le \epsilon\Big) > 1-\epsilon, \qquad \forall\, n>n_\epsilon.$$

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

$$\text{(4.28)}\qquad P\Big(\inf_{\|t\|>b}\mathcal{K}(\theta+n^{-1/2}t)\ge\alpha\Big)\ge 1-\epsilon, \quad \forall\, n>N_\epsilon.$$

Let

$$\bar{\dot\nu} := n^{-1}\sum_{i=1}^n \dot\nu(Z_i), \quad \dot\nu^c(Z_i) := \dot\nu(Z_i)-\bar{\dot\nu}, \quad \widehat\Gamma_\theta(u) := E\Big(\dot\nu^c(Z)\dot\nu^c(Z)'\ell_Z(L_Z^{-1}(u))\Big),$$

$$\widehat{\mathcal{U}}(u) := n^{-1/2}\sum_{i=1}^n \dot\nu^c(Z_i)\big\{I(L_{Z_i}(\zeta_i)\le u)-u\big\}, \quad \widehat\Omega_\theta := \int_0^1 \widehat\Gamma_\theta(u)\widehat\Gamma_\theta(u)'d\Psi(u),$$

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

We need to compute the covariance matrix of $\widehat{T}_n$. Let

$$\kappa_z(v) := \int_0^v \ell_z(L_z^{-1}(u))d\Psi(u), \quad z\in\mathbb{R}^p,\ 0\le v\le 1.$$

By (4.24), $\kappa_z$ is a continuous increasing and bounded function on $[0,1]$, for all $z\in\mathbb{R}^p$. Let $\mu^c(z) := \dot\nu^c(z)\dot\nu^c(z)'$. Argue as for (4.22) and use the fact that $\sum_{i=1}^n \dot\nu_c(Z_i)\equiv 0$ to obtain

$$\begin{aligned}
\widehat{T}_n &= \int\int_0^1 \mu^c(z)\widehat{\mathcal{U}}(u)\ell_z(L_z^{-1}(u))d\Psi(u)dQ(z)\\
&= -n^{-1/2}\sum_{i=1}^n \int \mu^c(z)\dot\nu^c(Z_i)\kappa_z\big(L_{Z_i}(\zeta_i)\big)dQ(z).
\end{aligned}$$

Note that the conditional distribution of $L_Z(\zeta)$, given $Z$, is uniform on $[0,1]$. Let $U$ be such a r.v. Define $\widehat{C}_z(s,t) := E\big[\kappa_s(L_Z(\zeta))\kappa_t(L_Z(\zeta))\big|Z=z\big] = E\big[\kappa_s(U)\kappa_t(U)\big]$ and $\widehat{K}(s,t) := E\big(\dot\nu^c(Z)\dot\nu^c(Z)'\widehat{C}_Z(s,t)\big)$. Then arguing as in (4.23), we 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 4.2** *Suppose the nonlinear Berkson measurement error model (4.1) and the assumptions (4.2), (4.3),* (4.24)–(4.27) *hold. Then the following holds.*

$$\sup_{\|t\|\le b}\big|\mathcal{K}(\theta+n^{-1/2}t)-\widehat{K}(t)\big| = o_p(1). \tag{4.29}$$

*In addition, if* (4.28) *holds and* $\widehat{\Omega}_\theta$ *is positive definite then* $n^{1/2}(\widehat{\theta}_R-\theta)\to_d N\big(0,\widehat{\Omega}_\theta^{-1}\widehat{\Sigma}_\theta\widehat{\Omega}_\theta^{-1}\big)$.

The proof of this lemma is similar to that of Theorem 1.2 of Koul (1996), hence no details are given here.

**Remark 4.1** Because of the importance of the estimators $\widehat{\theta}$ when $G(x)=x$, $G(x)=\delta_0(x)$ and $\widehat{\theta}_R$ when $\Psi(u)\equiv u$, it is of interest to restate the given assumptions under the Berskon ME set up for these three estimators. Consider first $\widehat{\theta}$ when $G(x)\equiv x$ and the following assumptions.

(4.30) $\quad E\Big(\|\dot\nu(Z)\|^2E\big(|\zeta|\big|Z\big)\Big) < \infty.$

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

(4.32) $\quad$ Density $\ell_z$ of $L_z$ exists for all $z\in\mathbb{R}^p$ and satisfies $\int \ell_z^2(x)dx<\infty$, $\forall\, z\in\mathbb{R}^p$,

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

(4.33) $\quad E\big(\|\dot\nu_{nt}(Z)\|^2|\nu_{nt}(Z)-\nu(Z)|\big)\to 0.$

In the case $G(x)\equiv x$, (4.6) and (4.7) are implied by (4.30) and (4.31), while (4.32) implies (4.9) and (4.10). Assumption (4.11) is trivially satisfied and (4.12) reduces to (4.33).

Next consider the analog of the LAD estimator, i.e., when $G(x)=\delta_0(x)$ and the following assumptions.

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

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

(4.36) $\quad \Gamma_\theta(0)$ is positive definite.

In this case (4.3), (4.33) and (4.34)–(4.36) together imply the assumptions (4.6)–(4.13). Not much simplification occurs in the remaining three assumptions (4.14)–(4.16).

As far as $\widehat{\theta}_R$ is concerned, there are no changes in the assumptions (4.24) to (4.27), as they do not involve $\Psi(u)$.

**Remark 4.2 Polynomial regression**. Let $h(x) := (h_1(x), \cdots, h_q(x))'$ be a vector of $q$ functions from $\mathbb{R}$ to $\mathbb{R}$ such that $E\|h(X)\|^2 < \infty$. Consider the model (4.1) with $p = 1$ and $m_\theta(x) = \theta' h(x)$. An example of this is the polynomial regression model with Berkson ME, where $h_j(x) = x^j, j = 1, \cdots, q$. Then $\nu_\vartheta(z) = \vartheta'\gamma(z)$, where $\gamma(z) := E\big(h(X)|Z = z\big)$. Thus $\dot{\nu}_\vartheta(z) \equiv \gamma(z)$ and the assumptions (4.2), (4.3), (4.7), (4.8), (4.14) and (4.15) are *a priori* satisfied. Argue as in the proof of Lemma 5.5.4, Koul (2002), to verify that assumption (4.16) holds in this case. To summarize, in this example the asymptotic normality of $\widehat{\theta}$ holds if the following assumptions hold.

$$\int E\big(\|\gamma(Z)\|^2(1 - L_Z(x))dG(x) < \infty, \qquad 0 < \int \ell_z^2(x)dG(x) < \infty, \forall\, z \in \mathbb{R}^p,$$

$$\int E\big(\|\gamma(Z)\|^2\ell_Z^j(x)\big)dG(x) < \infty,\ j = 1, 2, \qquad \int E\big(\ell_Z^2(x)\big)dG(x) < \infty.$$

$$\lim_{u\to 0}\int_{-\infty}^{\infty} E\Big(\|\gamma(Z)\|^2\ell_Z(x+u)\Big)dG(x) = \int E\big(\|\gamma(Z)\|^2\ell_Z(x)\big)dG(x).$$

With $\xi_t(z) := n^{-1/2}\gamma(z)'t$, $E\big(\int_{-|\xi_t(Z)|}^{|\xi_t(Z)|}\|\gamma(Z)\|^2\int \ell_Z(x+u)dG(x)du\big) \to 0$.

With $\Gamma(x) := E\big(\gamma(Z)\gamma(Z)'\ell_Z(x)\big)$, $\Omega := \int \Gamma(x)\Gamma(x)'dG(x)$ is positive definite.

Then $n^{1/2}(\widehat{\theta} - \theta) \to_D N(0, \Omega^{-1}\Sigma\Omega^{-1})$, with $\Sigma := \int\int h(u)h(u)'\mathcal{K}(u, v)h(v)h'(v)dQ(u)dQ(v)$, $\mathcal{K}(u, v) := E\big(h(Z)h(Z)'C_Z(u, v)\big), u, v \in \mathbb{R}^p$, not depending on $\theta$.

In contrast, for the asymptotic normality of $\widehat{\theta}_R$ here, one only needs (4.24) and $\Psi$ to be a d.f. such that $\widehat{\Omega}$ is positive definite. Note that here $\mu^c(z) = \gamma^c(z) := \gamma(z) - \bar{\gamma}$, $\widehat{K}(s, t) := E\big(\gamma^c(Z)\gamma^c(Z)'\widehat{C}_Z(s, t)\big)$, $\widehat{\Sigma} = \int\int \gamma^c(z)\gamma^c(z)'\widehat{K}(z, v)\gamma^c(v)\gamma^c(v)'dQ(z)dQ(v)$, $\widehat{\Gamma}(u) := E\Big(\gamma^c(Z)\gamma^c(Z)'\ell_Z(L_Z^{-1}(u))\Big)$, and $\widehat{\Omega} = \int_0^1 \widehat{\Gamma}(u)\widehat{\Gamma}(u)'d\Psi(u)$ do not depend on $\theta$.

# 5 Appendix

**Proof of (4.19).** Recall the definition of $\widetilde{D}(\vartheta)$ from (4.5). Let $\widetilde{M}(t) = \widetilde{D}(\theta + n^{-1/2}t)$ and define

$$\begin{aligned}
(5.1)\quad & \nu_{nt}(z) := \nu_{\theta+n^{-1/2}t}(z), \qquad \xi_{it} := \nu_{nt}(Z_i) - \nu_\theta(Z_i),\\
& V_s(x, t) := n^{-1/2}\sum_{i=1}^n \dot{\nu}_{ns}(Z_i)I(Y_i - \nu_{nt}(Z_i) \le x) = n^{-1/2}\sum_{i=1}^n \dot{\nu}_{ns}(Z_i)I\big(\zeta_i \le x + \xi_{it}\big),\\
& V(x, t) := n^{-1/2}\sum_{i=1}^n \dot{\nu}(Z_i)I\big(\zeta_i \le x + \xi_{it}\big),\\
& J_s(x, t) := n^{-1/2}\sum_{i=1}^n E\Big(\dot{\nu}_{ns}(Z_i)I\big(\zeta_i \le x + \xi_{it}\big)\big|Z_i\Big) = n^{-1/2}\sum_{i=1}^n \dot{\nu}_{ns}(Z_i)L_{Z_i}(x + \xi_{it}),\\
& J(x, t) := n^{-1/2}\sum_{i=1}^n \dot{\nu}(Z_i)L_{Z_i}(x + \xi_{it}),\\
& W_s(x, t) := V_s(x, t) - J_s(x, t), \qquad W(x, t) := V(x, t) - J(x, t), \quad s, t \in \mathbb{R}^q.
\end{aligned}$$

Note that $EV_s(x,t) \equiv EJ_s(x,t)$, $EW_s(x,t) \equiv 0$. Also, by (4.4),

$$n^{-1/2}\sum_{i=1}^{n} \dot{\nu}_{ns}(Z_i)\big\{L_{Z_i}(x) + L_{Z_i}(-x)\big\} = n^{-1/2}\sum_{i=1}^{n} \dot{\nu}_{ns}(Z_i), \quad \forall\, s \in \mathbb{R}^q, x \in \mathbb{R}^p.$$

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 (4.4), $\gamma_{nt}(x) \equiv \gamma_{nt}(-x)$, $g_n(x) \equiv g_n(-x)$. Use the above notation and facts to rewrite

$$\begin{aligned}
\widetilde{M}(t) &= \int_0^\infty \Big\|V_t(x,t) + V_t(-x,t) - n^{-1/2}\sum_{i=1}^{n} \dot{\nu}_{nt}(Z_i)\Big\|^2 dG(x) \\
&= \int_0^\infty \Big\|\big\{W_t(x,t) - W_t(x,0)\big\} + \big\{W_t(x,0) - W(x,0)\big\} + \big\{W_t(-x,t) - W_t(-x,0)\big\} \\
&\qquad + \big\{W_t(-x,0) - W(-x,0)\big\} + \big\{J_t(x,t) - J_t(x,0) - \gamma_{nt}(x)\big\} \\
&\qquad + \big\{J_t(-x,t) - J_t(-x,0) - \gamma_{nt}(-x)\big\} \\
&\qquad + 2\big\{\gamma_{nt}(x) - g_n(x)t\big\} + \big\{W(x,0) + W(-x,0) + 2g_n(x)t\big\}\Big\|^2 dG(x).
\end{aligned}$$

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

$$\begin{aligned}
M_1(t) &:= \int_0^\infty \big\|W_t(x,t) - W_t(x,0)\big\|^2 dG(x), \quad M_2(t) := \int_0^\infty \big\|W_t(x,0) - W(x,0)\big\|^2 dG(x), \\
M_3(t) &:= \int_0^\infty \big\|W_t(-x,t) - W_t(-x,0)\big\|^2 dG(x), \\
M_4(t) &:= \int_0^\infty \big\|W_t(-x,0) - W(-x,0)\big\|^2 dG(x), \\
M_5(t) &:= \int_0^\infty \big\|J_t(x,t) - J_t(x,0) - \gamma_{nt}(x)\big\|^2 dG(x), \\
M_6(t) &:= \int_0^\infty \big\|J_t(-x,t) - J_t(-x,0) - \gamma_{nt}(-x)\big\|^2 dG(x), \\
M_7(t) &:= 4\int_0^\infty \big\|\gamma_{nt}(x) - g_n(x)t\big\|^2 dG(x), \\
M_8(t) &:= \int_0^\infty \big\|W(x,0) + W(-x,0) + 2g_n(x)t\big\|^2 dG(x).
\end{aligned}$$

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

**Lemma 5.1** *Under the above set up and the assumptions* (4.2) *to* (4.7), $\forall\, 0 < b < \infty$,

$$\sup_{t\in\mathcal{U}(b)} M_j(t) \to_p 0, \quad j = 1, 2, \cdots, 7, \tag{5.2}$$

$$\sup_{t\in\mathcal{U}(b)} M_8(t) = O_p(1). \tag{5.3}$$

Unless mentioned otherwise, all the supremum below are taken over $t \in \mathcal{U}(b)$. Lemma 5.1 together with the C-S inequality implies that the supremum over $t \in \mathcal{U}(b)$ of all the cross product terms tends to zero, in probability. For example, by the C-S inequality

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

by (5.2) used with $j = 1, 5$. Similarly, by (5.2) with $j = 1$ and (5.3),

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

Consequently, we obtain

$$\sup_t \Big|\widetilde{M}(t) - M_8(t)\Big| = o_p(1). \tag{5.4}$$

Now, expand the quadratic in $M_8$ to write

$$\begin{aligned} M_8(t) &:= \int_0^\infty \big\|W(x,0) + W(-x,0)\big\|^2 dG(x) \\ &\quad +4t' \int_0^\infty g_n(x)\big\{W(x,0) + W(-x,0)\big\} dG(x) + 4\int_0^\infty \big(t' g_n(x)\big)^2 dG(x) \\ &= \widetilde{M}(0) + 4t'\widetilde{T}_n + 4\int_0^\infty \big(t' g_n(x)\big)^2 dG(x), \end{aligned} \tag{5.5}$$

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 converdence theorem

$$\sup_t \big\|t'(g_n(x) - \Gamma_\theta(x))\big\| \to_p 0, \quad \forall\, x \in \mathbb{R}; \quad \sup_t \int_0^\infty \big\|t'(g_n(x) - \Gamma_\theta(x))\big\|^2 dG(x) \to_p 0.$$

Moreover, recall $\widetilde{M}(0) = \widetilde{D}(\theta)$, so that by (4.17), $\widetilde{M}(0) = O_p(1)$. These facts together with the C-S inequality imply that

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

These facts combined with (4.13), (5.4), (5.5) yield that

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

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 (4.18). Hence the above expansion is equivalent to the expansion

$$\begin{aligned} &\sup_t \big|\widetilde{D}(\theta + n^{-1/2}t) - \widetilde{D}(\theta) - 2T_n t - 2t'\Omega_\theta t\big| = o_p(1), \\ &\sup_t \big|D(\theta + n^{-1/2}t) - D(\theta) - 4T_n t - 4t'\Omega_\theta t\big| = o_p(1), \end{aligned}$$

which is precisely the claim (4.19).

**Proof of Lemma 5.1.** Let $\delta_{it} := \xi_{it} - n^{-1/2} t' \dot{\nu}(Z_i)$. By (4.2) and (4.3),

$$\max_{1 \le i \le n,\, t} n^{1/2}\big|\delta_{it}\big| = o_p(1), \qquad \max_{1 \le i \le n} n^{-1/2}\|\dot{\nu}(Z_i)\| = o_p(1). \tag{5.6}$$

Hence,

$$\begin{aligned} \max_{1 \le i \le n,\, t} \big|\xi_{it}\big| &\le \max_{1 \le i \le n,\, \|t\| \le b} \big|\delta_{it}\big|. + \max_{1 \le i \le n,\, t} n^{-1/2}\big\|t\big\|\|\dot{\nu}(Z_i)\| \qquad (5.7) \\ &\le o_p(n^{-1/2}) + b \max_{1 \le i \le n} n^{-1/2}\|\dot{\nu}(Z_i)\| = o_p(1), \\ \sum_{i=1}^n \xi_{it}^2 &= \sum_{i=1}^n (\nu_{nt}(Z_i) - \nu(Z_i))^2 = \sum_{i=1}^n \delta_{it}^2 + n^{-1}\sum_{i=1}^n (t'\dot{\nu}(Z_i))^2, \\ \sup_t \sum_{i=1}^n \xi_{it}^2 &\le n \max_{1 \le i \le n,\, \|t\| \le b} \big|\delta_{it}\big|^2 + b^2 n^{-1} \sum_{i=1}^n \|\dot{\nu}(Z_i)\|^2 = O_p(1), \qquad (5.8) \end{aligned}$$

by (4.3).

Moreover, by (4.3) and the Law of Large Numbers,

$$\begin{aligned} &\sup_t \big\|n^{-1/2} \sum_{i=1}^n \dot{\nu}_\theta(Z_i)\xi_{it}\big\| \qquad (5.9) \\ &\le \max_{1 \le i \le n,\, \|t\| \le b} n^{1/2}|\delta_{it}| n^{-1} \sum_{i=1}^n \big\|\dot{\nu}_\theta(Z_i)\big\| + b n^{-1} \big\| \sum_{i=1}^n \dot{\nu}_\theta(Z_i)\dot{\nu}_\theta(Z_i)'\big\| \\ &= o_p(1) + O_p(1) = O_p(1). \end{aligned}$$

These facts will be use in the sequel.

Consider the term $M_7$. Write

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

Hence

$$M_7 = \int_0^\infty \big\|\gamma_{nt}(x)-g_n(x)t\big\|^2 dG(x) \;\le\; 4\{M_{71}(t)+M_{72}(t)+M_{73}(t)\} \tag{5.10}$$

where

$$\begin{aligned}M_{71}(t) &= n^{-1}\int_0^\infty \Big\|\sum_{i=1}^n \big[\dot\nu_{nt}(Z_i)-\dot\nu(Z_i)\big]\delta_{it}\ell_{Z_i}(x)\Big\|^2 dG(x), \\ M_{72}(t) &= n^{-2}\int_0^\infty \Big\|\sum_{i=1}^n \big[\dot\nu_{nt}(Z_i)-\dot\nu(Z_i)\big]\dot\nu(Z_i)'\ell_{Z_i}(x)t\Big\|^2 dG(x), \\ M_{73}(t) &= n^{-1}\int_0^\infty \Big\|\sum_{i=1}^n \dot\nu(Z_i)\delta_{it}\ell_{Z_i}(x)\Big\|^2 dG(x).\end{aligned}$$

But, by (4.8) and (5.6),

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

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 (4.8) and (4.9). Again, by (4.9) and (5.6),

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

These facts together with (5.10) prove (5.2) 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

(5.11) $M_5(t)$

$$
\begin{aligned}
&:= n^{-1}\int_0^\infty \Big\| \sum_{i=1}^n \Big[\dot{\nu}_{nt}(Z_i)L_{Z_i}(x+\xi_{it}) - \dot{\nu}_{nt}(Z_i)L_{Z_i}(x) - \dot{\nu}_{nt}(Z_i)\xi_{it}\ell_{Z_i}(x)\Big]\Big\|^2 dG(x) \\
&= n^{-1}\int_0^\infty \Big\|\sum_{i=1}^n \dot{\nu}_{nt}(Z_i)D_{it}(x)\Big\|^2 dG(x) \le n^{-1}\sum_{i=1}^n \Big\|\dot{\nu}_{nt}(Z_i)\Big\|^2 \int_0^\infty \sum_{i=1}^n D_{it}^2(x)dG(x).
\end{aligned}
$$

By (4.3) and (4.8),

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

By the C-S inequality, Fubini Theorem, (4.10) and (5.8),

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

Upon combining these facts with (5.11) we obtain $\sup_t M_5(t) = o_p(1)$, thereby proving (5.2) 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

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

by (4.12). Thus

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

To prove that this holds uniformly in $\|t\| \le b$, 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 $N_\epsilon$ such that for every $s \in \mathcal{U}(b)$,

$$P\Big(\sup_{\|t-s\|<\delta} \|M_1(t) - M_1(s)| \ge \epsilon\Big) \le \epsilon, \quad \forall\, n > N_\epsilon. \tag{5.13}$$

Let $\dot\nu_{ntj}(z)$ denote the $j$th coordinate of $\dot\nu_{nt}(z)$, $j = 1, \cdots, q$ 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

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

Thus it suffices to prove (5.13) with $M_1$ replaced by $M_{1j}$ for each $j = 1, \cdots, q$.

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 $j = 1, \cdots, q$, write $\dot\nu_{ntj}(Z_i) = \dot\nu^+_{ntj}(Z_i) - \dot\nu^-_{ntj}(Z_i)$ and define

$$\begin{aligned} W_j^\pm(x,t) &:= n^{-1/2}\sum_{i=1}^n \dot\nu^\pm_{ntj}(Z_i)\alpha_i(x,t), \quad D_j^\pm(x,s,t) := W_j^\pm(x,t) - W_j^\pm(x,s), \\ R_j^\pm(s,t) &:= \int_0^\infty \big(D_j^\pm(x,s,t)\big)^2 dG(x). \end{aligned}$$

Then

$$\begin{aligned} &\big|M_{1j}(t) - M_{1j}(s)\big| \\ &= \Big|\int_0^\infty \big(W_j^+(x,t) - W_j^-(x,t)\big)^2 dG(x) - \int_0^\infty \big(W_j^+(x,s) - W_j^-(x,s)\big)^2 dG(x)\Big| \\ &\le \int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x) + \int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x) \\ &\quad +2\Big\{\int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x)\int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x)\Big\}^{1/2} \\ &\quad +2\Big[\Big\{\int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x)\Big\}^{1/2} + \Big\{\int_0^\infty \big(D_j^-(x,s,t)\big)^2 dG(x)\Big\}^{1/2}\Big] M_{1j}^{1/2}(s) \\ &= R_j^+(s,t) + R_j^-(s,t) + 2\big(R_j^+(s,t)R_j^-(s,t)\big)^{1/2} + \big\{(R_j^+(s,t))^{1/2} + (R_j^-(s,t)\big)^{1/2}\big\} M_{1j}^{1/2}(s). \end{aligned} \tag{5.14}$$

Write

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

Hence

$$R_j^+(s,t) \le 2\int_0^\infty \big(D_{j1}^+(x,s,t)\big)^2 dG(x) + 2\int_0^\infty \big(D_{j2}^+(x,s,t)\big)^2 dG(x). \tag{5.15}$$

By (4.14), the first term here satisfies (5.13). 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 (4.8), there exists an $N_\epsilon$ 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

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

Let

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

Using the above inequalities and $\dot\nu_{njs}^+(Z_i)$ being nonnegative we obtain that on $B_n$,

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

Note that $\max_{1\le i\le n}(|\xi_{is}| + \Delta_{ni}) = o_p(1)$. Argue as for (5.12) 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

$$\begin{aligned}
&\int_0^\infty \Big(n^{-1/2}\sum_{i=1}^n \dot\nu_{njs}^+(Z_i)\int_{\xi_{is}-\Delta_{ni}}^{\xi_{is}+\Delta_{ni}} \ell_{Z_i}(x+u)du\, dG(x)\Big)^2 \qquad (5.16) \\
&\le 2n^{-1}\sum_{i=1}^n (\dot\nu_{njs}^+(Z_i))^2 \sum_{i=1}^n \Delta_{ni}\int_{\xi_{is}-\Delta_{ni}}^{\xi_{is}+\Delta_{ni}}\int_0^\infty \big[\ell_{Z_i}^2(x+u) - \ell_{Z_i}^2(x)\big]\, dG(x)\, du \\
&\qquad + 4n^{-1}\sum_{i=1}^n (\dot\nu_{njs}^+(Z_i))^2 \sum_{i=1}^n \Delta_{ni}^2 \int_0^\infty \ell_{Z_i}^2(x) dG(x).
\end{aligned}$$

The first summand in the above bound is bounded above by

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

because the first factor tends to zero in probability by (4.10) 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 summand in the upper bound of (5.16) is bounded from the above by

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

Since the factor multiplying $\delta^2$ is positive, the above term can be made smaller than $\epsilon$ by the choice of $\delta$. Hence (5.13) is satisfied by the second term in the upper bound of (5.15). This then completes the proof of $R_j^+$ satisfying (5.13). The details of the proof for verifying (5.13) for $R_j^-$ are exactly similar. These facts together with the upper bound of (5.14) show that (5.13) is satisfied by $M_{1j}$ for each $j=1,\cdots,q$. We have thus completed the proof of showing $\sup_t M_1(t)=o_p(1)$, thereby proving (5.2) 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

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

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

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

by (4.7). Thus

$$M_2(t)=o_p(1),\qquad \forall\, t\in\mathbb{R}^q. \tag{5.17}$$

To prove this holds uniformly we shall verify (5.13) for $M_2$. Accordingly, let $\delta>0$, $s\in\mathcal{U}(b)$ be fixed and consider $t\in\mathcal{U}(b)$ such that $\|t-s\|<\delta$. Then

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

This bound, (4.15) and (5.17) now readily verifies (5.13) for $M_2$, which also completes the proof of (5.2) for $j = 2$. The proof of (5.2) for $j = 4$ is precisely similar. This in turn completes the proof of Lemma 5.1.

## References


Beran, R.J. (1977). Minimum Helinger distance estimates for parametric models. *Ann. Statist.*, **5**, 445–463.

Beran, R.J. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.*, **6**, 292–313.

Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.P. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective. Second Edition.* Chapman & Hall/CRC, Bota Raton, FL.

Cheng, C.L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error.* Wiley, New York.

Donoho, D.L. and Liu, R.C. (1988a). Pathologies of some minimum distance estimators. *Ann. Statist.*, **16**, 587–608.

Donoho, D.L. and Liu, R.C. (1988b). The "automatic" robustness of minimum distance functionals. *Ann. Statist.*, **16**, 552–586.

Fuller, W.A. (1987). *Measurement Error Models.* Wiley, New York.

Hodges, J.L., Jr. and Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.*, **34**, 598–611.

Koul, Hira L. (1979). Weighted empirical processes and the regression model. *J. Indian Statist. Assoc.*, **17**, 83–91.

Koul, Hira L. (1985a). Minimum distance estimation in multiple linear regression. *Sankhya, Ser. A.*, **47, Part 1**, 57–74.

Koul, Hira L. (1985b). Minimum distance estimation in linear regression with unknown errors. *Statist. & Prob. Letters*, **3**, 1–8.

Koul, Hira L. (1992). *Weighted emprcals and Linear Models.* IMS Lecture Notes-Monograph Series, Hayward CA.

Koul, Hira L. (1996). Asymptotics of some estimators and sequential residual empiricals in non-linear time series. *Ann. Statist.*, **24**, 380–404.

Koul, Hira L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models. 2nd Edition.* **Lecture Notes Series in Statistics, 166,** Springer, New York, N.Y., USA.

Rudin, W. (1974). *Real and Complex Analysis (2nd edition).* New York: McGraw-Hill.

Yi, Grace. (2017). *Statistical analysis with measurement error or misclassification. Strategy, method and application.* With a foreword by Raymond J. Carroll. **Springer Series in Statistics.** Springer, New York.