Weighted empirical minimum distance estimators in linear errors-in-variables regression models

Pei Geng[a,*], Hira L. Koul[b]

[a] *Department of Mathematics, Illinois State University, Normal, IL, US, 61790*
[b] *Department of Statistics and Probability, Michigan State University, East Lansing, 48824*

Abstract

We develop analogs of a class of weighted empirical minimum distance estimators of the underlying parameters in errors-in-variables linear regression models, when the regression error distribution and the conditional distribution of conditionally centered measurement error, given the surrogate, are symmetric around the origin. This class of estimators is defined as the minimizers of integrals of the square of a certain symmetrized weighted empirical process of the residuals. It includes the least absolute deviation and an analog of the Hodges-Lehmann estimators. In this paper we first develop this class of estimators when the distributions of the true covariates and measurement errors are known, and then extend them to the case when these distributions are unknown but validation data is available. Findings of a simulation study that is included show significant superiority of some members of the proposed class of estimators over the bias corrected least squares estimator.

Keywords: Analog of Hodges-Lehmann estimator, Pitman efficiency, validation data.

1. Introduction

Donoho and Liu (1988a,b) argue that in the one and two sample location models the minimum distance estimators based on L_2 distances involving residual empirical distribution functions have some desirable finite sample properties and tend to be automatically robust against some contaminated models. In regression models without measurement error in the covariates, analogs of these estimators of the underlying regression parameters based on certain weighted residual empirical processes were developed in Koul (1979, 1985, 1996). These estimators include least absolute deviation (LAD), analogs of Hodges-Lehmann (H-L) estimators and several other estimators that are robust against outliers in regression errors and asymptotically efficient at some error distributions.

There are numerous practical situations where covariates are not accurately observed. Instead one observes their surrogates with additive errors. The regression models with such covariates are known as the errors-in-variables (EIVs) regression models. Fuller (1987), Cheng and Van Ness (1999), and Carroll et al. (2006) discuss numerous practical examples of these models. In the linear EIVs models, Stefanski (1985) developed the bias corrected least square estimator based on M-estimation, given the measurement error variance is known. Cook and Stefanski (1994) constructed a simulation-based estimation method for parametric measurement error models which is asymptotically equivalent to method-of-moments estimation in linear EIVs modeling. Buonaccorsi (2010) summarizes the moment-based bias corrected estimation for different settings in linear EIVs models. When the covariate is univariate, AI-Sharadqah (2018) proposed an adjusted maximum

[*] Corresponding author
Email addresses: pgeng@ilstu.edu (Pei Geng), koul@stt.msu.edu (Hira L. Koul)

likelihood estimator for Gaussian measurement error when the variance ratio of regression error and measurement error is known.

Given the importance of the EIVs regression models and the above mentioned properties of the above minimum distance (m.d.) estimators, it is desirable to develop their analogs for the EIVs regression models. Section 2 describes the m.d. estimators of interest in the EIVs linear regression model and the assumptions needed for their asymptotic normality when distributions of the true covariate vector X and the measurement error vector U are known. It also includes a discussion of these assumptions for some important cases, an example of distributions of X and U that satisfy the needed assumptions, and a discussion of the Pitman asymptotic relative efficiency of some members of the proposed class of estimators. This relative efficiency of the analog of H-L estimator, relative to the bias corrected least squares (BCLS) estimator, is seen to increase to infinity as the measurement error variance σ_U^2 increases to infinity, when X, U and the regression error are Gaussian r.v.'s. Section 3 generalizes the m.d. estimators under the setting that the distributions of X and U are unknown but validation data is available. Some proofs are deferred to the Appendix.

Section 4 presents the findings of a simulation study that assesses the performance of the empirical bias and root mean squared error (RMSE) of the two members of the proposed class of m.d. estimators, viz, the analog of H-L and LAD estimators, calibrated least squares (LS) and BCLS estimators. In these simulations, the regression error distributions are taken to be Gaussian(0,1), Laplace(0,1) and the t-distribution with 2 degree of freedom denoted by t_2. To assess the effect of the measurement error U on these estimators, we used several values of the measurement error variance σ_U^2. Tables 1–3 and Tables 4–6 below report the findings when the distributions of X and U are known and unknown but validation data is available, respectively.

The analog of H-L and LAD estimators are seen to be relatively much more stable in terms of the empirical bias and RMSE for all chosen values of σ_U^2 and regression error distributions. The analog of the H-L estimator is relatively more stable than the other three estimators when the regression error distribution is t_2. In comparison, RMSE of BCLS estimator is seen to be much larger than that of the other three estimators for the chosen larger values of σ_U^2 and error distributions. A practical example where σ_U^2 can be large is that of self-reported daily dietary intake, which may be recorded by survey participants with large variation due to the lack of precise nutritional knowledge of food intake. In such scenarios, it would be then desirable to use one of the proposed m.d. estimators. For more on simulations see Section 4.

2. Estimators in the linear EIVs model

In this section, we introduce the linear EIVs regression model and a class of m.d. estimators in this model. Consider the linear regression model where for some $\theta \in \mathbb{R}^p$, the response variable Y and the p dimensional predicting covariate vector X obey the relation

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

For an $x \in \mathbb{R}^p$, x' and $\|x\|$ denote its transpose and Euclidean norm, respectively. In the EIVs model of interest, X is the true covariate. Instead one observes a surrogate Z obeying the relation

$$Z = X + U, \tag{2.2}$$

where X, U, ε are assumed to be mutually independent, U is $p \times 1$ vector of errors with $E(U) = 0$, $E\|U\|^2 < \infty$, $E\|X\|^2 < \infty$.

Let $\gamma(z) := E(U|Z = z)$, $h(z) := E(X|Z = z) = z - \gamma(z)$, $z \in \mathbb{R}^p$. Rewrite (2.1)–(2.2) as

$$
\begin{aligned}
(2.3) \qquad Y &= \theta'(Z - E(U|Z)) + \varepsilon - \theta'(U - E(U|Z)) \\
&= \theta' h(Z) + \zeta, \qquad \zeta = \varepsilon - \theta' V, \quad V := U - \gamma(Z), \quad E(\zeta|Z = z) = 0, \ \forall\, z \in \mathbb{R}^p.
\end{aligned}
$$

Thus we have a regression model regressing Y on Z, with the error r.v. ζ uncorrelated with Z and the regression function $\theta' h(Z)$ with h satisfying

$$
(2.4) \qquad E\|h(Z)\|^2 < \infty.
$$

For the time being, assume that the distributions of X and U are known. Then the functions $\gamma(z)$, $h(z)$ and the conditional distribution function (d.f.) H_z of the r.v. V, given $Z = z$, are all known. The d.f. F of ε need not be known. We assume F to have Lebesgue density f and to be symmetric around zero and the conditional d.f. H_z to be symmetric around the origin, i.e., for every $z \in \mathbb{R}^p$, $-dH_z(v) = dH_z(-v)$, for all $v \in \mathbb{R}^p$. Then, because ε is independent of Z and $V = U - \gamma(Z)$, the conditional d.f. and density of ζ, given $Z = z$, respectively, are

$$
\begin{aligned}
K_z(x) &:= P(\zeta \le x|Z = z) = P(\varepsilon - \theta' V \le x|Z = z) = \int F(x + \theta' v) dH_z(v), \\
\kappa_z(x) &:= \int f(x + \theta' v) dH_z(v), \qquad x \in \mathbb{R},\, z \in \mathbb{R}^p.
\end{aligned}
$$

Both satisfy

$$
(2.5) \qquad K_z(x) = 1 - K_z(-x), \quad \kappa_z(x) = \kappa_z(-x), \quad \forall\, x \in \mathbb{R},\, z \in \mathbb{R}^p.
$$

The symmetry of $K_z(\cdot)$ motivates the following definition of a class of m.d. estimators of θ in the model (2.3), similar to the definition in Chapter 5.2 of Koul (2002) when there is no measurement error. Let G be as in (2.9) below and $\{(Y_i, Z_i), 1 \le i \le n\}$ be a random sample from the model (2.1)–(2.2). For $x \in \mathbb{R}, t \in \mathbb{R}^p$, define

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

Before proceeding further, we describe the estimator $\widetilde{\theta}$ corresponding to $G(x) \equiv \delta_0(x)$ – the measure degenerate at 0 and $G(x) \equiv x$. In the case $G(x) \equiv \delta_0(x)$, because of the continuity of the distribution of $\{Y_i, i \ge 1\}$,

$$
(2.7) \qquad M(t) = \Big\| n^{-1/2} \sum_{i=1}^n h(Z_i) \mathrm{sgn}(Y_i - t' h(Z_i)) \Big\|^2, \quad \text{uniformly in } t \in \mathbb{R}^p, \text{ with prob. 1},
$$

so that the corresponding $\widetilde{\theta}$ is the LAD estimator.

If $\big|G(b) - G(a)\big| = \big|G(-b) - G(-a)\big|$, for all $a, b \in \mathbb{R}$, then G is continuous and symmetric around zero and

$$
\begin{aligned}
M(t) = n^{-1} \sum_{i=1}^n \sum_{j=1}^n h(Z_i)' h(Z_j) \Big[&\big| G\big(Y_i - t' h(Z_i)\big) - G\big(-Y_j + t' h(Z_j)\big) \big| \\
&- \big| G\big(Y_i - t' h(Z_i)\big) - G\big(Y_j - t' h(Z_j)\big) \big| \Big].
\end{aligned}
$$

In particular if $G(x) \equiv x$, then

$$(2.8) \qquad M(t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h(Z_i)' h(Z_j) \big[\big| Y_i + Y_j - t'\big(h(Z_i) + h(Z_j)\big) \big| - \big| Y_i - Y_j - t'\big(h(Z_i) - h(Z_j)\big) \big| \big].$$

From this representation one sees that when the EIVs linear regression model is reduced to the one sample location model, i.e., when $p = 1$, $U_i \equiv 0$, $h(Z_i) \equiv Z_i \equiv X_i \equiv 1$, then

$$M(t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Big[\big| Y_i + Y_j - 2t \big| - \big| Y_i - Y_j \big| \Big], \quad \widetilde{\theta} = \text{median}\Big\{ \frac{Y_i + Y_j}{2}, \ 1 \le i \le j \le n \Big\}.$$

This estimator is the well celebrated H-L estimator of the one sample location parameter, see Hodges and Lehmann (1963). For this reason, in general, we call $\widetilde{\theta}$ corresponding to the $G(x) \equiv x$ an analog of the H-L estimator in the EIVs linear regression model. Perhaps it is worth emphasizing that the asymptotic distribution of this estimator under the EIVs linear regression setup does not seem to be currently available in the literature.

We now state the assumptions needed for establishing the asymptotic normality of $\widetilde{\theta}$. In this section, all limits are taken as $n \to \infty$, unless mentioned otherwise.

(2.9) G is a nondecreasing right continuous function on $\mathbb{R}$ to $\mathbb{R}$ having left limits and $dG(x) = -dG(-x)$, for all $x \in \mathbb{R}$.

$$(2.10) \qquad A := \int_0^\infty E\Big(\|h(Z)\|^2 (1 - K_Z(x)) \Big) dG(x) < \infty.$$

(2.11) $dH_z(v) = -dH_z(-v)$, for all $z, v \in \mathbb{R}^p$.

(2.12) F is symmetric around zero and has Lebesgue density f.

$$(2.13) \qquad 0 < \int \kappa_z(x) dG(x) < \infty, \forall\, z \in \mathbb{R}^p, \quad 0 < \int E\big(\|h(Z)\|^2 \kappa_Z^2(x) \big) dG(x) < \infty.$$

$$(2.14) \qquad \lim_{u \to 0} \int E\big(\|h(Z)\|^j \kappa_Z(x + u\|h(Z)\|) \big) dG(x) = \int E\big(\|h(Z)\|^j \kappa_Z(x) \big) dG(x) < \infty, \quad j = 2, 3.$$

$$0 < \lim_{|u| \to 0} \limsup_{n \to \infty} \int E\Big(\|h(Z)\|^4 \kappa_Z^2\big(x + n^{-1/2} s' h(Z) + u\|h(Z)\|\big) \Big) dG(x) = \int E\big(\|h(Z)\|^4 \kappa_Z^2(x) \big) dG(x) < \infty, \quad \forall\, s \in \mathbb{R}^p.$$

Let

$$(2.15) \qquad \Gamma_n(x) := n^{-1} \sum_{i=1}^{n} h(Z_i) h(Z_i)' \kappa_{Z_i}(x), \quad \Gamma(x) := E\big(h(Z) h(Z)' \kappa_Z(x) \big), \quad x \in \mathbb{R},$$

$$\mathcal{G}_n := \int \Gamma_n(x)' \Gamma_n(x) dG(x), \quad \mathcal{G} := \int \Gamma(x)' \Gamma(x) dG(x).$$

Assume

$$(2.16) \qquad \mathcal{G} \text{ is positive definite and } \int \big\| \Gamma_n(x) - \Gamma(x) \big\|^2 dG(x) \to_p 0.$$

Let $c_{ni} \equiv n^{-1/2}h(Z_i)$. Let h_j denote the jth component of h. Write $h_j = h_j^+ - h_j^-$, for $1 \le j \le p$. Assume there exists a constant $0 < C < \infty$ such that $\forall\, \delta > 0$, $0 < b < \infty$, $\|s\| \le b$,

$$\limsup_n \sum_{j=1}^{p} \int E\Big[n^{-1/2} \sum_{i=1}^{n} h_j^{\pm}(Z_i)\Big\{K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) - K_{Z_i}(x + s'c_{ni} - \delta\|c_{ni}\|)\Big\}\Big]^2 dG(x) \le C\delta^2. \tag{2.17}$$

Remark 2.1. We shall now discuss some sufficient conditions for the above assumptions for the three types of the integrating measure G, viz, when G is a d.f. symmetric around zero, when $G(x) \equiv \delta_0(x)$ and when $G(x) \equiv x$. In the absence of measurement error, the estimators corresponding to the latter two choices of G are known to be asymptotically efficient at Laplace and logistic error distributions, respectively, while the estimator $\widetilde{\theta}$ corresponding to a d.f. G is typically robust against gross errors in the error distribution.

Consider the following assumptions.

$$\int \kappa_z(x)dG(x) > 0, \quad \forall\, z \in \mathbb{R}^p; \quad \int E\big(\kappa_Z^2(x)\big)dG(x) > 0. \tag{2.18}$$

$$E\|h(Z)\|^4 < \infty, \quad \mathcal{G} \text{ is positive definite.} \tag{2.19}$$

(2.20) Density f of F is uniformly continuous, bounded and $f(x) \equiv f(-x)$.

Consider the case when G is a d.f. symmetric around zero. Then, because $0 \le 1 - K_z(x) \le 1$, for all x, z, $A \le E\|h(Z)\|^2 < \infty$, by (2.19), thereby verifying (2.10). By (2.20), $\kappa_z(x) = \int f(x + \theta'v)dH_z(v)$ is uniformly continuous in x, uniformly in z, and

$$\sup_{z,x} \kappa_z(x) \le \|f\|_\infty < \infty, \tag{2.21}$$

where for any function ℓ from $\mathbb{R}$ to $\mathbb{R}$, $\|\ell\|_\infty := \sup_{x\in\mathbb{R}} |\ell(x)|$. Thus (2.18) and (2.19) imply (2.13) here. Moreover, by (2.20), $\forall\, s \in \mathbb{R}^p$, $\kappa_Z(x + n^{-1/2}s'h(Z) + u\|h(Z)\|) - \kappa_Z(x) \to_p 0$, as first $n \to \infty$ and then $u \to 0$, and $|\kappa_Z(x+n^{-1/2}s'h(Z)+u\|h(Z)\|)-\kappa_Z(x)| \le 2\|f\|_\infty$. Hence, by the DCT, (2.14) holds. Under (2.19) and (2.20), $\|\Gamma_n(x)-\Gamma(x)\| \le \|f\|_\infty[n^{-1}\sum_{i=1}^n \|h(Z_i)\|^2+E(\|h(Z)\|^2)] = O_p(1)$. By the LLNs, $\|\Gamma_n(x) - \Gamma(x)\| \to_p 0$, for every $x \in \mathbb{R}$. Hence the DCT ensures the satisfaction of (2.16) here. Fact (2.21), G being a d.f. and the Cauchy-Schwarz (C-S) inequality yield that the left hand side of (2.17) is bounded from the above by

$$\begin{aligned} &\limsup_n \sum_{j=1}^{p} E\Big[2n^{-1/2} \sum_{i=1}^{n} h_j^{\pm}(Z_i)\|c_{ni}\|\delta\|f\|_\infty\Big]^2 \\ \le\ & 4\delta^2\|f\|_\infty^2 \limsup_n E\Big(n^{-1}\sum_{i=1}^{n}\|h(Z_i)\|^4\Big) = 4\delta^2\|f\|_\infty^2 E\big(\|h(Z)\|^4\big), \end{aligned}$$

by the LLNs, thereby verifying (2.17). We also use (2.19) here.

Next, consider the case when $G(x) \equiv \delta_0(x)$. Even though this is included in the case of G being a d.f., one can directly verify that in this case the conditions (2.10)–(2.17) are implied by $\kappa_z(0) := \int f(\theta'v)dH_z(v) > 0$, $\forall\, z \in \mathbb{R}^p$, $E\kappa_Z^2(0) > 0$, (2.19) and that f is continuous and bounded on $\mathbb{R}$. The latter condition is less restrictive than (2.20).

Finally, consider the case $G(x) \equiv x$. Here (2.18) is trivially satisfied. Assume (2.19) and (2.20) hold. Note that $E|\zeta|^2 \le E\varepsilon^2 + \|\theta\|^2E\|U\|^2 < \infty$. Then by the C-S inequality, $A \le$

$E\Big(\|h(Z)\|^2 E\big(|\zeta|\big|Z\big)\Big) \le E^{1/2}(\|h(Z)\|^4)E^{1/2}\big(|\zeta|^2\big) < \infty$, by (2.19). Argue as in the case of when G is a d.f. to see that (2.19) and (2.20) imply (2.12)–(2.14) here also.

To verify (2.16), by the LLNs, $\|\Gamma_n(x) - \Gamma(x)\| \to_p 0$, $\forall\, x \in \mathbb{R}$. Also, $\|\Gamma_n(x) - \Gamma(x)\|^2 \le 2n^{-1}\sum_{i=1}^n \|h(Z_i)\|^4 \kappa_{Z_i}^2(x) + 2E(\|h(Z)\|^4\kappa_Z^2(x))$. The fact (2.21) and $\int \kappa_z(x)dx = 1$ imply $\int \kappa_z^2(x)dx \le \|f\|_\infty$, for all $z \in \mathbb{R}^p$. Hence (2.16) follows from the DCT and (2.19).

Next, consider (2.17). By the C-S inequality used twice, once with the sum and once with the integrals and the Fubini Theorem and the fact that $\int \kappa_z^2(x)dx \le \|f\|_\infty, \forall z \in \mathbb{R}^p$, the left hand side of (2.17) is bounded from the above by

$$\limsup_n E\Big(n^{-1}\sum_{i=1}^n \|h(Z_i)\|^2 \int \sum_{i=1}^n \Big\{\int_{-\delta\|c_{ni}\|}^{\delta\|c_{ni}\|} \kappa_{Z_i}(x + s'c_{ni} + u)du\Big\}^2 dx\Big)$$
$$\le 2\limsup_n E\Big(n^{-1}\sum_{i=1}^n \|h(Z_i)\|^2 \sum_{i=1}^n \delta\|c_{ni}\| \int_{-\delta\|c_{ni}\|}^{\delta\|c_{ni}\|} \int \kappa_{Z_i}^2(x + s'c_{ni} + u)dxdu\Big)$$
$$\le 4\delta^2\|f\|_\infty \limsup_n E\Big(n^{-1}\sum_{i=1}^n \|h(Z_i)\|^2\Big)^2 \le 4\delta^2\|f\|_\infty E\|h(Z)\|^4, \quad \because\ c_{ni} = n^{-1/2}h(Z_i),$$

where the last inequality again follows from the C-S inequality. Because of (2.19), this verifies (2.17) in the case $G(x) \equiv x$.

We shall now return to the derivation of the asymptotic normality of $\widetilde{\theta}$. 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 $M(t)$ is AULQ (asymptotically uniformly locally quadratic) in $n^{1/2}(t-\theta)$ for $t \in \mathcal{N}_n(b) := \{t \in \mathbb{R}^p, n^{1/2}\|t-\theta\| \le b\}$, for every $0 < b < \infty$. The second step requires to show that $n^{1/2}\|\widetilde{\theta} - \theta\| = O_p(1)$.

In the current setup, the above mentioned second step is in part implied by having

$$M(\theta) = O_p(1). \tag{2.22}$$

Proof of (2.22). Because of (2.9) and because $V(x,t) \equiv V(-x,t)$,

$$M(t) \equiv 2\int_0^\infty \big\|V(x,t)\big\|^2 dG(x), \quad \forall\, t \in \mathbb{R}^p. \tag{2.23}$$

Moreover, by (2.5),

$$\begin{aligned} EV(x,\theta) &= n^{1/2}E\Big(h(Z)E\big[\big(I(\zeta \le x) - I(-\zeta < x)\big)\big|Z\big]\Big) \\ &= n^{1/2}E\Big(h(Z)\big[K_Z(x) - 1 + K_Z(-x)\big]\Big) = 0, \quad x \in \mathbb{R}, \\ E\big\|V(x,\theta)\big\|^2 &= E\Big(\|h(Z)\|^2 E\big[\big(I(\zeta \le x) - I(-\zeta < x)\big)^2\big|Z\big]\Big) \\ &= 2E\Big(\|h(Z)\|^2(1 - K_Z(x))\Big), \quad x > 0. \end{aligned}$$

Hence, by the Fubini Theorem and (2.10),

$$EM(\theta) = 4\int_0^\infty E\Big(\|h(Z)\|^2(1 - K_Z(x))\Big)dG(x) < \infty.$$

The claim (2.22) follows from this fact and the Markov inequality.

The fact $n^{1/2}(\widetilde{\theta} - \theta) = \text{arginf}_{s\in\mathbb{R}^p} M(\theta + sn^{-1/2})$ motivates the following notation. Let

$$W(x,s) := n^{-1/2}\sum_{i=1}^{n} h(Z_i)\big[I(\zeta_i \le x + s'c_{ni}) - K_{Z_i}(x + s'c_{ni})\big], \quad s \in \mathbb{R}^p, \tag{2.24}$$

$$T_n := \int \Gamma(x)'\big[W(x,0) + W(-x,0)\big]dG(x),$$

$$\widetilde{M}(s) = M(\theta) + 4T_n's + 4s'\mathcal{G}s, \qquad \widetilde{s} := \operatorname{argmin}_s \widetilde{M}(s) = -\frac{1}{2}\mathcal{G}^{-1}T_n.$$

For a positive integer m, let $\mathcal{N}_m(\mu,\Sigma)$ denote the m-dimensional normal distribution with mean vector μ and covariance matrix Σ, $\mathcal{N} = \mathcal{N}_1$. The following theorem describes the asymptotic distribution of $\widetilde{\theta}$.

Theorem 2.1. *Under* (2.1), (2.2), (2.4) *and assumptions (2.9)–(2.17), the following results hold.*

$$\sup_{\|s\|\le b} \big|M(\theta + n^{-1/2}s) - \widetilde{M}(s)\big| \to_p 0, \quad \forall\, 0 < b < \infty. \tag{2.25}$$

$$\text{(a)} \quad \|n^{1/2}(\widetilde{\theta} - \theta) - \widetilde{s}\,\| \to_p 0, \quad \text{(b)} \quad n^{1/2}(\widetilde{\theta} - \theta) \to_D \mathcal{N}_p\big(0, 4^{-1}\mathcal{G}^{-1}\Sigma\mathcal{G}^{-1}\big), \tag{2.26}$$

where Σ *is defined at* (2.28) *below.*

Proof. The proof of (2.25) appears in the Appendix. The proof of the claim (2.26)(a) is similar to that of Theorem 5.4.4 of Koul (2002). The details are omitted for the sake of brevity. To prove (2.26)(b), let

$$\psi_z(x) := \int_{-\infty}^{x} \kappa_z(y)dG(y), \qquad \mu(z) := h(z)h(z)', \quad x \in \mathbb{R},\, z \in \mathbb{R}^p. \tag{2.27}$$

By (2.13), for every $z \in \mathbb{R}^p$, $\psi_z(x)$ is continuous and uniformly bounded in x, $\psi_z(-x) \equiv \psi_z(\infty) - \psi_z(x)$, for all $x \in \mathbb{R}$. Let $\varphi_z(x) := \psi_z(-x) - \psi_z(x) = \psi_z(\infty) - 2\psi_z(x)$. Moreover, with Q denoting the d.f. of Z and using the Fubini Theorem and the definition of $W(x,0) + W(-x,0)$ from (2.24), we obtain the following.

$$\begin{aligned}
T_n &= n^{-1/2}\sum_{i=1}^{n} \int E\big(\mu(Z)\kappa_Z(x)\big)h(Z_i)\big[I(\zeta_i \le x) - I(-\zeta_i < x)\big]dG(x) \\
&= n^{-1/2}\sum_{i=1}^{n} \int\int \mu(z)\kappa_z(x)dQ(z)h(Z_i)\big[I(\zeta_i \le x) - I(-\zeta_i < x)\big]dG(x) \\
&= n^{-1/2}\sum_{i=1}^{n} \int \mu(z)h(Z_i) \int \big[I(\zeta_i \le x) - I(-\zeta_i < x)\big]d\psi_z(x)dQ(z) \\
&= n^{-1/2}\sum_{i=1}^{n} \int \mu(z)h(Z_i)\varphi_z(\zeta_i)dQ(z).
\end{aligned}$$

Let

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

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

$$(2.28)\qquad \Sigma := ET_nT_n' = E\Big\{\Big(\int \mu(z)h(Z)\varphi_z(\zeta)dQ(z)\Big)\Big(\int \mu(v)h(Z)\varphi_v(\zeta)dQ(v)\Big)'\Big\} = \int\int \mu(z)\mathcal{K}(z,v)\mu(v)'dQ(z)dQ(v).$$

Thus T_n is a $p \times 1$ vector of independent centered finite variance r.v.'s. By the classical CLT, $T_n \to_D \mathcal{N}_p(0, \Sigma)$. Hence, the minimizer $\widetilde{s} = (-1/2)\mathcal{G}^{-1}T_n \to_D \mathcal{N}_p\big(0, 4^{-1}\mathcal{G}^{-1}\Sigma\mathcal{G}^{-1}\big)$. The claim (2.26)(b) now follows from this result and (2.26)(a). □

Remark 2.2. Here we shall discuss an example where the conditional distribution of U, given Z, is known and that of V, given Z, does not depend on Z. We shall also discuss the Pitman's asymptotic relative efficiencies of some of the m.d. estimators, relative to the least squares and maximum likelihood estimators, at some error d.f. F.

Example 2.1. Suppose $p = 1$, $X \sim_D \mathcal{N}(a, \sigma_X^2)$, $U \sim_D \mathcal{N}(0, \sigma_U^2)$, with a, σ_X^2 and σ_U^2 known, and X and U are independent r.v.'s. Then $Z = X + U \sim_D \mathcal{N}(a, \sigma_X^2 + \sigma_U^2)$, $\mathrm{Cov}(Z, U) = \sigma_U^2$ so that

$$(Z, U) \sim_D \mathcal{N}_2\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 + \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 \end{pmatrix}\right).$$

Hence, the conditional distribution of U, given $Z = z$, is $\mathcal{N}((z-a)r^2, r^2\sigma_X^2)$, where $r^2 := \sigma_U^2/(\sigma_X^2 + \sigma_U^2)$. Thus, $\gamma(z) = E(U|Z = z) = (z - a)r^2$, $h(z) = z - \gamma(z) = z - r^2(z - a)$. Because $0 \le r^2 < 1$ and $\sigma_X^2 > 0$,

$$(2.29)\qquad Eh^2(Z) = (1 - r^2)^2(\sigma_X^2 + \sigma_U^2) + a^2 \ge (1 - r^2)^2(\sigma_X^2 + \sigma_U^2) > 0.$$

Next, the conditional distribution of $V = U - \gamma(Z) = U - (Z - a)r^2$, given Z, is $\mathcal{N}(0, r^2\sigma_X^2)$, which does not depend on Z and hence V is independent of Z. Thus

$$K_z(x) = P(\zeta \le x|Z = z) = P(\varepsilon - \theta'V|Z = z) = \int F(x + \theta v)d\Phi\big(v/r\sigma_X\big),$$

also does not depend on z, i.e., ζ is also independent of Z. Here Φ denotes the d.f. of a $\mathcal{N}(0, 1)$ r.v.. Write K and κ for K_z and κ_z in this case. Then many assumptions and entities involved in the statement of the asymptotic normality of $\widetilde{\theta}$ simplify as follows. Note that now (2.11) and (2.16) are *a priori* satisfied. The other assumptions (2.10), (2.13), (2.14) and (2.17) are equivalent to the following conditions.

$$(2.30)\qquad A := \int_0^\infty \big(1 - K(x)\big)dG(x) < \infty.$$

(2.31) F has Lebesgue density f, symmetric around zero, and such that the density $\kappa(x) = \int f(x + \theta'v)dH(v)$ of K satisfies the following:

$$0 < \int \kappa^j(x)dG(x) < \infty, \qquad \lim_{u\to 0}\int \big(\kappa(x + u) - \kappa(x)\big)^j dG(x) = 0,\ j = 1, 2.$$

(2.32) Assumption (2.17) with $K_z \equiv K$ holds.

The simplification in the statement of the asymptotic normality of $\widetilde{\theta}$ occurs as follows. To begin with, $\psi_z(x) \equiv \psi(x) \equiv \int_{-\infty}^x \kappa(y)dG(y)$, $\varphi_z(x) \equiv \varphi(x) = \psi(-x) - \psi(x)$, $\Gamma(x) = \Gamma\kappa(x)$, where $\Gamma := Eh^2(Z) > 0$, (see (2.29)), and

$$T_n = \Gamma n^{-1/2}\sum_{i=1}^n h(Z_i)\varphi(\zeta_i), \quad \mathcal{G} = \int \Gamma(x)'\Gamma(x)dG(x) = \Gamma^2 \int \kappa^2(x)dG(x),$$
$$\Sigma = \mathrm{Var}(T_n) = \Gamma^3\mathrm{Var}(\varphi(\zeta)), \ \ \frac{1}{4}\mathcal{G}^{-1}\Sigma\mathcal{G}^{-1} = \frac{\mathrm{Var}(\varphi(\zeta))}{4\Gamma(\int \kappa^2(x)dG(x))^2} = \frac{\mathrm{Var}(\psi(\zeta))}{\Gamma(\int \kappa^2(x)dG(x))^2}.$$

To summarize, under the above normality assumption of X, U, the conditions (2.9) and (2.30)–(2.32) imply that

$$n^{1/2}\big(\widetilde{\theta} - \theta\big) \to_D \mathcal{N}\Big(0, \frac{\tau_G^2}{Eh^2(Z)}\Big), \quad \tau_G^2 := \frac{\mathrm{Var}\big(\int_{-\infty}^{\zeta}\kappa(x)dG(x)\big)}{(\int \kappa^2(x)dG(x))^2}. \tag{2.33}$$

Consider the case when $G(x) \equiv x$. Write $\widetilde{\theta}_I$ for the corresponding m.d. estimator and τ_I^2 for τ_G^2 in this case. Because $K(\zeta)$ is a uniformly distributed r.v. on $[0,1]$,

$$\mathrm{Var}\Big(\int_{-\infty}^{\zeta}\kappa(x)dG(x)\Big) = \mathrm{Var}\Big(\int_{-\infty}^{\zeta}\kappa(x)dx\Big) = \mathrm{Var}(K(\zeta)) = 1/12,$$
$$\int \kappa^2(x)dG(x) = \int \kappa^2(x)dx, \quad \tau_I^2 = \frac{1}{12(\int \kappa^2(x)dx)^2}.$$

In particular, if there is no measurement error, $U = 0$, H is degenerate at zero, $\zeta = \varepsilon$, $\kappa = f$ and $\tau_I^2 = 1/\big(12(\int f^2(x)dx)^2\big)$, the very familiar expression related to the H-L estimator.

Next, consider the case $G(x) \equiv \delta_0(x)$. Write $\widetilde{\theta}_0$ for the corresponding m.d. estimator and τ_0^2 for τ_G^2 in this case. Because ζ is symmetrically distributed around zero, $\mathrm{Var}(I(\zeta > 0)) = 1/4$, $\int \kappa^2(x)dG(x) = \kappa^2(0)$ and

$$\mathrm{Var}\Big(\int_{-\infty}^{\zeta}\kappa(x)dG(x)\Big) = \mathrm{Var}\big(\kappa(0)I(\zeta > 0)\big) = \frac{\kappa^2(0)}{4}, \quad \tau_0^2 = \frac{1}{4\kappa^2(0)}.$$

Again if there is no measurement error then $\tau_0^2 = 1/(4f^2(0))$, which is the well celebrated asymptotic variance of the one sample median or the factor that appears in the asymptotic variance of the LAD estimator in regression models.

We shall now describe the extension of (2.33) when $p > 1$. Let $X \sim_D \mathcal{N}_p(\mu_X, \Sigma_X)$, $U \sim_D \mathcal{N}_p(0, \Sigma_U)$, X, U independent, Σ_X, Σ_U both positive definite and known. Also μ_X is known. Then $Z = X + U \sim_D \mathcal{N}_p(\mu_X, \Sigma_X + \Sigma_U)$, $\mathrm{Cov}(Z, U) = \Sigma_U$ so that

$$(Z, U) \sim_D \mathcal{N}_{2p}\left(\begin{pmatrix} \mu_X \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_X + \Sigma_U & \Sigma_U \\ \Sigma_U & \Sigma_U \end{pmatrix}\right).$$

Hence, the conditional distribution of U, given $Z = z$, is

$$\mathcal{N}_p\Big(\Sigma_U(\Sigma_X + \Sigma_U)^{-1}(Z - \mu_X), \, \Sigma_U - \Sigma_U(\Sigma_X + \Sigma_U)^{-1}\Sigma_U\Big).$$

Let $R := \Sigma_U(\Sigma_X + \Sigma_U)^{-1}$. Then $\gamma(z) = E(U|Z = z) = R(z - \mu_Z)$ and the conditional distribution of $V = U - R(Z - \mu_Z)$, given Z, is $\mathcal{N}_p(0, \Sigma_U - R\Sigma_U)$, which again does not depend on Z and hence V is independent of Z. Now $\Gamma = Eh(Z)h(Z)'$. Like (2.33), under (2.9) and (2.30)–(2.32), $n^{1/2}\big(\widetilde{\theta} - \theta\big) \to_D \mathcal{N}_p\big(0, \tau_G^2\Gamma^{-1}\big)$.

Pitman's Asmyptotic Relative Efficiency (ARE). We shall now compute Pitman's asymptotic relative efficiency of $\widetilde{\theta}_I, \widetilde{\theta}_0$ relative to the least squares estimator and the maximum likelihood estimator in the current setup when $F(x) = \Phi(x/\sigma_\varepsilon)$, i.e., $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. See Lehmann (1999) for the definition of the ARE.

Assume $p = 1$. Let $w^2 := \sigma_\varepsilon^2 + \theta^2 r^2 \sigma_X^2$. Because $V \sim \mathcal{N}(0, r^2\sigma_X^2)$, $\zeta = \varepsilon - \theta V \sim \mathcal{N}(0, w^2)$. Hence $\kappa^2(0) = (2\pi w^2)^{-1}$, $\tau_0^2 = (4\kappa^2(0))^{-1} = (\pi/2)w^2$ and

$$\int \kappa^2(x)dx = \frac{1}{2\sqrt{\pi\, w^2}}, \quad \tau_I^2 = \frac{1}{12(\int \kappa^2(x)dx)^2} = \frac{\pi}{3}w^2.$$

Next, by the CLT and the Slutsky's Theorem, $n^{1/2}\big(\widetilde{\theta}_{\ell se} - \theta\big) \to_D \mathcal{N}\big(0, w^2/Eh^2(Z)\big)$, where $\widetilde{\theta}_{\ell se} := \sum_{i=1}^n h(Z_i)Y_i / \sum_{i=1}^n h^2(Z_i)$ is the least squares estimator of θ in the model (2.3). Hence, $\mathrm{ARE}(\widetilde{\theta}_I, \widetilde{\theta}_{\ell se}) = 3/\pi$, and $\mathrm{ARE}(\widetilde{\theta}_0, \widetilde{\theta}_{\ell se}) = 2/\pi$, which are the same as in the case of no measurement error regression models.

Let $\tau^2 := w^2/(Eh^2(Z) + 2\theta^2 r^4 \sigma_X^4)$ and $\rho := Eh^2(Z)/(Eh^2(Z) + 2\theta^2 r^4\sigma_X^4)$. Under the above normality assumptions on the distributions of ε, U, X, a consistent solution $\hat{\theta}_n$ of the likelihood equation exists and $n^{1/2}(\hat{\theta}_n - \theta) \to_D \mathcal{N}(0, \tau^2)$. Consequently,

$$\mathrm{ARE}(\widetilde{\theta}_I, \hat{\theta}_n) = 3\rho/\pi, \qquad \mathrm{ARE}(\widetilde{\theta}_0, \hat{\theta}_n) = 2\rho/\pi, \qquad \mathrm{ARE}(\widetilde{\theta}_{\ell se}, \hat{\theta}_n) = \rho.$$

Next, consider the bias corrected least squares estimator

$$\widetilde{\theta}_{bc\ell s} := \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) / \Big[\sum_{i=1}^n (Z_i - \bar{Z})^2 - \sigma_U^2\Big].$$

Let

$$\varphi^2 := \big[\sigma_X^2(\sigma_\varepsilon^2 + \theta^2\sigma_U^2) + \sigma_U^2\sigma_\varepsilon^2 + 2\theta^2\sigma_U^4\big]/\sigma_X^4 = \big[w^2(\sigma_X^2 + \sigma_U^2) + 2\theta^2\sigma_U^4\big]/\sigma_X^4.$$

A direct application of the classical CLT and Slutksy's Theorem yields that $n^{1/2}\big(\widetilde{\theta}_{bc\ell s} - \theta\big) \to_D \mathcal{N}\big(0, \varphi^2\big)$. See also Cheng and Van Ness (1999). Hence $\mathrm{ARE}(\widetilde{\theta}_I, \widetilde{\theta}_{bc\ell s}) = (3/\pi)(\varphi^2/w^2)$. Because $r^2 \to 0, 1$, as $\sigma_U^2 \to 0, \infty$, respectively, we obtain that $w^2 = \sigma_\varepsilon^2 + \theta^2 r^2 \sigma_X^2 \to \sigma_\varepsilon^2$, $\sigma_\varepsilon^2 + \theta^2\sigma_X^2$, as $\sigma_U^2 \to 0, \infty$, respectively. Moreover, the derivative of the ratio σ_U^4/w^2 with respect to σ_U^2 is positive for all values of σ_U^2 so that this ratio is an increasing function of σ_U^2. Thus the

$$\mathrm{ARE}(\widetilde{\theta}_I, \widetilde{\theta}_{bc\ell s}) \;=\; \frac{3}{\pi}\frac{\varphi^2}{w^2} = \frac{3}{\pi}\Big\{\frac{\sigma_X^2 + \sigma_U^2}{\sigma_X^4} + 2\theta^2\frac{\sigma_U^4}{\sigma_X^4 w^2}\Big\}$$

is an increasing function of σ_U^2 and $\mathrm{ARE}(\widetilde{\theta}_I, \widetilde{\theta}_{bc\ell s}) \to (3/\pi)\sigma_X^{-2}$, ∞, as $\sigma_U^2 \to 0$, $\sigma_U^2 \to \infty$, respectively. In other words, the estimator $\widetilde{\theta}_I$ is far more efficient, compared to the BCLS, against the increasing measurement error. A similar statement holds about the $\mathrm{ARE}(\widetilde{\theta}_0, \widetilde{\theta}_{bc\ell s})$.

3. M.D. Estimation with validation data

In this section, we develop analogs of the class of m.d. estimators $\{\widetilde{\theta}, G \text{ varies}\}$ of the previous section when $h(z)$ in (2.3) is unknown, but when validation data is available. Here assume that the validation sample $\{(\widetilde{X}_k, \widetilde{Z}_k), 1 \le k \le N\}$ obeys the equation (2.2) of the covariates and is independent of the primary data set $\{(Z_i, Y_i), 1 \le i \le n\}$. For instance, in the dietary intake example in Section 1, an independent validation data of dietary intake can be obtained by assessing urinary biomarkers in labs. Validation data are used to obtain an estimate $\widehat{h}(z)$ of $h(z)$. An analog of $\widetilde{\theta}$ is obtained by replacing $h(z)$ by $\widehat{h}(z)$ in its definition.

First, we define an estimator of the function $h(z)$ for $z \in \mathcal{C}$ where $\mathcal{C}$ is a compact set in $\mathbb{R}^p$ with $\inf_{\mathcal{C}} f_Z(z) \geq b_0$ for some $b_0 > 0$. Because $\widetilde{X}_k, \widetilde{Z}_k$ obey (2.2), we obtain a random sample $\{\widetilde{U}_k, 1 \leq k \leq N\}$ with $\widetilde{U}_k = \widetilde{Z}_k - \widetilde{X}_k$, of the measurement error U. Define the kernel density estimators of f_U and f_Z as follows. Let $K_1(\cdot)$, $K_2(\cdot)$ be the two symmetric density kernels, $w_1 \equiv w_1(N)$, $w_2 \equiv w_2(N)$ be bandwidth sequences depending on N and define

$$
\begin{aligned}
(3.1)\qquad \hat{f}_U(u) &:= \frac{1}{N w_1^p} \sum_{k=1}^N K_1\Big(\frac{\widetilde{U}_k - u}{w_1}\Big), \qquad \hat{f}_Z(z) := \frac{1}{N w_2^p} \sum_{k=1}^N K_2\Big(\frac{\widetilde{Z}_k - z}{w_2}\Big), \\
\tilde{h}(z) &:= \frac{1}{N} \sum_{k=1}^N \widetilde{X}_k \hat{f}_U(z - \widetilde{X}_k), \qquad h_0(z) := \int x f_U(z - x) dF_X(x).
\end{aligned}
$$

Let ϵ_0 be a known number satisfying $0 < \epsilon_0 < b_0$. Because,

$$
h(z) = E(X|Z = z) = \int x f_{X|Z}(x|z) dx = \frac{\int x f_U(z - x) dF_X(x)}{f_Z(z)} = \frac{h_0(z)}{f_Z(z)}, \quad z \in \mathcal{C},
$$

the function $h(z)$ can be estimated by

$$
(3.2)\qquad \widehat{h}(z) := \frac{N^{-1} \sum_{k=1}^N \widetilde{X}_k \hat{f}_U(z - \widetilde{X}_k)}{(\hat{f}_Z(z) \vee \epsilon_0)} = \frac{\tilde{h}(z)}{(\hat{f}_Z(z) \vee \epsilon_0)}, \quad z \in \mathcal{C}.
$$

Here ϵ_0 is introduced to avoid the vanishing denominator. For any two numbers $c \vee d := \max\{c, d\}$. The lemma below gives the asymptotic distribution of $\widehat{h}(z)$, which may be of independent interest.

Lemma 3.1. *Under model* (2.2)*, when f_Z and f_U are twicely continuously differentiable and $w_i \to 0$, $N w_i^p \to \infty$ as $N \to \infty$ for $i = 1, 2$, then*

$$
N^{1/2}\big(\widehat{h}(z) - h(z) - w_1^2 B(z)\big) \to_D \mathcal{N}_p(0, \Omega(z)), \ \forall\, z \in \mathbb{R}^p, \ f_Z(z) > 0,
$$

where $B(z) = \frac{1}{2} \int x f_X(x) u' f_U''(z - x) u K_1(u) du dx / f_Z(z)$, $f_U''(u)$ is the $p \times p$ matrix of second order partial derivatives of $f_U(u)$ and $\Omega(z) = [f_Z(z)]^{-2}\{\mathrm{Cov}(X f_U(z - X)) + \mathrm{Cov}((z - U) f_X(z - U))\}$.

Analogous to (2.6), we propose the estimator $\widehat{\theta}$ based on $\widehat{h}(z)$ as follows. Define $I_{\mathcal{C}}(z) = 1$, if $z \in \mathcal{C}$, otherwise 0. For $x \in \mathbb{R}$, $t \in \mathbb{R}^p$, let

$$
\begin{aligned}
(3.3)\qquad \widehat{V}(x, t) &:= n^{-1/2} \sum_{i=1}^n I_{\mathcal{C}}(Z_i) \widehat{h}(Z_i) \big[I(Y_i - t'\widehat{h}(Z_i) \leq x) - I(-Y_i + t'\widehat{h}(Z_i) < x)\big], \\
\widehat{M}(t) &:= \int \big\|\widehat{V}(x, t)\big\|^2 dG(x), \qquad \widehat{\theta} := \mathrm{argmin}_{t \in \mathbb{R}^p} \widehat{M}(t).
\end{aligned}
$$

To state the asymptotic normality of $\widehat{\theta}$, we need some more notation and the following additional assumptions. In this section, all limits are taken as $n \wedge N \to \infty$, unless mentioned otherwise. Let

$$
\begin{aligned}
(3.4)\qquad \widehat{\Gamma}_n(x) &:= n^{-1} \sum_{i=1}^n I_{\mathcal{C}}(Z_i) \widehat{h}(Z_i) \widehat{h}(Z_i)' \kappa_{Z_i}(x), \quad \widehat{\Gamma}(x) := E\big(I_{\mathcal{C}}(Z) h(Z) h(Z)' \kappa_Z(x)\big), \\
\widehat{\mathcal{G}}_n &:= \int \widehat{\Gamma}_n(x)' \widehat{\Gamma}_n(x) dG(x), \quad \widehat{\mathcal{G}} := \int \widehat{\Gamma}(x)' \widehat{\Gamma}(x) dG(x), \\
L(Z, \zeta) &:= \int I_{\mathcal{C}}(Z) \mu(z) h(Z) \varphi_z(\zeta) dQ(z), \qquad \boldsymbol{\alpha} := E \int I_{\mathcal{C}}(Z) \mu(z) h(Z) \psi_z(\zeta) dQ(z),
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\beta}(\theta) &:= E\Big\{I_{\mathcal{C}}(Z)\theta' B(Z)\int \mu(z)h(Z)\psi_z(\zeta)dQ(z)\Big\},\\
R_\theta(x) &:= E\Big\{I_{\mathcal{C}}(Z)\theta'\big[\widetilde{X} f_U(Z-\widetilde{X})/f_Z(Z) - h(Z)\big]\Big|\widetilde{X}=x\Big\},\\
S_\theta(u) &:= E\Big\{I_{\mathcal{C}}(Z)\theta'\big[(Z-\widetilde{U})f_X(Z-\widetilde{U})/f_Z(Z) - h(Z)\big]\big|\widetilde{U}=u\Big\}.
\end{aligned}
$$

We need the following additional assumptions, where ψ_z and μ are as in (2.27).

(3.5) $E\|X\|^4, E\|B(Z)\|^2, E\zeta^2$ and $E\big(\|\mu(Z)\psi_Z(\zeta)\|\big)$ are finite.

(3.6) The measure G satisfies (2.9) and G is either a distribution function or is absolutely continuous with a.e. derivative g bounded, i.e., $dG(x) = g(x)dx$, $\|g\|_\infty < \infty$.

(3.7) The density f of ε satisfies$\|f\|_\infty < \infty$.

(3.8) $\inf_{z\in\mathcal{C}} f_Z(z) \geq b_0 > 0, \quad \sup_{z\in\mathcal{C}} f_Z(z) < \infty$.

(3.9) f_U has an absolutely integrable characteristic function on $\mathbb{R}^p$ and $\|f_U\|_\infty < \infty$.

(3.10) The second partial derivative matrix $f''_Z(z)$ satisfies that $\sup\{\lambda_{\max}(f''_Z(z)); z \in \mathbb{R}^p\} < \infty$, where $\lambda_{\max}(f''_Z(z))$ is the maximum of the absolute eigenvalues of $f''_Z(z)$. Assume that the same holds for f''_U.

(3.11) The kernels K_1, K_2 are positive symmetric square integrable densities on $[-1,1]^p$. In addition, K_2 satisfies a Lipschitz condition.

(3.12) The bandwidths $w_i \to 0$, $\dfrac{Nw_i^p}{|\log w_i|} \to \infty$, $\dfrac{|\log w_i|}{\log\log N} \to \infty$, $w_i^p(N) \leq c\, w_i^p(2N)$, for some $c > 0, i = 1, 2$.

(3.13) $\widehat{\mathcal{G}}$ is positive definite and $\displaystyle\int \|\widehat{\Gamma}_n(x) - \widehat{\Gamma}(x)\|^2 dG(x) \to_p 0$.

(3.14) The assumption (2.17) holds with $h(Z)$ replaced by $\widehat{h}(Z)$.

(3.15) $\lim(n/N) = \lambda$, $0 \leq \lambda < \infty$. Moreover, $nw_1^4 \to C_1 < \infty$.

(3.16) $\lim(n/N) = \lambda = \infty$, $Nw_1^4 \to C_2 < \infty$.

The limiting distribution of $\widehat{\theta}$ is affected by the range of values of $\lambda = \lim(n/N)$ as is described in the following two theorems.

Theorem 3.1. $(0 \leq \lambda < \infty)$. *Under models* (2.1) *and* (2.2)*, when the assumptions* (2.9)–(2.14) *and* (3.5)–(3.15) *hold,* $n^{1/2}\big(\widehat{\theta} - \theta + w_1^2\widehat{\mathcal{G}}^{-1}\boldsymbol{\beta}(\theta)\big) \to_D \mathcal{N}_p\big(0, 4^{-1}\widehat{\mathcal{G}}^{-1}(\Sigma_0 + 4\lambda\Sigma_\theta)\widehat{\mathcal{G}}^{-1}\big)$*, where* $\Sigma_0 = \mathrm{Cov}(L(Z,\zeta))$ *and* $\Sigma_\theta = [\mathrm{Var}(R_\theta(\widetilde{X})) + \mathrm{Var}(S_\theta(\widetilde{U}))]\boldsymbol{\alpha}\boldsymbol{\alpha}'$*.*

Theorem 3.2. $(\lambda = \infty)$. *Under models* (2.1) *and* (2.2)*, when the assumptions* (2.9)–(2.14), (3.5)–(3.14) *and* (3.16) *hold, then we have* $N^{1/2}\big(\widehat{\theta} - \theta + w_1^2\widehat{\mathcal{G}}^{-1}\boldsymbol{\beta}(\theta)\big) \to_D \mathcal{N}_p\big(0, \widehat{\mathcal{G}}^{-1}\Sigma_\theta\widehat{\mathcal{G}}^{-1}\big)$.

For all $0 \leq \lambda \leq \infty$, the asymptotic bias in $\widehat{\theta}$ is inherited from the asymptotic bias in the estimator $\widehat{h}(z)$, see Lemma 3.1. If $\lambda = 0$, then N is relatively much larger than n and the asymptotic covariance matrix in the limiting distribution of Theorem 3.1 becomes $4^{-1}\widehat{\mathcal{G}}^{-1}\Sigma_0\widehat{\mathcal{G}}^{-1}$. This covariance matrix is like the one when $h(z)$ is known, except here we have the set $\mathcal{C}$ appearing in this matrix. On the other extreme is the case $\lambda = \infty$. In this case, the validation sample size N is much limited compared to the primary sample size n. In practice, this case arises more often due to the high cost of validation studies. It is not surprising to see that the convergence rate becomes $\sqrt{N}$ with the limiting covariance matrix $\widehat{\mathcal{G}}^{-1}\Sigma_\theta\widehat{\mathcal{G}}^{-1}$. The proofs of both the theorems appear in the Appendix.

4. Simulation study

In this section, we report findings of a finite sample simulation study comparing the empirical bias and RMSE of the two m.d. estimators corresponding to the integrating measure $G(x) \equiv x$ and $G(x) = \delta_0(x)$, with the calibrated least square and the bias corrected least square estimators. We conducted these simulations for the two cases: $h(z)$ known and $h(z)$ unknown but validation data available.

4.1. M.D. estimation when the function $h(z)$ is known

Here, data is generated from the regression model of Example 2.1, i.e., $p = 1$ and

$$Y_i = \theta X_i + \varepsilon_i,\ \theta = 2,\ Z_i = X_i + U_i,\ X_i \sim \mathcal{N}(a, \sigma_X^2),\ U_i \sim \mathcal{N}(0, \sigma_U^2),\ \varepsilon_i \sim F, \tag{4.1}$$
$$a = 1,\ \sigma_X^2 = 1,\ h(z) = z/(1+\sigma_U^2) + \sigma_U^2/(1+\sigma_U^2), \quad 1 \le i \le n.$$

Clearly, the regression function $h(z)$ is fully determined by σ_U^2. To assess the effect of measurement error on the estimators, in this simulation study we used $\sigma_U = 0.2, 0.5, 1, 1.5, 2$. We choose the sample size $n = 200, 500$ and F to be $\mathcal{N}(0,1)$, Laplace$(0,1)$ and t_2 d.f.'s.

Let $\widetilde{\theta}_{\ell ad}$ and $\widetilde{\theta}_I$ denote the m.d. estimators corresponding to $G(x) = \delta_0(x)$ and $G(x) \equiv x$, respectively. The formulas for the weighted empirical distances given in (2.7) and (2.8) were minimized, respectively, to obtain the numerical values of these estimators. The two other estimators we include in this study are

$$\widetilde{\theta}_{\ell se} = \sum h(Z_i)Y_i / \sum h^2(Z_i), \quad \widetilde{\theta}_{bc\ell s} = S_{YZ}/(S_{ZZ} - \sigma_U^2),$$

where $S_{YZ} = n^{-1}\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})$ and $S_{ZZ} = n^{-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$.

Tables 1– 3 give the empirical bias and RMSE based on 500 randomly simulated trials for the four estimators $(\widetilde{\theta}_I, \widetilde{\theta}_{\ell ad}, \widetilde{\theta}_{\ell se}, \widetilde{\theta}_{bc\ell s})$ in the case of known distributions of X and U. In addition, we present the boxplot of each estimator for the sample size $n = 500$ in Figure 1. From these tables, we see that there is little empirical bias in $\widetilde{\theta}_I$, $\widetilde{\theta}_{\ell ad}$ and $\widetilde{\theta}_{\ell se}$ for all chosen values of σ_U, n and all chosen F. The empirical bias of $\widetilde{\theta}_{bc\ell s}$ increases significantly, compared to the other three estimators, with the increasing σ_U. This is true for all chosen sample sizes and F. Regarding RMSE, for any fixed value of σ_U and n, $\widetilde{\theta}_I$ and $\widetilde{\theta}_{\ell se}$ achieve minimal RMSE among the four estimators for Gaussian and Laplace error distributions, as seen in Tables 1 and 2, while from Table 3 we see the superior performance of the m.d. estimators $\widetilde{\theta}_I$ and $\widetilde{\theta}_{\ell ad}$ for t_2 error distribution. The estimator $\widetilde{\theta}_{bc\ell s}$ displays much larger RMSE for the larger values of $\sigma_U = 1.5$ and 2 and for all chosen F, which is also indicated in Figure 1. Not unexpectedly, the RMSE of each estimator increases with increasing σ_U. Moreover, from the right panel of Figure 1, we see that for $F = t_2$, $\widetilde{\theta}_{\ell se}$ and $\widetilde{\theta}_{bc\ell s}$ display not only larger variation but also unstable estimation presented by outliers. The bias and RMSE of $\widetilde{\theta}_I$ is especially robust against the larger σ_U.

4.2. M.D. estimation when $h(z)$ is unknown but validation data is available

We continue generating data from the model (4.1) in Section 4.1. We simulated both the primary sample $\{Y_i, Z_i, 1 \le i \le n\}$ and validation sample $\{\widetilde{X}_k, \widetilde{Z}_k, 1 \le k \le N\}$ following (4.1) with the two samples size choices $(n, N) = (200, 200)$ and $(500, 500)$. The function $h(z)$ is estimated by $\widehat{h}(z)$ in (3.2) based on the validation sample. Let $(\widehat{\theta}_I, \widehat{\theta}_{\ell ad}, \widehat{\theta}_{\ell se}, \widehat{\theta}_{bc\ell s})$ denote the analog of H-L, LAD, LS and BCLS estimators when $h(z)$ and σ_U^2 are replaced by $\widehat{h}(z)$ and $\hat{\sigma}_U^2$. Here $\hat{\sigma}_U^2$ is the sample variance of $\{\widetilde{U}_k := \widetilde{Z}_k - \widetilde{X}_k, 1 \le k \le N\}$. The kernels are chosen as $K_1(x) = K_2(x) = 0.75(1-x^2)I(|x| \le 1)$. To meet the assumptions, we specify the choices of the bandwidths w_1 and w_2 as follows. Similar to the rule-of-thumb bandwidth in kernel density estimators of Silverman

(1986), we set $w_1 = \hat{\sigma}_U^2 N^{-1/(2+p)}$. We adapt the univariate plug-in selector of Wand and Jones (1994) for w_2. We chose $\mathcal{C}$ as the interval between the 5th and 95th percentiles of $\{Z_i, 1 \le i \le n\}$ and $\epsilon_0 = 10^{-4}$.

The empirical bias and RMSE summary is presented in Tables 4–6 for $\sigma_U = 0.2, 0.5, 1, 1.5, 2$ and for Gaussian, Laplace and t_2 regression errors, respectively. Boxplots of the four estimators for sample sizes $(n, N) = (500, 500)$ are shown in Figure 2. From these tables we see that for fixed n, N, σ_U, F, the empirical bias and RMSE of all four estimators are larger than those in Tables 1–3 due to the estimation error in $\widehat{h}(z)$ and $\widehat{\sigma}_U$. The bias and RMSE of the two m.d. estimators $\widehat{\theta}_I$ and $\widehat{\theta}_{\ell ad}$ increase only slightly with the increasing σ_U, while bias and RMSE of $\widehat{\theta}_{bc\ell s}$ increases significantly with increasing σ_U. Moreover, from Figure 2 we see that the estimators $\widehat{\theta}_{\ell se}$ and $\widehat{\theta}_{bc\ell s}$ display unstable estimation performance for the larger σ_U and all chosen F. Overall, the m.d. estimator $\widehat{\theta}_I$ achieves the smallest RMSE with controlled bias among the four estimators.

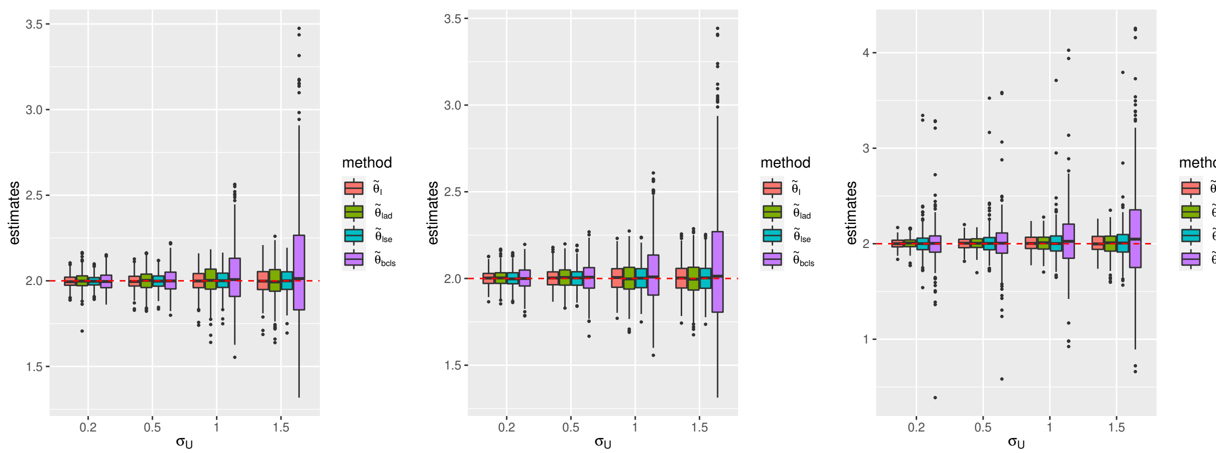

Figure 1: Boxplot of each estimator for known $h(z)$ and $n = 500$ with different values of σ_U under Gaussian errors (left panel), Laplace errors (central panel) and t_2 errors (right panel).

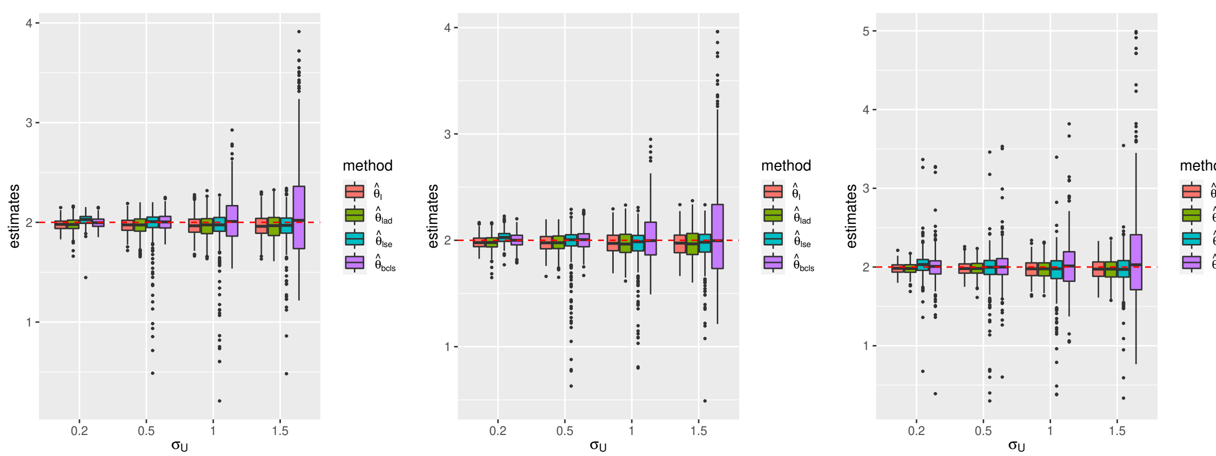

Figure 2: Boxplot of each estimator for unknown $h(z)$ and $(n, N) = (500, 500)$ with different values of σ_U under Gaussian errors (left panel), Laplace errors(central panel) and t_2 errors (right panel).

$\varepsilon \sim \mathcal{N}(0,1)$	$n=200$				$n=500$			
	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	0.0021	0.0102	0.0022	0.0049	-0.0029	0.0013	-0.0022	-0.0026
RMSE	0.0533	0.0791	0.0528	0.0788	0.0355	0.0525	0.0346	0.0509
$\sigma_U = 0.5$								
BIAS	0.0017	0.0138	0.0016	0.0107	-0.0026	0.0018	-0.0017	0.0018
RMSE	0.0690	0.0968	0.0673	0.1218	0.0466	0.0676	0.0452	0.0757
$\sigma_U = 1$								
BIAS	0.0030	0.0160	0.0019	0.0477	-0.0022	0.0073	-0.0012	0.0301
RMSE	0.0953	0.1337	0.0912	0.2930	0.0689	0.0932	0.0664	0.1800
$\sigma_U = 1.5$								
BIAS	0.0046	-0.0056	0.0033	0.2054	$< 10^{-4}$	-0.0034	0.0002	0.0841
RMSE	0.1132	0.1479	0.1092	0.8464	0.0792	0.0992	0.0777	0.3829
$\sigma_U = 2$								
BIAS	0.0046	0.0129	0.0035	0.4586	0.0008	-0.0008	0.0010	0.2368
RMSE	0.1237	0.1661	0.1194	1.8504	0.0879	0.1156	0.0860	0.8585

Table 1: Bias and RMSE of each estimator for known $h(z)$ under Gaussian regression error.

$\varepsilon \sim \text{Laplace}(0,1)$	$n=200$				$n=500$			
	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	0.0014	-0.0009	0.0039	-0.0019	0.0011	0.0040	0.0010	0.0012
RMSE	0.0658	0.0979	0.0737	0.1040	0.0418	0.0511	0.0473	0.0690
$\sigma_U = 0.5$								
BIAS	-0.0019	0.0021	-0.0025	0.0070	0.0016	0.0086	0.0012	0.0052
RMSE	0.0860	0.1131	0.0902	0.1452	0.0543	0.0722	0.0568	0.0929
$\sigma_U = 1$								
BIAS	0.0015	0.0063	-0.0011	0.0520	0.0020	0.0039	0.0019	0.0265
RMSE	0.1159	0.1554	0.1179	0.3224	0.0765	0.1001	0.0773	0.1878
$\sigma_U = 1.5$								
BIAS	0.0039	-0.0086	0.0007	0.2484	0.0034	-0.0003	0.0033	0.0824
RMSE	0.1346	0.1684	0.1362	1.0327	0.0854	0.1035	0.0855	0.4044
$\sigma_U = 2$								
BIAS	0.0070	0.0070	0.0072	0.3120	0.0037	0.0046	0.0033	0.2308
RMSE	0.1459	0.1902	0.1438	1.3663	0.0964	0.1157	0.0963	0.9375

Table 2: Bias and RMSE of each estimator for known $h(z)$ under Laplace regression error.

$\varepsilon \sim t_2$	$n=200$				$n=500$			
	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$	$\widetilde{\theta}_I$	$\widetilde{\theta}_{\ell ad}$	$\widetilde{\theta}_{\ell se}$	$\widetilde{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	-0.0024	0.0006	0.0101	0.0136	0.0020	0.0055	0.0052	0.0036
RMSE	0.0774	0.1032	0.2925	0.3371	0.0476	0.0717	0.1300	0.1912
$\sigma_U = 0.5$								
BIAS	-0.0030	-0.0019	0.0084	0.0189	0.0024	0.0068	0.0061	0.0088
RMSE	0.0966	0.1291	0.2861	0.3386	0.0607	0.0812	0.1378	0.2159
$\sigma_U = 1$								
BIAS	-0.0025	-0.0084	0.0077	0.0600	0.0035	0.0118	0.0071	0.0306
RMSE	0.1256	0.1777	0.2915	0.4611	0.0821	0.1094	0.1525	0.3032
$\sigma_U = 1.5$								
BIAS	-0.0009	-0.0202	0.0083	0.2199	0.0041	0.0032	0.0079	0.0921
RMSE	0.1450	0.1825	0.3028	0.9348	0.0971	0.1190	0.1651	0.4947
$\sigma_U = 2$								
BIAS	0.0001	-0.0061	0.0090	0.5872	0.0047	0.0044	0.0087	0.2986
RMSE	0.1571	0.2074	0.3129	2.0090	0.1058	0.1261	0.1737	1.2157

Table 3: Bias and RMSE of each estimator for known $h(z)$ under t_2 regression error.

$\varepsilon \sim \mathcal{N}(0,1)$	$n=200, N=200$				$n=500, N=500$			
	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	-0.0380	-0.0025	0.0033	0.0057	-0.0241	0.0173	0.0236	-0.0027
RMSE	0.0884	0.0989	0.0811	0.0790	0.0589	0.0650	0.0592	0.0513
$\sigma_U = 0.5$								
BIAS	-0.0439	-0.0082	-0.0485	0.0139	-0.0291	0.0065	-0.0231	0.0014
RMSE	0.1167	0.1468	0.2412	0.1309	0.0821	0.0951	0.1714	0.0834
$\sigma_U = 1$								
BIAS	-0.0551	-0.0535	-0.0960	0.0775	-0.0335	-0.0070	-0.0567	0.0288
RMSE	0.1603	0.2020	0.3048	0.3764	0.1085	0.1281	0.2246	0.2267
$\sigma_U = 1.5$								
BIAS	-0.0593	-0.0612	-0.0833	0.4932	-0.0371	-0.0378	-0.0480	0.1371
RMSE	0.1843	0.2075	0.2699	2.0647	0.1231	0.1390	0.1830	0.6050
$\sigma_U = 2$								
BIAS	-0.0629	-0.0567	-0.0655	0.7770	-0.0412	-0.0421	-0.0390	0.4812
RMSE	0.1977	0.2315	0.2252	8.1122	0.1302	0.1440	0.1487	3.1191

Table 4: Bias and RMSE of each estimator for unknown $h(z)$ with validation data under Gaussian errors.

$\varepsilon \sim$ Laplace$(0,1)$	$n=200, N=200$				$n=500, N=500$			
	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	-0.0384	-0.0134	0.0038	0.0011	-0.0199	0.0175	0.0269	0.0009
RMSE	0.0991	0.1057	0.1146	0.1051	0.0620	0.0669	0.06418	0.0693
$\sigma_U = 0.5$								
BIAS	-0.0447	-0.0113	-0.0411	0.0131	-0.0216	0.0131	-0.0265	0.0039
RMSE	0.1355	0.1591	0.2362	0.1565	0.0838	0.0975	0.1809	0.0980
$\sigma_U = 1$								
BIAS	-0.0483	-0.0474	-0.0968	0.1015	-0.0267	-0.0062	-0.0479	0.0249
RMSE	0.1713	0.2194	0.3460	0.4507	0.1079	0.1253	0.1863	0.2333
$\sigma_U = 1.5$								
BIAS	-0.0498	-0.0534	-0.0808	0.4352	-0.0304	-0.0334	-0.0415	0.1166
RMSE	0.1917	0.2126	0.2947	2.8417	0.1224	0.1430	0.1733	0.5827
$\sigma_U = 2$								
BIAS	-0.0545	-0.0470	-0.0628	-0.2593	-0.0329	-0.0198	-0.0304	0.3198
RMSE	0.2036	0.2326	0.2567	9.8919	0.1290	0.1449	0.1501	4.5430

Table 5: Bias and RMSE of each estimator for unknown $h(z)$ with validation data under Laplace errors.

$\varepsilon \sim t_2$	$n=200, N=200$				$n=500, N=500$			
	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$	$\widehat{\theta}_I$	$\widehat{\theta}_{\ell ad}$	$\widehat{\theta}_{\ell se}$	$\widehat{\theta}_{bc\ell s}$
$\sigma_U = 0.2$								
BIAS	-0.0452	-0.0516	0.0110	0.0131	-0.0171	-0.0173	0.0294	0.0035
RMSE	0.1138	0.1283	0.2842	0.3367	0.0691	0.0759	0.1568	0.1909
$\sigma_U = 0.5$								
BIAS	-0.0488	-0.0536	-0.0434	0.0167	-0.0201	-0.0167	-0.0221	0.0083
RMSE	0.1469	0.1657	0.3661	0.3400	0.0906	0.0996	0.2394	0.2153
$\sigma_U = 1$								
BIAS	-0.0600	-0.0678	-0.0905	0.0690	-0.0266	-0.0270	-0.0493	0.0350
RMSE	0.1834	0.2076	0.3814	0.5206	0.1197	0.1271	0.2498	0.3269
$\sigma_U = 1.5$								
BIAS	-0.0593	-0.0663	-0.0847	0.2926	-0.0245	-0.0243	-0.0530	0.1398
RMSE	0.2038	0.2272	0.3469	2.8562	0.1094	0.1179	0.2644	0.6613
$\sigma_U = 2$								
BIAS	-0.0620	-0.0694	-0.0983	0.6243	-0.0305	-0.0294	-0.0335	0.6066
RMSE	0.1983	0.2789	0.3219	7.4859	0.1369	0.1462	0.2225	3.6414

Table 6: Bias and RMSE of each estimator for unknown $h(z)$ with validation data under t_2 errors.

Appendix

In this section some proofs are presented.

Proof of (2.25). Let $D(s) := \int_0^\infty \|V(x,\theta+n^{-1/2}s)\|^2 dG(x)$. By (2.23), $M(\theta+n^{-1/2}s) = 2D(s)$. Recall that $c_{ni} := n^{-1/2}h(Z_i)$ and define

$$U(x,s) := n^{-1/2}\sum_{i=1}^n h(Z_i)I\big(\zeta_i \le x + s'c_{ni}\big), \quad J(x,s) := n^{-1/2}\sum_{i=1}^n h(Z_i)K_{Z_i}(x+s'c_{ni}),$$
$$W(x,s) := U(x,s) - J(x,s), \quad x \in \mathbb{R}^p, s \in \mathbb{R}^p.$$

Note that $EU(x,s) \equiv EJ(x,s)$, $EW(x,s) \equiv 0$. By (2.5),

$$n^{-1/2}\sum_{i=1}^n h(Z_i)\big\{K_{Z_i}(x) + K_{Z_i}(-x)\big\} = n^{-1/2}\sum_{i=1}^n h(Z_i), \quad \forall\, x \in \mathbb{R}^p.$$

Recall the definition of $\Gamma_n(x), \Gamma(x), \mathcal{G}_n$ and $\mathcal{G}$ from (2.15). By (2.5), $\Gamma_n(x) \equiv \Gamma_n(-x)$, $\Gamma(x) \equiv \Gamma(-x)$. Use the above notation and facts to rewrite

$$\begin{aligned} D(s) &= \int_0^\infty \Big\|U(x,s) + U(-x,s) - n^{-1/2}\sum_{i=1}^n h(Z_i)\Big\|^2 dG(x) \\ &= \int_0^\infty \Big\|\big\{W(x,s) - W(x,0)\big\} + \big\{W(-x,s) - W(-x,0)\big\} \\ &\qquad + \big\{J(x,s) - J(x,0) - \Gamma_n(x)s\big\} + \big\{J(-x,s) - J(-x,0) - \Gamma_n(-x)s\big\} \\ &\qquad + \big\{W(x,0) + W(-x,0) + 2\Gamma_n(x)s\big\}\Big\|^2 dG(x). \end{aligned}$$

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

$$D(s) = D_1(s) + D_2(s) + \cdots + D_5(s) + 2 \times (10 \text{ cross product terms}), \tag{4.2}$$

where

$$\begin{aligned} D_1(s) &:= \int_0^\infty \big\|W(x,s) - W(x,0)\big\|^2 dG(x), \\ D_2(s) &:= \int_0^\infty \big\|W(-x,s) - W(-x,0)\big\|^2 dG(x), \\ D_3(s) &:= \int_0^\infty \big\|J(x,s) - J(x,0) - \Gamma_n(x)s\big\|^2 dG(x), \\ D_4(s) &:= \int_0^\infty \big\|J(-x,s) - J(-x,0) - \Gamma_n(-x)s\big\|^2 dG(x), \\ D_5(s) &:= \int_0^\infty \big\|W(x,0) + W(-x,0) + 2\Gamma_n(x)s\big\|^2 dG(x). \end{aligned}$$

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

Lemma 4.1. *Under the above setup and the assumptions* (2.9)–(2.17), $\forall\, 0 < b < \infty$,

$$\sup_{s\in\mathcal{U}(b)} D_k(s) = o_p(1), \quad k = 1, 2, \cdots, 4, \tag{4.3}$$

$$E\Big(\sup_{s\in\mathcal{U}(b)} D_5(s)\Big) = O(1). \tag{4.4}$$

Unless mentioned otherwise, all the supremum below are taken over $s \in \mathcal{U}(b)$. Lemma 4.1 together with the C–S inequality yield that

$$\sup_s \Big| \int_0^\infty \big\{W(x,s) - W(x,0)\big\}\big\{W(x,0) + W(-x,0) + 2\Gamma_n(x)s\big\} dG(x) \Big|^2 \\ \le \sup_s D_1(s) \sup_s D_5(s) = o_p(1),$$

by (4.3) applied with $k = 1$ and (4.4). Similarly, the supremum over s of all other cross product terms in the right hand side of (4.2) tend to zero, in probability.

Consequently, because $D(s) = M(\theta + n^{-1/2}s)/2$, we obtain

$$\sup_s \big| M(\theta + n^{-1/2}s) - 2D_5(s) \big| \to_p 0. \tag{4.5}$$

Let

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

Because of (2.5) and (2.9), $\int_0^\infty \Gamma_n(x)\Gamma_n(x)' dG(x) = \mathcal{G}_n/2$, $\int_0^\infty \Gamma(x)\Gamma(x)' dG(x) = \mathcal{G}/2$. Use these facts when expanding the quadratic in D_5 to write

$$D_5(s) = 2^{-1}M(\theta) + 4s'T_n^* + 2s'\mathcal{G}_n s. \tag{4.6}$$

Because $T_n = 2\int_0^\infty \Gamma(x)\big\{W(x,0) + W(-x,0)\big\} dG(x)$, by the C-S inequality, (2.16) and (2.22),

$$\begin{aligned} \big\|T_n^* - \frac{1}{2}T_n\big\|^2 &= \Big\| \int_0^\infty \big\{\Gamma_n(x) - \Gamma(x)\big\}\big\{W(x,0) + W(-x,0)\big\} dG(x) \Big\|^2 \\ &\le D(0) \int_0^\infty \big\|\Gamma_n(x) - \Gamma(x)\big\|^2 dG(x) \to_p 0, \\ \|\mathcal{G}_n - \mathcal{G}\| &\to_p 0. \end{aligned}$$

Hence (2.25) follows from (4.5) and (4.6), thereby completing the proof of Theorem 2.1. □

Proof of Lemma 4.1. Consider D_3. Let $D_{is}(x) := K_{Z_i}(x + s'c_{ni}) - K_{Z_i}(x) - s'c_{ni}\kappa_{Z_i}(x)$. Then

$$\begin{aligned} D_3(s) &:= \int_0^\infty \Big\| n^{-1/2} \sum_{i=1}^n h(Z_i) D_{is}(x) \Big\|^2 dG(x) \\ &\le \frac{1}{n} \sum_{i=1}^n \big\|h(Z_i)\big\|^2 \int_0^\infty \sum_{i=1}^n D_{is}^2(x) dG(x). \end{aligned} \tag{4.7}$$

By the LLN's and (2.4),

$$\sum_{i=1}^n \|c_{ni}\|^2 = n^{-1} \sum_{i=1}^n \|h(Z_i)\|^2 \to_p E\|h(Z)\|^2 < \infty. \tag{4.8}$$

By the C–S inequality and (2.14), for all $s \in \mathcal{U}(b)$,

$$\begin{aligned} \int_0^\infty \sum_{i=1}^n D_{is}^2(x) dG(x) &\le \int_0^\infty \sum_{i=1}^n \Big(\int_{-|s'c_{ni}|}^{|s'c_{ni}|} \big(\kappa_{Z_i}(x+u) - \kappa_{Z_i}(x)\big) du \Big)^2 dG(x) \\ &\le 2bn^{-1} \int_0^\infty \sum_{i=1}^n \|h(Z_i)\|^2 \int_{-bn^{-1/2}}^{bn^{-1/2}} \big(\kappa_{Z_i}(x + u\|h(Z_i)\|) - \kappa_{Z_i}(x)\big)^2 du dG(x). \end{aligned}$$

Therefore, by the Fubini Theorem and (2.14),

$$(4.9)\qquad E\Big(\sup_s \int_0^\infty \sum_{i=1}^n D_{is}^2(x)dG(x)\Big)$$
$$\le 2bn^{1/2}\int_{-bn^{-1/2}}^{bn^{-1/2}}\int E\big\{\|h(Z)\|^2\big(\kappa_Z(x+v\|h(Z)\|)-\kappa_Z(x)\big)^2\big\}dG(x)dv \to 0.$$

Upon combining this fact with (4.8) and (4.7), we obtain $\sup_s D_3(s) = o_p(1)$, thereby proving (4.3) for $j=3$. The proof for $j=4$ is exactly similar.

Now consider D_1. Because the ith summand in $W(x,s)-W(x,0)$ is a conditionally centered Bernoulli type r.v. and the summands are i.i.d., by the Fubini Theorem and (2.14),

$$(4.10)\qquad \begin{aligned} ED_1(s) &\le \int_{-\infty}^\infty E\big\|W(x,s)-W(x,0)\big\|^2 dG(x) \\ &\le E\Big(\|h(Z)\|^2\int_{-\infty}^\infty \big|K_Z(x+n^{-1/2}s'h(Z))-K_Z(x)\big|dG(x)\Big) \\ &\le \int_{-bn^{-1/2}}^{bn^{-1/2}}\int E\Big(\|h(Z)\|^3\kappa_Z\big(x+u\|h(Z)\|\big)\Big)dG(x)du \to 0. \end{aligned}$$

In view of (4.10), to prove (4.3) for $j=1$, 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 $s\in\mathcal{U}(b)$,

$$(4.11)\qquad P\Big(\sup_{\|t-s\|<\delta}\big|D_1(t)-D_1(s)\big|>\epsilon\Big)<\epsilon, \quad \forall\, n>N_\epsilon.$$

Let $h_j(z)$ denote the jth coordinate of $h(z)$, $j=1,\cdots,p$ and let $\alpha_i(x,t) := I(\zeta_i\le x+t'c_{ni}) - I(\zeta_i\le x) - K_{Z_i}(x+t'c_{ni}) + K_{Z_i}(x)$. Then

$$\begin{aligned} D_1(s) &= \int_0^\infty \big\|W(x,s)-W(x,0)\big\|^2 dG(x) \\ &= \sum_{j=1}^p\int_0^\infty\Big(n^{-1/2}\sum_{i=1}^n h_j(Z_i)\alpha_i(x,s)\Big)^2 dG(x) = \sum_{j=1}^p D_{1j}(s), \quad \text{say.} \end{aligned}$$

Thus it suffices to prove (4.11) with D_1 replaced by D_{1j} for each $1\le j\le p$.

Fix a $1\le j\le p$ and write $h_j(Z_i) = h_j^+(Z_i) - h_j^-(Z_i)$, where $h_j^+ = \max(0,h_j)$, $h_j^- = \max(0,-h_j)$. Let

$$W_j^\pm(x,s) := n^{-1/2}\sum_{i=1}^n h_j^\pm(Z_i)\alpha_i(x,s), \quad \mathcal{D}_j^\pm(x,s,t) := W_j^\pm(x,t)-W_j^\pm(x,s),$$
$$R_j^\pm(s,t) := \int_0^\infty \big(\mathcal{D}_j^\pm(x,s,t)\big)^2 dG(x), \quad \alpha_i(x,s,t) := \alpha_i(x,t)-\alpha_i(x,s).$$

Then

$$(4.12)\qquad \begin{aligned} &\big|D_{1j}(t)-D_{1j}(s)\big| \\ &= \int_0^\infty\Big(n^{-1/2}\sum_{i=1}^n\big[h_j^+(Z_i)-h_j^-(Z_i)\big]\alpha_i(x,s,t)\Big)^2 dG(x) \\ &\le \int_0^\infty\big(\mathcal{D}_j^+(x,s,t)\big)^2 dG(x) + \int_0^\infty\big(\mathcal{D}_j^-(x,s,t)\big)^2 dG(x) \\ &\quad +2\Big\{\int_0^\infty\big(\mathcal{D}_j^+(x,s,t)\big)^2 dG(x)\int_0^\infty\big(\mathcal{D}_j^-(x,s,t)\big)^2 dG(x)\Big\}^{1/2} \end{aligned}$$

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

Note that

$$\mathcal{D}_j^+(x,s,t) \quad = \quad n^{-1/2}\sum_{i=1}^n h_j^+(Z_i)\big[\alpha_i(x,t) - \alpha_i(x,s)\big].$$

Fix $s \in \mathcal{U}_b$, $\epsilon > 0$ and $\delta > 0$. Then, $\forall\, t \in \mathcal{U}(b)$, $\|t-s\| < \delta$ implies $s'c_{ni} - \delta\|c_{ni}\| \le t'c_{ni} \le s'c_{ni} + \delta\|c_{ni}\|$, for all i. By the nondecreasing property of the indicator function and the d.f.,

$$\begin{aligned}
& I(\zeta_i \le x + s'c_{ni} - \delta\|c_{ni}\|) - I(\zeta_i \le x + s'c_{ni}) - K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) \\
&\qquad + K_{Z_i}(x + s'c_{ni}) + K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) - K_{Z_i}(x + s'c_{ni} - \delta\|c_{ni}\|) \\
& \le \alpha_i(x,t) - \alpha_i(x,s) = I(\zeta_i \le x + t'c_{ni}) - I(\zeta_i \le x + s'c_{ni}) \\
&\qquad\qquad\qquad\qquad - K_{Z_i}(x + t'c_{ni}) + K_{Z_i}(x + s'c_{ni}) \\
& \le I(\zeta_i \le x + s'c_{ni} + \delta\|c_{ni}\|) - I(\zeta_i \le x + s'c_{ni}) - K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) \\
&\qquad + K_{Z_i}(x + s'c_{ni}) + K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) - K_{Z_i}(x + s'c_{ni} - \delta\|c_{ni}\|).
\end{aligned}$$

Let, for any $a \in \mathbb{R}, x \in \mathbb{R}$,

$$\begin{aligned}
\mathcal{D}_{j1}^{\pm}(x,s,a) \quad := \quad & n^{-1/2}\sum_{i=1}^n h_j^{\pm}(Z_i)\big\{I(\zeta_i \le x + s'c_{ni} + a\|c_{ni}\|) - I(\zeta_i \le x + s'c_{ni}) \\
& \qquad - K_{Z_i}(x + s'c_{ni} + a\|c_{ni}\|) + K_{Z_i}(x + s'c_{ni})\big\}.
\end{aligned}$$

Using the above inequalities and $h_j^+(Z_i)$ being nonnegative we obtain that

$$\begin{aligned}
& \mathcal{R}_j^+(s,t) := \int_0^\infty \big(D_j^+(x,s,t)\big)^2 dG(x) \\
& \le \int_0^\infty \big(\mathcal{D}_{j1}^+(x,s,\delta)\big)^2 dG(x) + \int_0^\infty \big(\mathcal{D}_{j1}^+(x,s,-\delta)\big)^2 dG(x) \\
& + \int_0^\infty \Big(n^{-1/2}\sum_{i=1}^n h_j^+(Z_i)\big\{K_{Z_i}(x + s'c_{ni} + \delta\|c_{ni}\|) - K_{Z_i}(x + s'c_{ni} - \delta\|c_{ni}\|\big\} dG(x)\Big)^2.
\end{aligned}$$

Argue as for (4.3) when $j = 3$ to see that the first two terms in the above bound tend to 0, in probability. Let $\delta_n := n^{-1/2}\delta$ and $\Delta_j(s,\delta)$ denote the last term in the above bound. Argue as for (4.10) to obtain that $\Delta_j(s,\delta)$ is bounded from the above by

$$\begin{aligned}
& \int_0^\infty \Big(n^{-1/2}\sum_{i=1}^n h_j^+(Z_i)\|h(Z_i)\| \int_{-\delta_n}^{\delta_n} \kappa_{Z_i}(x + n^{-1/2}s'h(Z_i) + u\|h(Z_i)\|)du\, dG(x)\Big)^2 \\
& \le n^{-1}\int_0^\infty \Big(\sum_{i=1}^n \|h(Z_i)\|^2 \int_{-\delta_n}^{\delta_n} \kappa_{Z_i}(x + n^{-1/2}s'h(Z_i) + u\|h(Z_i)\|)du\, dG(x)\Big)^2 \\
& \le n^{-1}\int \sum_{i=1}^n \|h(Z_i)\|^4 \sum_{i=1}^n (2\delta_n) \int_{-\delta_n}^{\delta_n} \kappa_{Z_i}^2(x + n^{-1/2}s'h(Z_i) + u\|h(Z_i)\|)du\, dG(x).
\end{aligned}$$

Hence, by the Fubini Theorem and (2.14),

$$\begin{aligned}
E\Delta_j(s,\delta) \quad & \le \quad 2\delta n^{1/2}\int_{-n^{-1/2}\delta}^{n^{-1/2}\delta}\int E\Big(\|h(Z)\|^4\kappa_Z^2\big(x + n^{-1/2}s'h(Z) + u\|h(Z)\|\big)\Big)dG(x)\, du \\
& \to \quad 4\delta^2 \int E\big(\|h(Z)\|^4\kappa_Z^2(x)\big)dG(x), \quad \text{as } n \to \infty, \forall\, 1 \le j \le p.
\end{aligned}$$

Since the factor multiplying δ^2 is positive, the above term can be made smaller than ϵ by the choice of δ. This then completes the proof of R_j^+ satisfying (4.11). The details of the proof for verifying (4.11) for R_j^- are exactly similar. These facts together with the upper bound of (4.12) show that (4.11) is satisfied by D_{1j} for each $j = 1, \cdots, p$ thereby completing the proof of (4.3) for D_1. The proof of (4.3) for D_2 is similar. This completes the proof of (4.3).

Next, consider (4.4). By the C-S inequality, $\|\Gamma_n(x)\|^2 \le \frac{1}{n}\sum_{i=1}^n \|h(Z_i)\|^4 \times \kappa_{Z_i}^2(x)$, for all $x \in \mathbb{R}$. Recall $D(0) = M(\theta)/2$. Hence, by (2.13) and (2.22),

$$\begin{aligned} E\big(\sup_s D_5(s)\big) &\le 2ED(0) + 2b\int E\|\Gamma_n(x)\|^2 dG(x) \\ &\le 2ED(0) + 2b\int E\Big(\|h(Z)\|^4\kappa_Z^2(x)\Big)dG(x) < \infty. \qquad \square \end{aligned}$$

Proof of Lemma 3.1. It is well known that the kernel density estimator $\hat f_Z(z) \to_p f_Z(z)$ under (3.12) for all $z \in \mathcal{C}$. Because $\epsilon_0 < b_0 \le \inf_{z\in\mathcal{C}} f_Z(z)$, we have that $(\hat f_Z(z) \vee \epsilon_0) \to_p f_Z(z)$, for all $z \in \mathcal{C}$. Slutsky's theorem implies that it suffices to show that $\tilde h(z)$ is asymptotically normally distributed. Rewrite

$$\tilde h(z) = \frac{1}{N^2 w_1^p}\sum_{k=1}^N\sum_{l=1}^N \widetilde X_k K_1\Big(\frac{\widetilde U_l + \widetilde X_k - z}{w_1}\Big) := N^{-2}\sum_{k=1}^N\sum_{l=1}^N \phi(z, \widetilde X_k, \widetilde U_l),$$

where $\phi(z, \widetilde X_k, \widetilde U_l) = \widetilde X_k K_w(\widetilde U_l + \widetilde X_k - z)$ and $K_w(\cdot) = K(\cdot/w_1)/w_1^p$. We see that $\tilde h(z)$ is in fact a two-sample U-statistic based on the function ϕ. Recall $h_0(z)$ in (3.1). The conditional expectations can be calculated as follows.

$$\begin{aligned} &E\phi(\widetilde X_1, \widetilde U_1) = h_0(z) + w_1^2\widetilde B(z) + o(w_1^2), \\ &\widetilde B(z) := \frac{1}{2}\int\int x f_X(x) u' f_U''(z-x) u K_1(u)\,du\,dx, \\ &E(\phi|\widetilde X_1) = \widetilde X_1 f_U(z - \widetilde X_1) + O(w_1^2),\ E(\phi|\widetilde U_1) = (z - \widetilde U_1) f_X(z - \widetilde U_1) + O(w_1^2). \end{aligned}$$

Similar to the derivation of Lemma 6.4 in Geng and Koul (2017), we obtain that

$$\mathrm{Cov}(\tilde h(z)) = O\Big(\frac{1}{N^2 w_1^p} + N^{-1}\mathrm{Cov}\big(E(\phi|\widetilde X_1)\big) + N^{-1}\mathrm{Cov}(E\big(\phi|\widetilde U_1)\big)\Big).$$

Under (3.12), we see that the asymptotic covariance of $\tilde h(z)$ is of the order N^{-1}, because of $N w_1^p \to \infty$ implied by (3.12). Furthermore, Theorem 6.1.4 in Lehmann (1999) and $w_1 \to 0$ imply that

$$N^{1/2}(\tilde h(z) - h_0(z) - w_1^2\widetilde B(z)) \to_D \mathcal{N}_p\big(0, \widetilde\Omega\big),\ \ \widetilde\Omega := \mathrm{Cov}(X f_U(z - X)) + \mathrm{Cov}((z-U) f_X(z-U)).$$

This completes the proof of Lemma 3.1. $\square$

To study the asymptotic distribution of $\widehat\theta$, we need the following preliminary result about the uniform consistency of $\widehat h(z)$ over $\mathcal{C}$.

$$\sup_{z\in\mathcal{C}} \|\widehat h(z) - h(z)\| = o_p(1). \tag{4.13}$$

Proof of (4.13). Recall the definitions of $\widehat h$ and h_0 from (3.1). Rewrite

$$\widehat h(z) - h(z) = \frac{1}{(\hat f_Z(z) \vee \epsilon_0)}(\tilde h(z) - h_0(z)) + h_0(z)\Big(\frac{1}{(\hat f_Z(z) \vee \epsilon_0)} - \frac{1}{f_Z(z)}\Big).$$

By (3.5) and (3.9), $\sup_z \|h_0(z)\| < \infty$. Moreover, $(\hat{f}_Z(z) \vee \epsilon_0)^{-1} \leq \epsilon_0^{-1}$. Therefore, it suffices to show that

$$(4.14) \qquad \text{(a)} \sup_{z\in\mathcal{C}} \|\tilde{h}(z) - h_0(z)\| = o_p(1); \qquad \text{(b)} \sup_{z\in\mathcal{C}} \Big|\frac{1}{(\hat{f}_Z(z) \vee \epsilon_0)} - \frac{1}{f_Z(z)}\Big| = o_p(1).$$

First, we prove (4.14)(a). Let

$$F_N(x) := N^{-1} \sum_{k=1}^{N} I(\widetilde{X}_k \leq x), \quad h_1(z) := N^{-1} \sum_{k=1}^{N} \widetilde{X}_k f_U(z - \widetilde{X}_k).$$

Now write $\tilde{h}(z) - h_0(z) = \tilde{h}(z) - h_1(z) + h_1(z) - h_0(z)$. Giné and Guillou (2002) showed that if f_U is bounded with the bandwidth w_1 satisfying (3.12), then

$$(4.15) \qquad \sup_{u\in\mathbb{R}^p} \big|\hat{f}_U(u) - E\hat{f}_U(u)\big| = O_p\Big(\sqrt{\frac{\log w_1^{-1}}{N w_1^p}}\Big).$$

On the other hand, Taylor expansion, the symmetry of K_1 and the assumption (3.10) yield that

$$(4.16) \qquad \sup_{u\in\mathbb{R}^p} \big|E\hat{f}_U(u) - f_U(u)\big| = w_1^2 \sup_{u^*\in\mathbb{R}^p} \Big|\int u' f''_U(u^*) u K_1(u) du\Big| = O_p(w_1^2).$$

The facts (4.15) and (4.16) together imply

$$(4.17) \quad \sup_{u\in\mathbb{R}^p} \big|\hat{f}_U(u) - f_U(u)\big| \leq \sup_{u\in\mathbb{R}^p} \big|\hat{f}_U(u) - E\hat{f}_U(u)\big| + \sup_{u\in\mathbb{R}^p} \big|E\hat{f}_U(u) - f_U(u)\big| = o_p(1).$$

Hence, $E\|X\| < \infty$ implies

$$\sup_{z\in\mathcal{C}} \big\|\tilde{h}(z) - h_1(z)\big\| \leq \int \|x\| \big|\hat{f}_U(x-z) - f_U(z-x)\big| dF_N(x)$$
$$\leq N^{-1} \sum_{k=1}^{N} \|\widetilde{X}_k\| \sup_{u\in\mathbb{R}^p} \big|\hat{f}_U(u) - f_U(u)\big| = o_p(1).$$

It remains to show that $\sup_{z\in\mathcal{C}} \big\|h_1(z) - h_0(z)\big\| = o_p(1)$. By (3.9), f_U has an absolutely integrable characteristic function $\gamma(t)$. Similar to the argument in Bierens (1983), by the inversion formula,

$$f_U(u) = (2\pi)^{-p} \int_{\mathbb{R}^p} \exp(-it'u)\gamma(t) dt.$$

Because $E(h_1(z)) = h_0(z)$, we have

$$h_1(z) - h_0(z) = \frac{1}{(2\pi)^p N} \sum_{k=1}^{N} \Big\{\widetilde{X}_k \int \exp(-it'(z - \widetilde{X}_k))\gamma(t) dt$$
$$- E[\widetilde{X}_k \int \exp(-it'(z - \widetilde{X}_k))\gamma(t) dt]\Big\}.$$

Let $Z_k(t) := \widetilde{X}_k \exp(it'\widetilde{X}_k)$, $Y_k(t) := Z_k(t) - EZ_k(t), G_N(t) := N^{-1} \sum_{k=1}^{N} Y_k(t)$. Then

$$E \sup_{z\in\mathbb{R}^p} \big\|h_1(z) - h_0(z)\big\| = \frac{1}{(2\pi)^p} E \sup_{z\in\mathbb{R}^p} \Big\|\int \exp(-it'z) \times G_N(t)\gamma(t) dt\Big\|$$
$$\leq \frac{1}{(2\pi)^p} \int E\big\|G_N(t)\big\| |\gamma(t)| dt \leq \sup_{t\in\mathbb{R}^p} E\big\|G_N(t)\big\| \frac{1}{(2\pi)^p} \int |\gamma(t)| dt.$$

Because $G_N(t)$ is an average of i.i.d. mean zero complex valued r.v.'s $Y_k(t)$'s and because $E\|Y_1(t)\|^2 \le E\|Z_1(t)\|^2 = E(Z_1'(t)Z_1(-t)) = E\|\widetilde{X}\|^2$, for all t, we obtain

$$E\big\|G_N(t)\big\|^2 = N^{-1}E\|Y_1(t)\|^2 \le N^{-1}E\|\widetilde{X}\|^2, \quad \text{uniformly in } t \in \mathbb{R}^p.$$

Hence $E\sup_{z\in\mathcal{C}}\big\|h_1(z)-h_0(z)\big\| = O(N^{-1/2})$ and $\sup_{z\in\mathcal{C}}\big\|h_1(z)-h_0(z)\big\| = o_p(1)$, which together with (4.17) completes the proof of (4.14)(a).

Next we prove (4.14)(b). Because

$$\Big|\frac{1}{(\hat{f}_Z(z)\vee\epsilon_0)} - \frac{1}{f_Z(z)}\Big| \;\le\; \epsilon_0^{-1}b_0^{-1}\Big|(\hat{f}_Z(z)\vee\epsilon_0) - f_Z(z)\Big|,$$

it suffices to show that

$$\sup_{z\in\mathcal{C}}\Big|(\hat{f}_Z(z)\vee\epsilon_0) - f_Z(z)\Big| = o_p(1). \tag{4.18}$$

An argument similar to the one used for (4.17) and assumptions (3.10) and (3.12) yield

$$\sup_{z\in\mathbb{R}^p}\big|\hat{f}_Z(z) - f_Z(z)\big| = o_p(1). \tag{4.19}$$

Moreover, since

$$\sup_{z\in\mathcal{C}}\Big|(\hat{f}_Z(z)\vee\epsilon_0) - f_Z(z)\Big| \le \sup_{z\in\mathcal{C}}\Big|(\hat{f}_Z(z)\vee\epsilon_0) - \hat{f}_Z(z)\Big| + \sup_{z\in\mathcal{C}}\Big|\hat{f}_Z(z) - f_Z(z)\Big|,$$

it remains to show that

$$\sup_{z\in\mathcal{C}}\Big|(\hat{f}_Z(z)\vee\epsilon_0) - \hat{f}_Z(z)\Big| = o_p(1). \tag{4.20}$$

For a fixed $0<\epsilon<b_0-\epsilon_0$, let $A_{n,\epsilon} := \big\{f_Z(z)-\epsilon \le \hat{f}_Z(z) \le f_Z(z)+\epsilon,\ \ \forall\, z\in\mathcal{C}\big\}$. By (4.19), $\exists\, N_\epsilon$ such that $P(A_{n,\epsilon}) \ge 1-\epsilon$, $\forall\, n > N_\epsilon$. On $A_{n,\epsilon}$, since $\inf_{z\in\mathcal{C}} f_Z(z) > b_0$, we have

$$\hat{f}_Z(z) \ge f_Z(z) - \epsilon > f_Z(z) - b_0 + \epsilon_0 > \epsilon_0.$$

Hence, $\sup_{z\in\mathcal{C}}\big|(\hat{f}_Z(z)\vee\epsilon_0) - \hat{f}_Z(z)\big| = 0$ on $A_{n,\epsilon}$, thereby proving (4.20) and hence (4.14)(b). This also completes the proof of (4.13). $\square$

Let $\Delta(z) := \widehat{h}(z) - h(z)$ and $\delta_i := \theta'\Delta(Z_i)$, $1\le i\le n$. From (4.13) we readily obtain the following result for the later use.

$$\max_{1\le i\le n, Z_i\in\mathcal{C}}|\delta_i| \le \|\theta\|\sup_{z\in\mathcal{C}}\|\Delta(z)\| = o_p(1). \tag{4.21}$$

Proof of Theorem 3.1. The proof is similar to that of Theorem 2.1. First, we shall prove the following.

$$\widehat{M}(\theta) = O_p(1). \tag{4.22}$$

A major difference between $V(x,\theta)$ and $\widehat{V}(x,\theta)$ is that $V(x,\theta)$ is centered at 0 while $\widehat{V}(x,\theta)$ has an asymptotic bias of the order w_1^2 due to the estimation of $h(z)$. Rewrite

$$\widehat{V}(x,\theta) \;=\; n^{-1/2}\sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)\big[I(\zeta_i \le x + \theta'\Delta(Z_i)) - I(-\zeta_i < x - \theta'\Delta(Z_i))\big].$$

Let

$$\begin{aligned}
\xi(z,x) &:= [K_z(x+\theta'\Delta(z)) - K_z(x-\theta'\Delta(z))],\\
J_\theta(z,x) &:= I_{\mathcal{C}}(z)\frac{\tilde{h}(z)}{(\hat{f}_Z(z)\vee\epsilon_0)}\xi(z,x), \qquad \widetilde{J}_\theta(z,x) = I_{\mathcal{C}}(z)\frac{\tilde{h}(z)}{f_Z(z)}\xi(z,x).
\end{aligned}$$

Write $\widehat{V}(x,\theta) \equiv T(x,\theta) + n^{-1/2}\sum_{i=1}^n J_\theta(Z_i,x)$, where

$$T(x,\theta) := \widehat{V}(x,\theta) - n^{-1/2}\sum_{i=1}^n J_\theta(Z_i,x).$$

Note that

$$\begin{aligned}
\sum_{i=1}^n \big\|J_\theta(Z_i,x)\big\|^2 &= \sum_{i=1}^n I_{\mathcal{C}}(Z_i)\frac{\|\tilde{h}(Z_i)\|^2}{(\hat{f}_Z(z)\vee\epsilon_0)^2}|\xi(Z_i,x)|^2\\
&\le \sup_{z\in\mathcal{C}}\Big(\frac{f_Z(z)}{(\hat{f}_Z(z)\vee\epsilon_0)}\Big)^2\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\frac{\|\tilde{h}(Z_i)\|^2}{f_Z^2(Z_i)}|\xi(Z_i,x)|^2\\
&= \sup_{z\in\mathcal{C}}\Big(\frac{f_Z(z)}{(\hat{f}_Z(z)\vee\epsilon_0)}\Big)^2\sum_{i=1}^n \big\|\widetilde{J}_\theta(Z_i,x)\big\|^2.
\end{aligned}$$

Then

$$\begin{aligned}
(4.23)\qquad \int \big\|\widehat{V}(x,\theta)\big\|^2 dG(x) &\le 2\int \big\|T(x,\theta)\big\|^2 dG(x) + 2\int n^{-1}\big\|\sum_{i=1}^n J_\theta(Z_i,x)\big\|^2 dG(x)\\
&\le 2\int \big\|T(x,\theta)\big\|^2 dG(x) + 2\int\sum_{i=1}^n \big\|J_\theta(Z_i,x)\big\|^2 dG(x) \quad \text{(by the C–S ineq.)}\\
&\le 2\int \big\|T(x,\theta)\big\|^2 dG(x) + 2\sup_{z\in\mathcal{C}}\Big(\frac{f_Z(z)}{(\hat{f}_Z(z)\vee\epsilon_0)}\Big)^2\int\sum_{i=1}^n \big\|\tilde{J}_\theta(Z_i,x)\big\|^2 dG(x).
\end{aligned}$$

It thus suffices to show that the two terms on the right hand side above are $O_p(1)$.

Because $T(x,\theta) \equiv T(-x,\theta)$ and G is symmetric around zero, we obtain

$$\int \|T(x,\theta)\|^2 dG(x) = 2\int_0^\infty \|T(x,\theta)\|^2 dG(x).$$

Recall $\delta(z) = \theta'\Delta(z)$. Let $E_{\mathcal{C}}(L(Z)) := E\big(L(Z)I(Z\in\mathcal{C})\big)$, for any integrable function L of Z. The law of total expectation and (3.6) yield

$$\begin{aligned}
&E\int_0^\infty \big\|T(x,\theta)\big\|^2 dG(x)\\
&= E_{\mathcal{C}}\int_0^\infty \|\widehat{h}(Z)\|^2 \mathrm{Var}\big(I(\zeta\le x+\delta(Z)) - I(-\zeta< x-\delta(Z))\big)dG(x)\\
&\le \|g\|_\infty E_{\mathcal{C}}\Big(\|\widehat{h}(Z)\|^2\int_0^\infty \big\{2 - K_Z(x-\delta(Z)) - K_Z(x+\delta(Z))\big\}dx\Big).
\end{aligned}$$

Moreover, conditioning on the validation data and Z and using the independence between Z and validation data, we obtain

$$\begin{aligned}
\int_0^\infty \big\{1-K_Z(x-\delta(Z))\big\}dx &= \int_0^\infty P(\zeta+\delta(Z)>x)dx \le E\big(|\zeta|\big|Z\big)+|\delta(Z)|,\\
\int_0^\infty \big\{1-K_Z(x+\delta(Z))\big\}dx &= \int_0^\infty P(\zeta-\delta(Z)>x)dx \le E\big(|\zeta|\big|Z\big)+|\delta(Z)|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& E\int_0^\infty \big\|T(x,\theta)\big\|^2 dG(x)\\
&\le 2\|g\|_\infty E_{\mathcal{C}}\Big(\|\widehat{h}(Z)\|^2[E(|\zeta||Z)+|\delta(Z)|]\Big)\\
&= 2\|g\|_\infty E_{\mathcal{C}}\Big\{\|\widehat{h}(Z)\|^2|\zeta|\Big\} + 2\|g\|_\infty E_{\mathcal{C}}\Big\{\|\widehat{h}(Z)\|^2|\delta(Z)|\Big\}\\
&\le 2\|g\|_\infty \epsilon_0^{-2}\Big(E_{\mathcal{C}}\Big\{\|\tilde{h}(Z)\|^2|\zeta|\Big\} + \|\theta\|E_{\mathcal{C}}\Big\{\|\tilde{h}(Z)\|^2\|\Delta(Z)\|\Big\}\Big)\\
&\le 2\|g\|_\infty \epsilon_0^{-2}\Big(\sqrt{E_{\mathcal{C}}\|\tilde{h}(Z)\|^4 E\zeta^2} + \|\theta\|\sqrt{E_{\mathcal{C}}\|\tilde{h}(Z)\|^4 E_{\mathcal{C}}\|\Delta(Z)\|^2}\Big).
\end{aligned}$$

To proceed further, we need the following results obtained by direct calculations.

$$\begin{aligned}
(4.24)\qquad E\tilde{h}(z) &= h_0(z)+\frac{1}{2}w_1^2\widetilde{B}(z)+o(w_1^2),\\
E\|\tilde{h}(z)\|^2 &= N^{-4}E\sum_{i,j=1}^N\sum_{q,r=1}^N \widetilde{X}_i'\widetilde{X}_j K_w(\widetilde{U}_q+\widetilde{X}_i-z)K_w(\widetilde{U}_r+\widetilde{X}_j-z)\\
&= \|h_0(z)\|^2 + w_1^2 h_0(z)'\widetilde{B}(z)+o(w_1^2)+O(N^{-1}) = \|h_0(z)\|^2+o(1),\\
E\|\tilde{h}(z)\|^4 &= N^{-8}E\Big\{\sum_{i,j,k,l}\sum_{q,r,s,t}\widetilde{X}_i'\widetilde{X}_j K_w(\widetilde{U}_q+\widetilde{X}_i-z)K_w(\widetilde{U}_r+\widetilde{X}_j-z)\\
&\qquad\times \widetilde{X}_k'\widetilde{X}_l K_w(\widetilde{U}_s+\widetilde{X}_k-z)K_w(\widetilde{U}_t+\widetilde{X}_l-z)\Big\}\\
&= \|h_0(z)\|^4 + 2w_1^2\|h_0(z)\|^2 h_0(z)'\widetilde{B}(z)+o(w_1^2)+O(N^{-1}) \to \|h_0(z)\|^4.
\end{aligned}$$

We further obtain that

$$\begin{aligned}
(4.25)\qquad & E\|\Delta(z)\|^2 = E\|\widehat{h}(z)-h(z)\|^2\\
&\le 2E\Big\{\Big\|\frac{1}{(\hat{f}_Z(z)\vee\epsilon_0)}(\tilde{h}(z)-h_0(z))\Big\|^2\Big\}\\
&\qquad + 2E\Big\{\Big\|h_0(z)\Big(\frac{1}{(\hat{f}_Z(z)\vee\epsilon_0)}-\frac{1}{f_Z(z)}\Big)\Big\|^2\Big\}\\
&\le 2\epsilon_0^{-2}E\|\tilde{h}(z)-h_0(z)\|^2 + 2(\epsilon_0^{-1}+b_0^{-1})^2\|h_0(z)\|^2\\
&\le 4\epsilon_0^{-2}E\|\tilde{h}(z)\|^2 + \{4\epsilon_0^{-2}+2(\epsilon_0^{-1}+b_0^{-1})^2\}\|h_0(z)\|^2\\
&= \Big\{\frac{8}{\epsilon_0^2}+2\Big(\frac{1}{\epsilon_0}+\frac{1}{b_0}\Big)^2\Big\}\|h_0(z)\|^2+o(1).
\end{aligned}$$

Combine (4.24) and (4.25) with the assumption (3.5), which implies $E\|h_0(Z)\|^4<\infty$, to conclude that $E\int_0^\infty \big\|T(x,\theta)\big\|^2 dG(x) = O(1)$.

Next, consider the second term in the bound of (4.23). By (4.18), $\sup_{z\in\mathcal{C}}\big\{f_Z(z)/(\hat f_Z(z)\vee\epsilon_0)\big\}^2 = O_p(1)$. It remains to show that

$$\int\sum_{i=1}^n \big\|\tilde J_\theta(Z_i,x)\big\|^2 dG(x) = O_p(1). \tag{4.26}$$

Recall $\delta_i = \theta'\Delta(Z_i)$. We have

$$\begin{aligned}
&\int\sum_{i=1}^n \big\|\tilde J_\theta(Z_i,x)\big\|^2 dG(x)\\
&= \sum_{i=1}^n \int I_{\mathcal{C}}(Z_i)\frac{\|\tilde h(Z_i)\|^2}{f_Z^2(Z_i)}\Big(K_{Z_i}(x+\delta_i)-K_{Z_i}(x-\delta_i)\Big)^2 dG(x)\\
&\le \sum_{i=1}^n \int I_{\mathcal{C}}(Z_i)\frac{\|\tilde h(Z_i)\|^2}{f_Z^2(Z_i)}\Big(\int_{-|\delta_i|}^{|\delta_i|}\kappa_{Z_i}(x+s)ds\Big)^2 dG(x)\\
&\le 2\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\frac{\|\tilde h(Z_i)\|^2}{f_Z^2(Z_i)}|\delta_i|\int_{-|\delta_i|}^{|\delta_i|}\int \kappa_{Z_i}^2(x+s)dG(x)ds\\
&\le 4\max_{1\le i\le n, Z_i\in\mathcal{C}}\Big((2|\delta_i|)^{-1}\int_{-|\delta_i|}^{|\delta_i|}\int\kappa_{Z_i}^2(x+s)dG(x)ds\Big)\times\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\frac{\|\tilde h(Z_i)\|^2}{f_Z^2(Z_i)}\delta_i^2.
\end{aligned}$$

The last but one inequality is obtained by using the C–S inequality and the Fubini Theorem. Under (3.7), we have $\sup_{z,x} k_z(x)\le\|f\|_\infty<\infty$. Hence

$$\begin{aligned}
\big(2\delta_i\big)^{-1}\int_{-|\delta_i|}^{|\delta_i|}\int\kappa_{Z_i}^2(x+s)dG(x) &\le \|f\|_\infty^2, && \text{if } G \text{ is a d.f.}\\
&\le \|g\|_\infty\|f\|_\infty, && \text{if } dG(x)=g(x)dx, \|g\|_\infty<\infty.
\end{aligned}$$

Let

$$D_n := \sum_{i=1}^n I_{\mathcal{C}}(Z_i)\|\tilde h(Z_i)\|^2\delta_i^2.$$

Using the above bounds we obtain

$$\begin{aligned}
\int\sum_{i=1}^n\big\|\tilde J_\theta(Z_i,x)\big\|^2 dG(x) &\le \frac{4}{b_0^2}\|f\|_\infty^2 D_n, && \text{if } G \text{ is a d.f.}\\
&\le \frac{4}{b_0^2}\|f\|_\infty\|g\|_\infty D_n, && \text{if } dG(x)=g(x)dx, \|g\|_\infty<\infty.
\end{aligned}$$

It thus remains to show that $D_n = O_p(1)$. By the definitions of $\tilde h(z)$ and δ_i and (3.8),

$$\begin{aligned}
&E(D_n) = nE\big\{I_{\mathcal{C}}(Z_1)\tilde h(Z_1)'\tilde h(Z_1)\delta_1^2\big\}\\
&= \frac{n}{N^8}\sum_{i,j,k,l=1}^N\sum_{q,r,s,t=1}^N E\Big\{I_{\mathcal{C}}(Z_1)\widetilde X_i'\widetilde X_j K_w(\widetilde U_q+\widetilde X_i-Z_1)K_w(\widetilde U_r+\widetilde X_j-Z_1)\\
&\qquad\times\theta'\big[\widetilde X_k K_w\big(\frac{\widetilde U_s+\widetilde X_k-Z_1}{f_Z(Z_1)}\big)-h(Z_1)\big]\theta'\big[\widetilde X_l K_w\big(\frac{\widetilde U_t+\widetilde X_l-Z_1}{f_Z(Z_1)}\big)-h(Z_1)\big]\Big\}\\
&= \frac{n}{N^8}\sum_{i\ne j\ne k=l}\sum_{q\ne r\ne s\ne t}+\sum_{i\ne k\ne j\ne l}\sum_{q\ne r\ne s=t}+\sum_{i\ne j\ne k\ne l}\sum_{q\ne r\ne s\ne t}E\Big\{...\Big\}+\text{ other terms},\\
&:= \tilde D_1+\tilde D_2+\tilde D_3+D_R.
\end{aligned}$$

In $\tilde{D}_1$, $k=l$ and all other indices different. Hence, $\tilde{D}_1$ is bounded from the above by

$$\frac{n[N(N-1)(N-2)]^2(N-3)}{b_0^2N^8}E\Big\{I_{\mathcal{C}}(Z_1)\widetilde{X}_1'\widetilde{X}_2K_w(\widetilde{U}_1+\widetilde{X}_1-Z_1)K_w(\widetilde{U}_2+\widetilde{X}_2-Z_1)$$
$$\times\,\theta'\big[\widetilde{X}_3K_w\big(\widetilde{U}_3+\widetilde{X}_3-Z_1\big)-h(Z_1)\big]\theta'\big[\widetilde{X}_3K_w\big(\widetilde{U}_4+\widetilde{X}_3-Z_1\big)-h(Z_1)\big]\Big\}$$
$$=O(nN^{-1})E\Big\{I_{\mathcal{C}}(Z)\|h(Z)\|^2\Big[E\big((\theta'\widetilde{X})^2f_U^2(Z-\widetilde{X})\big|Z\big)\Big]\Big\}:=O\big(nN^{-1}L_1(\theta)\big).$$

Similarly,

$$\tilde{D}_2=O(nN^{-1})E\Big\{I_{\mathcal{C}}(Z)\|h(Z)\|^2E\big((\theta'(Z-\widetilde{U}))^2f_X^2(Z-\widetilde{U})\big|Z\big)\Big\}$$
$$:=O(nN^{-1})L_2(\theta).$$

In $\tilde{D}_3$, with indices $i\neq j\neq k\neq l$ and $q\neq r\neq s\neq t$, we get

$$\tilde{D}_3=O(nw_1^4)\big(\theta'E[I_{\mathcal{C}}(Z)B(Z)]\big)^2=O(nw_1^4\|\theta\|^2E\|B(Z)\|^2):=O(nw_1^4)L_3(\theta).$$

Similar calculations show that all other terms in $E(D_n)$ are $o(\tilde{D}_1+\tilde{D}_2+\tilde{D}_3)$. The assumptions (3.5) and (3.9) imply that $L_i(\theta), i=1,2,3$ are finite. Therefore, under (3.5) and (3.15), we get, for sufficiently large $n\wedge N$,

$$\text{(4.27)}\qquad E(D_n)\;=\;O(nN^{-1})\big[L_1(\theta)+L_2(\theta)\big]+nw_1^4L_3(\theta)+o(1)$$
$$\to\;\lambda(L_1(\theta)+L_2(\theta))+CL_3(\theta)<\infty.$$

Since $D_n\geq 0$, by the Markov inequality, $D_n=O_p(1)$. This completes the proof of (4.22).

Next, we prove the asymptotic normality of the estimator $\widehat{\theta}$. Let $\hat{c}_{ni}=n^{-1/2}\widehat{h}(Z_i)$ and

$$\text{(4.28)}\qquad \widehat{W}(x,s):=n^{-1/2}\sum_{i=1}^nI_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)\Big[I\big(\zeta_i\leq x+\delta_i+s'\hat{c}_{ni}\big)-K_{Z_i}(x+\delta_i+s'\hat{c}_{ni})\Big],$$
$$\widehat{T}_n:=\int\widehat{\Gamma}(x)'\big[\widehat{W}(x,0)+\widehat{W}(-x,0)\big]dG(x),$$
$$\widehat{M}_1(s):=\widehat{M}(\theta)+4\widehat{T}_n's+4\,s'\widehat{\mathcal{G}}s,\qquad \widehat{s}:=\text{argmin}_s\,\widehat{M}_1(s)=-\frac{1}{2}\widehat{\mathcal{G}}^{-1}\widehat{T}_n,$$

where $\widehat{\Gamma}_n(x)$, $\widehat{\Gamma}(x)$ and $\widehat{\mathcal{G}}$ are defined in (3.4).

Similar to the proof of Theorem 2.1, we aim to show that

$$\text{(4.29)}\qquad \sup_{\|s\|\leq b}\big|\widehat{M}(\theta+n^{-1/2}s)-\widehat{M}_1(s)\big|\to_p 0,\quad\forall\,0<b<\infty,\quad\|n^{1/2}(\widehat{\theta}-\theta)-\widehat{s}\|\to_p 0.$$

Define

$$\widehat{U}(x,s)=n^{-1/2}\sum_{i=1}^nI_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)I(\zeta_i\leq x+\delta_i+s'\hat{c}_{ni}),$$
$$\widehat{J}(x,s)=n^{-1/2}\sum_{i=1}^nI_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)K_{Z_i}(x+\delta_i+s'\hat{c}_{ni}).$$

Then $\widehat{W}(x,s) = \widehat{U}(x,s) - \widehat{J}(x,s)$, and

$$\begin{aligned}
\widehat{M}(\theta + n^{-1/2}s) &= \int \Big\| \widehat{U}(x,s) + \widehat{U}(-x,s) - n^{-1/2}\sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)\Big\|^2 dG(x) \\
&= \int \Big\| \Big\{\widehat{W}(x,s) - \widehat{W}(x,0)\Big\} + \Big\{\widehat{W}(-x,s) - \widehat{W}(-x,0)\Big\} \\
&\quad + \Big\{\widehat{J}(x,s) - \widehat{J}(x,0) - \widehat{\Gamma}_n(x)s\Big\} + \Big\{\widehat{J}(-x,s) - \widehat{J}(-x,0) - \widehat{\Gamma}_n(-x)s\Big\} \\
&\quad + \Big\{\widehat{W}(x,0) + \widehat{W}(-x,0) + 2\widehat{\Gamma}_n(x)s\Big\}\Big\|^2 dG(x).
\end{aligned}$$

Use this decomposition and an argument similar to the one use in the proof of (2.25) to obtain (4.29), under the given assumptions. Details are skipped for the sake of brevity.

Now it suffices to derive the asymptotic normality of $\widehat{T}_n$ to complete the proof. Similar to T_n, $\widehat{T}_n$ can be rewritten as

$$\begin{aligned}
\widehat{T}_n &= n^{-1/2}\int \Gamma(x)' \sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)\Big[I(\zeta_i \le x + \delta_i) + I(-\zeta_i < x - \delta_i)\Big] dG(x) \\
&= n^{-1/2}\sum_{i=1}^{n}\int I_{\mathcal{C}}(Z_i)\mu(z)\widehat{h}(Z_i)\varphi_z\big(\zeta_i - \theta'\Delta(Z_i)\big) dQ(z) \\
&= n^{-1/2}\sum_{i=1}^{n}\int I_{\mathcal{C}}(Z_i)\mu(z)\widehat{h}(Z_i)\varphi_z(\zeta_i) dQ(z) \\
&\quad + 2n^{-1/2}\sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)[\theta'\Delta(Z_i)]\int \mu(z)\widehat{h}(Z_i)\psi_z(\zeta_i) dQ(z) + R_n \\
&:= \widehat{T}_{n1} + \widehat{T}_{n2} + R_n, \quad \text{(say)},
\end{aligned}$$

where R_n is the reminder of the Taylor expansion and $R_n = o_p(\widehat{T}_{n1} + \widehat{T}_{n2})$. Therefore, it suffices to analyze $\widehat{T}_{n1}$ and $\widehat{T}_{n2}$. First, we rewrite

$$\widehat{T}_{n1} \;=\; \frac{n^{1/2}}{nN^2}\sum_{i=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N}\int \frac{I_{\mathcal{C}}(Z_i)}{(\hat{f}_Z(Z_i)\vee\epsilon_0)}\mu(z)\widetilde{X}_k K_w(\widetilde{U}_l + \widetilde{X}_k - Z_i)\varphi_z(\zeta_i) dQ(z)$$

Define

$$\begin{aligned}
\phi_1(Z_i, \zeta_i, \widetilde{X}_k, \widetilde{U}_l) &= \int \frac{I_{\mathcal{C}}(Z_i)}{f_Z(Z_i)}\mu(z)\widetilde{X}_k K_w(\widetilde{U}_l + \widetilde{X}_k - Z_i)\varphi_z(\zeta_i) dQ(z) \\
\widetilde{T}_{n1} &= \frac{n^{1/2}}{nN^2 w_1^p}\sum_{i=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{N}\phi_1(Z_i, \zeta_i, \widetilde{X}_k, \widetilde{U}_l).
\end{aligned}$$

Then (4.18) implies that $\widehat{T}_{n1} = \widetilde{T}_{n1} + o_p(\widetilde{T}_{n1})$. Therefore it suffices to study the asymptotic property of $\widetilde{T}_{n1}$ for $\widehat{T}_{n1}$. It can be seen that $\widetilde{T}_{n1}$ is a 3-sample U statistic based on $\{(Z_i, \zeta_i), 1 \le k \le n\}$, $\{\widetilde{X}_k, 1 \le k \le N\}$ and $\{\widetilde{U}_l, 1 \le l \le N\}$. We further compute that

$$\begin{aligned}
E(\phi_1|\widetilde{X}_1) &= E(\phi_1|\widetilde{U}_1) = E\phi_1 = 0, \\
E(\phi_1|Z_1, \zeta_1) &= \int I_{\mathcal{C}}(Z_1)\mu(z)h(Z_1)\varphi_z(\zeta_1) dQ(z) + O_p(w_1^2).
\end{aligned}$$

So asymptotically $\widehat{T}_{n1}$ only depends on the sample $\{(Z_i, \zeta_i), 1 \le i \le n\}$. Then Theorem 6.1.4 in Lehmann (1999) implies that, with $n/N \to \lambda$, $L(Z, \zeta) = \int I_{\mathcal{C}}(Z)\mu(z)h(Z)\varphi_z(\zeta)dQ(z)$ and $\Sigma_0 = \mathrm{Cov}(L(Z, \zeta))$, $\widetilde{T}_{n1} \to_D N\big(0, \Sigma_0\big)$. Therefore, for $0 \le \lambda < \infty$,

$$\widehat{T}_{n1} \to_D \mathcal{N}_p\big(0, \Sigma_0\big). \tag{4.30}$$

Similarly, $\widehat{T}_{n2}$ is also a 3-sample U statistic. Specifically,

$$\begin{aligned}
\widehat{T}_{n2} &= 2n^{-1/2} \sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)[\theta' \Delta(Z_i)] \int \mu(z)\widehat{h}(Z_i)\psi_z(\zeta_i)dQ(z) \\
&= \frac{2n^{1/2}}{nN^4} \sum_{i,k,l,s,t} \frac{I_{\mathcal{C}}(Z_i)}{f_Z(Z_i)} \theta' \big[\widetilde{X}_k K_w(\widetilde{U}_s + \widetilde{X}_k - Z_i) - h(Z_i) f_Z(Z_i)\big] \\
&\qquad \times \int \mu(z)\widetilde{X}_l \frac{K_w(\widetilde{U}_t + \widetilde{X}_l - Z_i)}{f_Z(Z_i)} \psi_z(\zeta_i)dQ(z) \\
&=: \frac{2n^{1/2}}{nN^4} \sum_{i,k,l,s,t} g_1(Z_i, \widetilde{X}_k, \widetilde{U}_s) g_2(Z_i, \zeta_i, \widetilde{X}_l, \widetilde{U}_t) \\
&= \frac{2n^{1/2}}{nN^4} \Big\{ \sum_{i,k\neq l,s\neq t} + \sum_{i,k\neq l,s=t} + \sum_{i,k=l,s\neq t} + \sum_{i,k=l,s=t} \Big\} g_1(Z_i, \widetilde{X}_k, \widetilde{U}_s) g_2(Z_i, \zeta_i, \widetilde{X}_l, \widetilde{U}_t) \\
&=: S_{n1} + S_{n2} + S_{n3} + S_{n4}.
\end{aligned}$$

We will show that S_{n1} is asymptotically normally distributed while $S_{nj}, j = 2, 3, 4$ are asymptotically negligible compared to S_{n1}. Consider

$$\begin{aligned}
S_{n1} &:= \frac{2n^{1/2}}{nN^4} \sum_{i,k\neq l,s\neq t} g_1(Z_i, \widetilde{X}_k, \widetilde{U}_s) g_2(Z_i, \zeta_i, \widetilde{X}_l, \widetilde{U}_t) \\
&= \frac{n^{1/2}}{nN^4} \sum_{i,k\neq l,s\neq t} \frac{1}{2} \Big\{ g_1(Z_i, \widetilde{X}_k, \widetilde{U}_s) g_2(Z_i, \zeta_i, \widetilde{X}_l, \widetilde{U}_t) + g_1(Z_i, \widetilde{X}_l, \widetilde{U}_s) g_2(Z_i, \zeta_i, \widetilde{X}_k, \widetilde{U}_t) \\
&\qquad + g_1(Z_i, \widetilde{X}_k, \widetilde{U}_t) g_2(Z_i, \zeta_i, \widetilde{X}_l, \widetilde{U}_s) + g_1(Z_i, \widetilde{X}_l, \widetilde{U}_t) g_2(Z_i, \zeta_i, \widetilde{X}_k, \widetilde{U}_s) \Big\} \\
&=: \frac{n^{1/2}}{nN^4} \sum_{i,k\neq l,s\neq t} \phi_2(Z_i, \zeta_i, \widetilde{X}_k, \widetilde{X}_l, \widetilde{U}_s, \widetilde{U}_t).
\end{aligned}$$

Note that ϕ_2 is symmetric within each sample. Recall the notation from (3.4). Direct calculation shows that

$$\begin{aligned}
&E(g_1) = w_1^2 E[I_{\mathcal{C}}(Z)\theta' B(Z)] = O(w_1^2), \quad E(g_2) = \boldsymbol{\alpha} + O(w_1^2), \\
&E(g_1|\widetilde{X}_1) = R_\theta(\widetilde{X}_1) + o_p(1), \quad E(g_1|\widetilde{U}_1) = S_\theta(\widetilde{U}_1) + o_p(1), \\
&E(\phi_2|\widetilde{X}_1) = E(g_1|\widetilde{X}_1)E(g_2) + E(g_1)E(g_2|\widetilde{X}_1) = R_\theta(\widetilde{X}_1)\boldsymbol{\alpha} + o_p(1), \\
&E(\phi_2|\widetilde{U}_1) = E(g_1|\widetilde{U}_1)E(g_2) + E(g_1)E(g_2|\widetilde{U}_1) = S_\theta(\widetilde{U}_1)\boldsymbol{\alpha} + o_p(1), \\
&E(\phi_2|Z_1, \zeta_1) = 2w_1^2 I_{\mathcal{C}}(Z_1)\theta' B(Z_1) \int \mu(z)h(Z_1)\psi_z(\zeta_1)dQ(z) + o_p(w_1^2).
\end{aligned}$$

Furthermore, applying Theorem 6.1.4 in Lehmann (1999) to S_{n1}, we obtain that for $\lambda < \infty$,

$$\begin{aligned}
\widetilde{S}_{n1} &:= \sqrt{\frac{n+2N}{n}}\Big\{S_{n1} - 2\,n^{1/2}w_1^2\boldsymbol{\beta}(\theta)\Big\} \to_D \mathcal{N}_p(0, \widetilde{\Sigma}),\\
\widetilde{\Sigma} &= (1+2/\lambda)\mathrm{Cov}(E(\phi_2|Z_1,\zeta_1)) + (2+\lambda)\mathrm{Cov}(E(\phi_2|\widetilde{X}_1)) + 4(2+\lambda)\mathrm{Cov}(E(\phi_2|\widetilde{U}_1))\\
&= (2+\lambda)\mathrm{Cov}(R_\theta(\widetilde{X}_1)\boldsymbol{\alpha}) + 4(2+\lambda)\mathrm{Cov}(S_\theta(\widetilde{U}_1)\boldsymbol{\alpha})\\
&= 4(2+\lambda)[\mathrm{Var}(R_\theta(\widetilde{X})) + \mathrm{Var}S_\theta(\widetilde{U})]\boldsymbol{\alpha}\boldsymbol{\alpha}'.
\end{aligned}$$

Therefore, under (3.15), we have

$$S_{n1} - 2\,n^{1/2}w_1^2\boldsymbol{\beta}(\theta) \to_D \mathcal{N}_p(0, 4\lambda\Sigma_\theta). \tag{4.31}$$

Next, we analyze S_{n2}, S_{n3} and S_{n4}. Similar to Lemma 6.4 in Geng and Koul (2017), tedious calculation shows that $E\|S_{n2}\|^2 = O(nN^{-2}) = E\|S_{n3}\|^2$, $E\|S_{n2}\|^2 = O(nN^{-4})$. Therefore, under (3.15), $S_{n2} + S_{n3} + S_{n4} = o_p(1)$. Combine this with (4.31) to obtain that

$$\widehat{T}_{n2} - 2\,n^{1/2}w_1^2\boldsymbol{\beta}(\theta) \to_D \mathcal{N}_p(0, 4\lambda\Sigma_\theta). \tag{4.32}$$

Finally, (4.30), (4.32) and the independence among the three samples imply that, under (3.15), $\widehat{T}_n - 2\,n^{1/2}w_1^2\boldsymbol{\beta}(\theta) \to_D \mathcal{N}_p(0, \Sigma_0 + 4\lambda\Sigma_\theta)$. This fact, together with (4.28) and (4.29), completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. With $\widehat{V}(x,t)$ and $\widehat{M}(t)$ as in (3.3), let

$$\widehat{\mathcal{V}}(x,t) = n^{-1/2}N^{1/2}\widehat{V}(x,t), \qquad \widehat{\mathcal{M}}(t) = n^{-1}N\widehat{M}(t).$$

Note that $\widehat{\theta} = \mathrm{argmin}_{t\in\mathbb{R}^p}\widehat{\mathcal{M}}(t)$. The proof is similar to the proof of Theorem 3.1 with $\widehat{M}(\theta)$ replaced by $\widehat{\mathcal{M}}(\theta)$.

First, we show that $\widehat{\mathcal{M}}(\theta) = O_p(1)$. In fact, from the proof of Theorem 3.1, we obtain that $\widehat{M}(\theta) = O_p(D_n)$. Therefore, it suffices to show that $Nn^{-1}D_n = O_p(1)$ for $\lambda = \infty$. In view of (4.27), under (3.16),

$$\begin{aligned}
E(Nn^{-1}D_n) &= Nn^{-1} \times O\Big(nN^{-1}[L_1(\theta) + L_2(\theta)] + nw_1^4L_3(\theta)\Big)\\
&= O\big(L_1(\theta) + L_2(\theta) + C_2L_3(\theta)\big) < \infty.
\end{aligned}$$

Let $\check{c}_{Ni} = N^{-1/2}\widehat{h}(Z_i)$, define

$$\begin{aligned}
\widetilde{W}(x,s) &:= \frac{N^{1/2}}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\widehat{h}(Z_i)\Big[I\Big(\zeta_i \le x + \delta_i + s'\check{c}_{Ni}\Big) - K_{Z_i}(x + \delta_i + s'\check{c}_{Ni})\Big], \quad (4.33)\\
\widetilde{T}_n &:= \int \widehat{\Gamma}(x)'\big[\widetilde{W}(x,0) + \widetilde{W}(-x,0)\big]dG(x),\\
\widetilde{\mathcal{M}}(s) &:= \widehat{\mathcal{M}}(\theta) + 4\widetilde{T}_n's + 4\,s'\widehat{\mathcal{G}}s, \quad \check{s} := \mathrm{argmin}_s\,\widetilde{\mathcal{M}}(s) = -\frac{1}{2}\widehat{\mathcal{G}}^{-1}\widetilde{T}_n,
\end{aligned}$$

Argue as in the proof of Theorem 2.1 to obtain that

$$\sup_{\|s\|\le b}\big|\widehat{\mathcal{M}}(\theta + N^{-1/2}s) - \widetilde{\mathcal{M}}(s)\big| \to_p 0, \forall\ 0 < b < \infty, \quad \|N^{1/2}(\widehat{\theta} - \theta) - \check{s}\| \to_p 0. \tag{4.34}$$

Finally, we derive the asymptotic distribution of $\widetilde{T}_n$. With $\widehat{T}_n$ as in (4.28), rewrite $\widetilde{T}_n = n^{-1/2}N^{1/2}\,\widehat{T}_n = n^{-1/2}N^{1/2}\,(\widehat{T}_{n1} + \widehat{T}_{n2}) + o_p(1)$. Argue as in the proof of Lemma 6.4 of Geng and Koul (2017) to obtain that $E\|\widehat{T}_{n1}\|^2 = O(1)$. Since $n/N \to \lambda = \infty$, we have $\sqrt{N/n}\,\widehat{T}_{n1} = o_p(1)$. Furthermore, Theorem 6.1.4 in Lehmann (1999) yields that, under (3.16), $\sqrt{N/n}\,\widehat{T}_{n2} - 2N^{1/2}w_1^2\boldsymbol{\beta}(\theta) \to_D \mathcal{N}_p(0, 4\Sigma_\theta)$. Therefore, $\widetilde{T}_n - 2N^{1/2}w_1^2\boldsymbol{\beta}(\theta) \to_D \mathcal{N}_p(0, 4\Sigma_\theta)$. Eventually, these facts, (4.33) and (4.34) complete the proof. □

References

Al-Sharadqah, A. (2018). A comparative study for two newly developed estimators for the slope in the functional EIV linear model. *Theory of Probability and Mathematical Statistics*, **97**, 211–236.

Bierens, H. J. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association*, **78(383)**, 699–707.

Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications.* CRC press.

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.

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

Cook, J. R., and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.*, **89(428)**, 1314–1328.

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.

Geng, P., Koul, H. L. (2017). Model checking in Tobit regression with measurement errors using validation data. *J. Statist. Planning & Inference*, **190**, 15–31.

Giné, E., and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *In Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **38(6)**, 907–921.

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

Koul, Hira L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear 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.

Lehmann, E.L. (1999). *Elements of Large-Sample Theory*, Springer, New York, N.Y., USA.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, New York.

Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, **72(3)**, 583–592.

Wand, M.P. and Jones, M.C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, **9**, 97–116.