Statistica Sinica **20** (2010), 1551-1579

# MODEL CHECKING IN PARTIAL LINEAR REGRESSION MODELS WITH BERKSON MEASUREMENT ERRORS

Hira L. Koul and Weixing Song

*Michigan State University and Kansas State University*

*Abstract:* This paper discusses the problem of fitting a parametric model to the nonparametric component in partially linear regression models when covariates in parametric and nonparametric parts are subject to Berkson measurement errors. The proposed test is based on the supremum of a martingale transform of a certain partial sum process of calibrated residuals. The asymptotic null distribution of this transformed process is shown to be the same as that of a time transformed standard Brownian motion. Consistency of this sequence of tests against some fixed alternatives and asymptotic power under some local nonparametric alternatives are also discussed. A simulation study is conducted to assess the finite sample performance of the proposed test. A Monte Carlo power comparison with some existing tests shows some superiority of the proposed test at the chosen alternatives.

*Key words and phrases:* Asymptotically distribution free, consistency, local alternatives, marked empirical process.

## 1. Introduction

In this paper, we are interested in developing a lack-of-fit test for checking if the nonparametric component takes on a parametric form in the partially linear regression model with Berkson measurement errors. More precisely, in the model under consideration one observes $(S, Z, Y)$ obeying the relations

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

where $X$ is a $p$-dimensional random vector, $T$ is a scalar random variable, $\beta$ is an unknown $p$-dimensional vector of regression parameters, and $g$ is an unknown real-valued measurable function. The random variables (r.v.'s) $\xi$ and $\eta$ are $p$-dimensional and 1-dimensional measurement errors, respectively. All r.v.'s $\varepsilon, (Z, S)$, $\xi$, and $\eta$ are assumed to be mutually independent, with $\varepsilon$, $\eta$ having zero means, finite variances, and $\xi$ having zero mean and known covariance matrix $\Sigma_\xi$. The distributions of $\varepsilon$ and $\xi$ are assumed to be otherwise unknown, while that of $\eta$ is assumed to be known. Under these assumptions, the above model is identifiable and the covariance of $X$ and $T$ is the same as that of $Z$ and $S$. See

Hu and Schennach (2008) for more on identifiability in Berkson and other nonclassical measurement error models. A discussion on the availability of density function or covariance matrices of measurement errors can be found in Delaigle, Hall, and Qiu (2006).

Traditionally, in some cases, the variables $Z$ and $S$ are called controlled variables as their values are deterministic. But as in Delaigle, Hall, and Qiu (2006), we also treat these variables as random. These authors cite many examples where the controlled variables are genuinely random, rather than deterministic, cf. Reeves et al. (1998), Thomas et al. (1999), Raaschou-Nielsen et al. (2001), Stram, Huberman and Wu (2002) and Lubin et al. (2005). See also Huwang and Huang (2000) and Wang (2004) for more on this point.

Here, we are interested in testing whether $g$ in (1.1) is of a parametric form or not. Thus, given a parametric family of functions $\{g_\gamma; \gamma \in \Gamma\}$, where $\Gamma$ is a subset of $\mathbb{R}^q$ with $q$ being a known positive integer, one is interested in testing $H_0 : g(t) = g_\gamma(t)$, for some $\gamma \in \Gamma$, and for all $t \in \mathbb{R}$, versus $H_1 : H_0$ is not true. This problem is of interest because knowing $g$ is parametric would lead to more accurate inference about the underlying parameters.

Lack-of-fit testing in other regression models without measurement errors has been widely studied in the literature, cf., Hart (1997), Stute, Thies, and Zhu (1998), Stute and Zhu (2002, 2005), Zhu and Ng (2003), Liang (2006), and Khmaladze and Koul (2004).

In this paper, we provide a test for $H_0$ based on a martingale transform, *a la* Khmaladze (1979) and Stute, Thies, and Zhu (1998) (STZ), of the marked empirical process of calibrated residuals. A similar idea is used to construct lack-of-tests in a purely nonparametric regression set-up with Berkson measurement error in Koul and Song (2008), but its extension to the above partial linear model set-up is far from trivial. It is not *a priori* clear how the presence of linear component in the model affects asymptotic properties of the martingale transformed process. In particular, the key lemma used in the purely nonparametric case obviously needs to be modified to account for the multidimensional covariates $X$, as is done in Lemma 5.2 below. We also have to deal with the additional difficulty that the linear part has to be estimated before constructing a test. Moreover, some quantities, such as the conditional variances of the residuals, are more complicated than in the purely nonparametric set-up.

Upon choosing $\Sigma_\xi = 0$ and $\sigma_\eta^2 = 0$, where $\sigma_\eta^2$ is the variance of $\eta$, we see that the proposed test is also applicable in the partial linear regression model with no measurement error. For such a model, Zhu and Ng (2003) have developed a procedure to test the hypothesis $E(Y|X = x, T = t) = \beta' x + g(t)$, for some $\beta$ and $g$, but if we know $X$ is linearly related to the response, this test will be less efficient than ours. Moreover, their test is not asymptotically distribution free.

They propose a variant of wild bootstrap approximation to implement their test. Liang (2006) developed two tests based on a residual-marked empirical process and a linear mixed effect framework for checking linearity of the non-parametric component. Again, because of the complicated limiting distributions, Liang uses bootstrap methodology to implement these tests. In contrast, the transformed marked residual empirical process discussed in this paper converges weakly to a time transformed Brownian motion in uniform metric. Consequently, any test based on a continuous functional of this process is asymptotically distribution free (ADF) and can be implemented at least for moderate to large samples without resorting to a resampling method.

The rest of the paper is organized as follows. The marked residual empirical process and its asymptotic null distribution is discussed in Section 2 under quite broad assumptions. Consistency and asymptotic power against $n^{-1/2}$-local nonparametric alternatives of the test based on the supremum of this process are discussed in Section 3. Section 4 contains a simulation study, and a Monte Carlo power comparison of the proposed test with the two tests of Liang.

All proofs are deferred to Section 5. In the sequel, $B$ denotes standard Brownian motion on $[0, \infty)$, and for any r.v. $U$, $F_U$ and $f_U$ denote its distribution and density function, respectively.

## 2. Main Results

The first subsection below discusses a test for a simple hypothesis, while testing for $H_0$ is discussed in the next subsection.

### 2.1. Testing for a simple hypothesis

Let $g_0$ be a known real-valued function with $Eg_0^2(T) < \infty$. Consider the simple hypothesis

$$H_{10} : g(t) = g_0(t), \quad \forall\, t \in \mathbb{R}, \quad \text{versus} \quad H_{11} : H_{10} \text{ is not true.}$$

The discussion of this simple case sheds some light on the more general hypothesis $H_0$ to be discussed later.

Let $\mu(s) := E(g(T)|S = s)$, $s \in \mathbb{R}$. Under the model assumptions, $E(Y|Z = z, S = s) = E(\beta' X + g(T) + \varepsilon|Z = z, S = s) = \beta' z + \mu(s)$. We are thus led to the calibrated partial linear regression model $Y = \beta' Z + \mu(S) + \zeta$, where the error variable $\zeta$ satisfies $E(\zeta|Z = z, S = s) = 0$, and hence is uncorrelated with $(Z, S)$. This technique of transforming the regression function of $Y$ on $(X, T)$ to the regression function of $Y$ on $(Z, S)$ is known as regression calibration, and is widely used when dealing with measurement error models, see, e.g., Carroll, Ruppert, and Stefanski (1995).

Let $\mu_0(s) := E(g_0(T)|S = s)$, $s \in \mathbb{R}$. Since $f_\eta$ is known, $\mu_0(s) = E(g_0(S + \eta)|S = s) = \int g_0(s + v)f_\eta(v)dv$ is known. Thus, a test of $H_{10}$ can be carried out by testing

$$H_{20}: \quad \mu(s) = \mu_0(s), \quad \forall s \in \mathbb{R}, \qquad \text{versus} \quad H_{21} : H_{20} \quad \text{is not true.}$$

The two hypotheses $H_{10}$ and $H_{20}$ are not equivalent in general. Clearly, $H_{10}$ implies $H_{20}$. The converse is not true in general, since $\int g_0(v)f_\eta(v - s)dv \equiv \int g_1(v)f_\eta(v - s)dv$ need not imply $g_0 = g_1$, but if the family of densities $\{f_\eta(\cdot - s), s \in \mathbb{R}\}$ is complete, then $g_0 = g_1$ almost everywhere.

To proceed further, let $\tau_0^2(s) = E[(g_0(T) - \mu_0(S))^2|S = s]$ and $\sigma_\varepsilon^2 := E(\varepsilon^2)$. The conditional variance of $\zeta$, given $(Z, S)$, is

$$\sigma_{\zeta,\beta}^2(z, s) := E(\zeta^2|Z = z, S = s) = \sigma_\varepsilon^2 + \beta^T\Sigma_\xi\beta + \tau_0^2(s).$$

Since $\sigma_{\zeta,\beta}^2(z, s)$ does not depend on $z$, write $\sigma_{\zeta,\beta}^2(s)$ for $\sigma_{\zeta,\beta}^2(z, s)$. Extend the definitions of $\mu_0$, $\tau_0^2$ to $\bar{\mathbb{R}} := [-\infty, \infty]$ by assigning the value 0 to these functions at $\pm\infty$. This convention will apply also to the analogs of these functions in the sequel. Note that $\sigma_{\zeta,\beta}^2(s) \geq \sigma_\varepsilon^2 > 0$ for all $s \in \bar{\mathbb{R}}$.

Under $H_{20}$ one has the regression model where the 'response' variable is $Y - \beta' Z$, the design variable is $S$, and the error $\zeta$ is uncorrelated with $S$ and heteroscedastic with the conditional variance function $\sigma_{\zeta,\beta}^2(S)$. Thus if $\beta$ were known, one could adapt the STZ testing procedure to this regression set-up. In the more realistic situation where $\beta$ is unknown, this procedure is modified as follows.

Let $\hat{\beta}_n$ be a $n^{1/2}$-consistent estimator of $\beta$ under $H_{10}$, $\zeta_i = Y_i - \beta' Z_i - \mu_0(S_i)$, and $\hat{\zeta}_i = Y_i - \hat{\beta}_n' Z_i - \mu_0(S_i)$. Because $E(\zeta^2) = \sigma_\varepsilon^2 + \beta^T\Sigma_\xi\beta + E\tau_0^2(S)$, consistent estimators of $\sigma_\varepsilon^2$ and $\sigma_{\zeta,\beta}^2(s)$ are given, respectively, by

$$\hat{\sigma}_{2\varepsilon}^2 = \Big|\frac{1}{n}\sum_{i=1}^n \hat{\zeta}_i^2 - \hat{\beta}_n'\Sigma_\xi\hat{\beta}_n - \frac{1}{n}\sum_{i=1}^n \tau_0^2(S_i)\Big|, \qquad \hat{\sigma}_2^2(s) = \hat{\sigma}_{2\varepsilon}^2 + \hat{\beta}_n^T\Sigma_\xi\hat{\beta}_n + \tau_0^2(s).$$

Tests of $H_{20}$ can be based on the marked residual process

$$W_{2n}(s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\hat{\zeta}_i}{\hat{\sigma}_2(S_i)} I(S_i \leq s), \qquad s \in \bar{\mathbb{R}}.$$

Tests of lack-of-fit based on analogs of this process have a long history beginning with von Neumann (1941).See An and Cheng (1991), Hart (1997), STZ, and Khmaladze and Koul (2004) for more on basing tests of lack-of-fit on these types of marked empirical processes.

The asymptotic null distribution of the process $\{W_{2n}(s), s \in \bar{\mathbb{R}}\}$ generally depends on the estimator $\hat{\beta}_n$ and the joint d.f. of $(Z, S)$, and hence is not known.

We next describe a transform of this process that converges to a time transformed Brownian motion. Because of the known parametric structure of the error variance $\sigma^2_{\zeta,\beta}(s)$, unlike in STZ, we do not use the split sample technique to construct a consistent estimator of $\sigma^2_{\zeta,\beta}(s)$.

Let $F_{Z,S}$ denote the joint d.f. of $(Z,S)$, and set

$$e_i = \frac{\zeta_i}{\sigma_{\zeta,\beta}(S_i)}, \quad e = \frac{\zeta}{\sigma_{\zeta,\beta}(S)}, \quad C_s = E\frac{ZZ'I(S \geq s)}{\sigma^2_{\zeta,\beta}(S)}, \quad s \in \mathbb{R}.$$

Assume $C_s$ is positive definite for all $s \in \mathbb{R}$ and let

$$K(s) := e\Big[I(S \leq s) - \int_{y \leq s}\int \frac{x'}{\sigma_{\zeta,\beta}(y)} C_y^{-1} I(S \geq y) dF_{Z,S}(x,y)\frac{Z}{\sigma_{\zeta,\beta}(S)}\Big].$$

One can verify that $EK(s) \equiv 0$ and $EK(s)K(t) = F_S(s \wedge t)$, $s,t \in \mathbb{R}$. Let $K_i(s)$ denote $K(s)$ when the r.v.'s $e$, $Z$, and $S$ are replaced by $e_i$, $Z_i$ and $S_i$, respectively. Define $\mathcal{W}_{\beta,F_{Z,S}}(s) = n^{-1/2}\sum_{i=1}^n K_i(s)$. From the classical CLT, we readily obtain that all finite dimensional distributions of $\mathcal{W}_{\beta,F_{Z,S}}$ converge weakly to those of $B \circ F_S$. But this transform is not useful as it depends on the unknown $\beta$ and $F_{Z,S}$.

Let $\hat{F}_{Z,S}$ denote the empirical d.f. of $(Z_i,S_i), 1 \leq i \leq n$, $\hat{C}_y$ denote the $C_y$ with $F_{Z,S}$ and $\sigma_{\zeta,\beta}$ replaced by $\hat{F}_{Z,S}$ and $\hat{\sigma}_2$, and let $\hat{K}_i$ denote the transform $K$ when $e$, $Z$, $S$, $C_y$ and $F_{Z,S}$ are replaced by $\hat{e}_i := \hat{\zeta}_i/\hat{\sigma}_2(S_i)$, $Z_i$, $S_i$, $\hat{C}_y$ and $\hat{F}_{Z,S}$, respectively. Then, the transformed process on which the proposed test is based takes the form $\hat{\mathcal{W}}_n(s) = n^{-1/2}\sum_{i=1}^n \hat{K}_i(s)$. Under some regularity conditions, we can show that $\hat{\mathcal{W}}_n \Rightarrow B \circ F_S$ in $D([-\infty,s])$, for every $s < \infty$, and in uniform metric. Details of the proof here are similar to those given for the general case in the next section and hence are omitted. A computational formula for $\hat{\mathcal{W}}_n(s)$ is

$$\begin{aligned}\hat{\mathcal{W}}_n(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{e}_i\Big\{I(S_i \leq s) - \frac{1}{n}\sum_{j=1}^n \frac{Z_j'}{\hat{\sigma}_2(S_j)}\hat{C}_{S_j}^{-1}\, I(S_j \leq s \wedge S_i)\,\frac{Z_i}{\hat{\sigma}_2(S_i)}\Big\},\\ \hat{C}_{S_j} &:= \frac{1}{n}\sum_{k=1}^n \frac{Z_k Z_k'}{\hat{\sigma}_2^2(S_k)} I(S_k \geq S_j).\end{aligned}$$

### 2.2. Tests for $H_0$

Let $\mu_\gamma(s) = E(g_\gamma(T)|S=s) = \int g_\gamma(s+v)f_\eta(v)dv$. Under $H_0$, by regression calibration, we obtain the calibrated partial linear regression model $Y = \beta'Z + \mu_\gamma(S) + \zeta$, where $\zeta$ is still used to denote the regression error. Thus to test $H_0$ vs. $H_1$, it suffices to test the hypothesis

$$H_{30}: \mu(S) = \mu_\gamma(S) \quad \text{for some } \gamma \in \Gamma, \quad \text{versus} \quad H_{31}: H_{30} \ \text{ is not true.}$$

Let $\beta_0$ denote the true value of $\beta$, $\gamma_0$ denote the true value of $\gamma$ under $H_0$, assumed to be in the interior of $\Gamma$, and let $\theta' = (\beta_0', \gamma_0')$. To proceed further, we need the following additional assumptions.

(e) $$E\varepsilon^4 + E\|\xi\|^4 + E\|Z\|^4 + Eg_{\gamma_0}^4(T) < \infty.$$

(g1) For some positive continuous function $r(t)$ with $Er^4(T) < \infty$,

$$|g_{\gamma_1}(t) - g_{\gamma_2}(t)| \le \|\gamma_1 - \gamma_2\| r(t), \quad \forall\, \gamma_1, \gamma_2 \in \Gamma, t \in \mathbb{R}.$$

(g2) For every $t \in \mathbb{R}$, $g_\gamma(t)$ is differentiable in $\gamma$ in a neighborhood of $\gamma_0$ with the vector of derivatives $\dot{g}_\gamma(t)$, such that $E\|\dot{g}_{\gamma_0}(T)\|^2 < \infty$, and for every $0 < k < \infty$,

$$\sup_{t\in\mathbb{R}, \sqrt{n}\|\gamma-\gamma_0\|\le k} \sqrt{n}|g_\gamma(t) - g_{\gamma_0}(t) - (\gamma - \gamma_0)'\dot{g}_{\gamma_0}(t)| = o(1).$$

(g3) Let $\dot{\mu}_\gamma(s) := \int \dot{g}_\gamma(s+y) f_\eta(y)dy$. For some $q \times q$ square matrix $\ddot{\mu}_{\gamma_0}(s)$ and a nonnegative function $k_{\gamma_0}(s)$, both measurable in the $s$ coordinate, the following hold: $E\|\ddot{\mu}_{\gamma_0}(S)\|^2 < \infty$, $E\|\ddot{\mu}_{\gamma_0}(S)\|\|\dot{\mu}_{\gamma_0}(S)\|^j < \infty$, $E\|\dot{\mu}_{\gamma_0}(S)\|^j k_{\gamma_0}(S) < \infty$, $j = 0, 1$, and for all $\delta > 0$, there exists an $\eta > 0$ such that $\|\gamma - \gamma_0\| \le \eta$ implies

$$\|\dot{\mu}_\gamma(s) - \dot{\mu}_{\gamma_0}(s) - \ddot{\mu}_{\gamma_0}(s)(\gamma - \gamma_0)\| \le \delta k_{\gamma_0}(s)\|\gamma - \gamma_0\|, \text{ a.s. } (F_S).$$

(m) $E\|\dot{\mu}_{\gamma_0}(S)\|^2 < \infty$, and with $\ell(z, s) := (z, \dot{\mu}_{\gamma_0}(s))'/\sigma_{\zeta,\theta}(s)$,

$$M_y := E\ell(Z, S)\ell(Z, S)'I(S \ge y) \text{ is positive definite for all } y \in \mathbb{R}.$$

The moment condition (e) is needed to bound some quantities when deriving their asymptotics. Conditions (g1)$-$(g3) require certain smoothness of $g_\gamma$ as a function of $\gamma$. These conditions are satisfied if either $g_\gamma(t)$, as a function of $\gamma$, has bounded second derivative, or the r.v. $T$ has a compact support. Condition (m) is a technical assumption to ensure that certain matrices used in the martingale transformation are invertible.

Now let $\tau_\gamma^2(s) := E[(g_\gamma(T) - \mu_\gamma(S))^2 | S = s]$, $s \in \mathbb{R}$. The analogs of $\tau^2$ and $\sigma_{\zeta,\beta}^2$ of the previous sub-section are, respectively, $\tau_{\gamma_0}^2$ and $\sigma_{\zeta,\theta}^2(s) = \sigma_\varepsilon^2 + \beta_0'\Sigma_\xi\beta_0 + \tau_{\gamma_0}^2(s)$. To estimate them, let $\hat\beta_n$, $\hat\gamma_n$ be any $\sqrt{n}$-consistent estimators for $\beta_0$, $\gamma_0$, under $H_0$, respectively. Let $\widetilde{\zeta}_i := Y_i - \hat\beta_n' Z_i - \mu_{\hat\gamma_n}(S_i)$. Because $\mu_\gamma$ is continuous in $\gamma$ at $\gamma_0$, consistent estimators of $\sigma_\varepsilon^2$ and $\sigma_{\zeta,\theta}^2(s)$ are, respectively,

$$\hat\sigma_{3\varepsilon}^2 = \Big|\frac{1}{n}\sum_{i=1}^n \widetilde{\zeta}_i^2 - \hat\beta_n'\Sigma_\xi\hat\beta_n - \frac{1}{n}\sum_{i=1}^n \tau_{\hat\gamma_n}^2(S_i)\Big|, \qquad \tilde\sigma_3^2(s) = \hat\sigma_{3\varepsilon}^2 + \hat\beta_n'\Sigma_\xi\hat\beta_n + \tau_{\hat\gamma_n}^2(s).$$

Let $\hat{W}_{3n}(s) = (1/\sqrt{n})\sum_{i=1}^{n} \widetilde{\zeta}_i I(S_i \le s)/\tilde{\sigma}_3(S_i)$. As in the simple hypothesis case, lack-of-fit tests based on $\hat{W}_{3n}$ are not ADF, but the ones based on its martingale transform are. To describe this transform, let $\hat{M}_y$ denote the estimate of $M_y$ obtained by the plug-in method where all parameters are replaced by their estimates:

$$\hat{M}_y = \iint_{s\ge y} \begin{pmatrix} zz' & z\dot{\mu}'_{\hat{\gamma}_n}(s) \\ \dot{\mu}_{\hat{\gamma}_n}(s)z' & \dot{\mu}_{\hat{\gamma}_n}(s)\dot{\mu}'_{\hat{\gamma}_n}(s) \end{pmatrix} \frac{1}{\tilde{\sigma}_3^2(s)} d\hat{F}_{Z,S}(z,s).$$

Under assumptions (e), (g1), (g2) and under $H_0$, $\sup_y \|\hat{M}_y - M_y\| = o_p(1)$. Consequently, with arbitrarily large probability, $\hat{M}_y^{-1}$ will exist for all $y < \infty$ and for all sufficiently large $n$. Let $\hat{\ell}(z,s)$ denote the $\ell(z,s)$ where $\gamma_0$ and $\sigma_{\zeta,\theta}$ are replaced by $\hat{\gamma}_n$, and $\tilde{\sigma}_3$, respectively. Define $\widetilde{e}_i := \widetilde{\zeta}_i/\tilde{\sigma}_3(S_i)$, and

$$\tilde{K}_i(s) := \widetilde{e}_i\Big[I(S_i \le s) - \int_{y\le s}\int \hat{\ell}(x,y)'\hat{M}_y^{-1} I(S_i \ge y)\, d\hat{F}_{Z,S}(x,y)\, \hat{\ell}(Z_i,S_i)\Big].$$

The proposed test is to be based on the process

$$\begin{aligned} \mathcal{W}_n(s) &:= n^{-1/2}\sum_{i=1}^{n} \tilde{K}_i(s) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widetilde{e}_i\Big\{I(S_i \le s) - \frac{1}{n}\sum_{j=1}^{n} \hat{\ell}(Z_j,S_j)'\hat{M}_{S_j}^{-1} I(S_i \wedge s \ge S_j)\, \hat{\ell}(Z_i,S_i)\Big\}, \\ \hat{M}_s &:= \frac{1}{n}\sum_{i=1}^{n} \hat{\ell}(Z_i,S_i)\hat{\ell}(Z_i,S_i)' I(S_i \ge s). \end{aligned}$$

The following theorem gives the needed weak convergence result.

**Theorem 2.1.** *Suppose, in addition to* (1.1) *and* $H_0$*, the conditions* (e), (g1)–(g3)*, and* (m) *hold, and* $\hat{\beta}_n$, $\hat{\gamma}_n$ *satisfy*

$$\sqrt{n}\|\hat{\beta}_n - \beta_0\| = O_p(1), \qquad \sqrt{n}\|\hat{\gamma}_n - \gamma_0\| = O_p(1), \qquad (H_0). \tag{2.1}$$

*Then, for every* $s_0 < \infty$, $\mathcal{W}_n \Rightarrow B \circ F_S$, *in* $D([-\infty, s_0])$ *and uniform metric.*

Although many estimation methods will provide estimators of $\beta_0, \gamma_0$ satisfying (2.1), in the Appendix, we show that, under certain mild conditions, the least square estimators of $\beta_0$, $\gamma_0$ satisfy (2.1).

As in STZ, we recommend applying the above result with $s_0$ equal to the 99th percentile of $\hat{F}_S$. Consequently, the test that rejects $H_0$ whenever $\sup_{s\le s_0} |\mathcal{W}_n(s)/0.995| > b_\alpha$ is of the asymptotic size $\alpha$, where $b_\alpha$ is such that $P(\sup_{0\le u\le 1} |B(y)| > b_\alpha) = \alpha$.

## 3. Consistency and Local Power

In this section we show, under some regularity conditions, that the above test is consistent for certain fixed alternatives, and has non-trivial asymptotic power against a large class of $n^{-1/2}$-local nonparametric alternatives.

### 3.1. Consistency

Let $h$ be a known real-valued function with $Eh^2(T) < \infty$ and $h \notin \{g_\gamma; \gamma \in \Gamma\}$. Consider the alternative $H_a : g(t) = h(t)$, for all $t \in \mathbb{R}$. Assume the estimators $\hat{\beta}_n$, $\hat{\gamma}_n$ used in the test statistic now satisfy

$$\sqrt{n}(\hat{\beta}_n - \beta_a) = O_p(1), \quad \sqrt{n}(\hat{\gamma}_n - \gamma_a) = O_p(1) \tag{3.1}$$

for some $\beta_a \in \mathbb{R}^p$, $\gamma_a \in \mathbb{R}^q$, under the alternative $H_a$.

One way to obtain these estimators and parameters is to proceed as follows. Let

$$(\hat{\beta}_n', \hat{\gamma}_n)' := \operatorname{argmin}_{\beta,\gamma} \frac{1}{n}\sum_{i=1}^n [Y_i - \beta' Z_i - \mu_\gamma(S_i)]^2, \tag{3.2}$$

$$(\beta_a', \gamma_a)' := \operatorname{argmin}_{\beta,\gamma} E_a[Y - \beta' Z - \mu_\gamma(S)]^2. \tag{3.3}$$

The Appendix provides some sufficient conditions under which the above $\hat{\beta}_n$, $\hat{\gamma}_n$, $\beta_a$, $\gamma_a$ satisfy (3.1).

Now, define new random variables

$$Y_i^a = \beta_a' X_i + g_{\gamma_a}(T_i) + \varepsilon_i, \quad \hat{e}_i^a = \frac{Y_i^a - \hat{\beta}_n' Z_i - \mu_{\hat{\gamma}_n}(S_i)}{\tilde{\sigma}_3(S_i)}, \quad i = 1, \ldots, n,$$

where $\hat{\beta}_n$, $\hat{\gamma}_n$ used in $\tilde{\sigma}_3(s)$ are as in (3.2). Also, let $\hat{\ell}_i := \hat{\ell}(Z_i, S_i)$, $\hat{M}_i := \hat{M}_{S_i}$, $1 \le i \le n$, where $\hat{M}_y$ is the same as in the previous section with $\hat{\beta}_n$, $\hat{\gamma}_n$ replaced by the ones defined in (3.2). Then, $\widetilde{e}_i = \hat{e}_i^a + [(Y_i - Y_i^a)/\tilde{\sigma}_3(S_i)]$ and $\mathcal{W}_n(s) = \mathcal{W}_n^a(s) + \mathcal{R}_n^a(s)$, where

$$\mathcal{W}_n^a(s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{e}_i^a \Big\{ I(S_i \le s) - \frac{1}{n}\sum_{j=1}^n \hat{\ell}_j' \hat{M}_j^{-1} I(S_i \wedge s \ge S_j)\, \hat{\ell}_i \Big\},$$

$$\mathcal{R}_n^a(s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\tilde{\sigma}_3(S_i)} \Big\{ I(S_i \le s) - \frac{1}{n}\sum_{j=1}^n \hat{\ell}_j' \hat{M}_j^{-1} I(S_i \wedge s \ge S_j)\, \hat{\ell}_i \Big\}.$$

Using (3.1), one can verify that $\sup_{s\in\bar{\mathbb{R}}} |\tilde{\sigma}_3^2(s) - \sigma_a^2(s)| = o_p(1)$, where

$$\sigma_a^2(s) := \Big| \sigma_\varepsilon^2 + E_a[\beta_0' X - \beta_a' Z + h(T) - \mu_{\gamma_a}(S)]^2$$

$$-\beta_a'\Sigma_\xi\beta_a - E_a[g_{\gamma_a}(T)-\mu_{\gamma_a}(S)]^2\Big| + \beta_a'\Sigma_\xi\beta_a + \tau^2_{\gamma_a}(s).$$

In particular, if $X$ and $T$ can be measured without error, then $X = Z$, $T = S$, and $\sigma_a^2(s) = \sigma_\varepsilon^2 + E_a[(\beta_0-\beta_a)'X + h(T) - g_{\gamma_a}(T)]^2$. We can also show that

$$\sup_{1\le i\le n} |\tilde\sigma_3^2(S_i) - \sigma_a^2(S_i)| = o_p(1), \quad (H_a), \tag{3.4}$$

in a similar fashion as showing (5.5) in Section 5.

Define

$$\begin{aligned}
\ell_a(z,s) &:= \begin{pmatrix} z \\ \dot\mu_{\gamma_a}(s)\end{pmatrix}\frac{1}{\sigma_a(s)}, \quad A_s := E(l_a(Z,S)l_a'(Z,S)I(S\ge s)),\\
\mathcal{D}_1(s) &:= E\Big[\frac{(\beta_0-\beta_a)'X + h(T) - g_{\gamma_a}(T)}{\sigma_a(S)} I(S\le s)\Big].\\
\rho(y) &:= E\Big[\frac{(\beta_0-\beta_a)'X + h(T) - g_{\gamma_a}(T)}{\sigma_a(S)} \ell_a(Z,S)I(S\ge y)\Big],\\
\mathcal{D}_2(s) &:= E\Big[\ell_a(Z,S)'A_S^{-1}\rho(S)I(S\le s)\Big].
\end{aligned}$$

The difference between $\mathcal{D}_1$ and $\mathcal{D}_2$ measures the discrepancy between the null and the alternative hypotheses as is reflected in the following theorem.

**Theorem 3.1.** *Suppose the conditions* (e), (g1)−(g3), (m), *and* (3.1) *hold under the alternative hypothesis* $H_a$. *Also assume the alternative is such that* $A_s$ *is positive definite for all* $s < \infty$. *Then, for every* $s_0 < \infty$, *the test that rejects* $H_0$ *whenever* $\sup_{s\le s_0} |\mathcal{W}_n(s)/\sqrt{\hat F_S(s_0)}| > b_\alpha$ *is consistent for* $H_a$, *provided* $\sup_{s\le s_0} |\mathcal{D}_1(s) - \mathcal{D}_2(s)| > 0$.

### 3.2. Local power

Let $\delta$ be a real-valued function with $E\delta^2(T) < \infty$. Here we study the asymptotic power of the proposed test against the local alternatives

$$H_{\text{Loc}} : g_n(t) = g_{\gamma_0}(t) + n^{-1/2}\delta(t), \quad \forall\, t\in\mathbb{R}. \tag{3.5}$$

Under $H_{\text{Loc}}$, the partial linear regression model becomes $Y_i = \beta_0'X_i + g_{\gamma_0}(T_i) + n^{-1/2}\delta(T_i) + \varepsilon_i, i = 1,\ldots,n$. Now assume that the estimators $\hat\beta_n$, $\hat\gamma_n$ used in the test statistic satisfy (2.1) under the local alternative (3.5). This in turn, with a similar argument as in showing (5.5), implies $\sup_{1\le i\le n}\Big|\tilde\sigma_3^2(S_i) - \sigma^2_{\zeta,\theta}(S_i)\Big| = o_p(1)$.

By introducing the notation $Y_i^L = \beta_0'X_i + g_{\gamma_0}(T_i) + \varepsilon_i$,

$$\hat e_i^L = \frac{Y_i^L - \hat\beta_n'Z_i - \mu_{\hat\gamma_n}(S_i)}{\tilde\sigma_3(S_i)}, \quad i = 1,\ldots,n,$$

the standardized residuals $\widetilde{e}_i$ have the decomposition

$$\widetilde{e}_i = \hat{e}_i^L + \frac{Y_i - Y_i^L}{\tilde{\sigma}_3(S_i)} = \hat{e}_i^L + \frac{\delta(T_i)}{\sqrt{n}\tilde{\sigma}_3(S_i)}, \qquad i = 1, \ldots, n.$$

Then $\mathcal{W}_n(s) = \mathcal{W}_n^L(s) + \mathcal{R}_n^L(s)$, where $\mathcal{W}_n^L(s)$ has the same form as $\mathcal{W}_n^a(s)$ with $\hat{e}_i^a$ replaced by $\hat{e}_i^L$, while $\mathcal{R}_n^L(s)$ is obtained by replacing $Y_i - Y_i^a$ by $\delta(T_i)/\sqrt{n}$ in $\mathcal{R}_n^a(s)$. Using these facts, asymptotic distribution of $\mathcal{W}_n$ under $H_{\text{Loc}}$ can be studied by similar arguments as in the case of fixed alternative. Define

$$\mathcal{D}_1^L(s) := E\Big[\frac{\delta(T)}{\sigma_{\zeta,\theta}(S)} I(S \le s)\Big], \quad \rho(y) := E\Big[\frac{\delta(T)}{\sigma_{\zeta,\theta}(S)} \ell_a(Z,S) I(S \ge y)\Big],$$
$$\mathcal{D}_2^L(s) := E\Big[\ell_a(Z,S)' M_S^{-1} \rho(S) I(S \le s)\Big].$$

Since $\delta(t)$ reflects the deviation of the local alternative from the null hypothesis, so $\mathcal{D}_1^L$ and $\mathcal{D}_2^L$ are measures of the difference between these two hypotheses. In fact, we have the following theorem.

**Theorem 3.2.** *Suppose the local alternatives* (3.5) *and the conditions* (e), (m), (g1)$-$(g3), (2.1) *hold. Then, for every $s_0 < \infty$,*

$$\lim_{n\to\infty} P\Big(\sup_{s\le s_0}\Big|\frac{\mathcal{W}_n(s)}{\sqrt{\hat{F}_S(s_0)}}\Big| > b_\alpha\Big) = P\Big(\sup_{s\le s_0}\frac{\Big|B(F_S(s)) + \mathcal{D}_1^L(s) - \mathcal{D}_2^L(s)\Big|}{\sqrt{F_S(s_0)}} > b_\alpha\Big).$$

**Remark 3.1.** *Unknown $f_\eta$ and $\Sigma_\xi$.* The structure of the null hypothesis on $\mu$ and the test statistic assume that the density function $f_\eta$ and the covariance matrix $\Sigma_\xi$ are known. The necessity of this assumption is mainly due to the identifiability issue, but its feasibility comes from the fact that, in some studies, we do have some prior information on $f_\eta$ and $\Sigma_\xi$. For example, in the data example of Delaigle, Hall, and Qiu (2006), the measurement error in the digitized aerial photography can be reasonably modeled as having a bi-weight density function.

If no prior knowledge about these entities is available, but there is a sufficiently large validation data set, larger than the main data set, in which the observations of both the true and the surrogate variables are available, then the conclusions of Theorems 2.1, 3.1 and 3.2 still hold after replacing $f_\eta$, $\Sigma_\xi$ in $\mathcal{W}_n$ by their consistent estimators obtained from the validation data. Currently nothing is known about asymptotic null distribution of this modified test when the size of the validation data set is smaller than or comparable to the size of the main data set.

## 4. Simulation

We first give a computational formula for $\mathcal{W}_n(s)$ that is used in simulation. Let $S_{(i)}, i = 1, \ldots, n$ be the order statistics of $S_i, i = 1, \ldots, n$. Let $\hat{e}_{(i)}$, $Z_{(i)}$, $\hat{\sigma}_{3(i)}$, $\dot{\mu}_{\hat{\gamma}_n(i)}$, be the sorted sequence of $\widetilde{e}_i, Z_i, \tilde{\sigma}_3(S_i)$, and $\dot{\mu}_{\hat{\gamma}_n}(S_i)$ according to $S_i$, $i = 1, \ldots, n$. Let $\hat{\nu}'_{(i)} := (Z'_{(i)}, \dot{\mu}'_{\hat{\gamma}_n(i)})/\hat{\sigma}_{3(i)}$. Then, with $S_{(0)} := -\infty$, $S_{(n+1)} := \infty$,

$$\mathcal{W}_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{l} \hat{e}_{(i)} \Big\{ 1 - \frac{1}{n} \sum_{j=1}^{i} \hat{\nu}'_{(j)} \hat{M}_{(j)}^{-1} \hat{\nu}_{(i)} \Big\}, \quad S_{(l)} \le s < S_{(l+1)}, \ l = 1, \cdots, n,$$

$$\hat{M}_{(j)} = \frac{1}{n} \begin{pmatrix} \displaystyle\sum_{k=j}^{n} \frac{Z_{(k)} Z'_{(k)}}{\hat{\sigma}^2_{3(k)}} & \displaystyle\sum_{k=j}^{n} \frac{Z_{(k)} \dot{\mu}'_{\hat{\gamma}_n(k)}}{\hat{\sigma}^2_{3(k)}} \\ \displaystyle\sum_{k=j}^{n} \frac{\dot{\mu}_{\hat{\gamma}_n(k)} Z'_{(k)}}{\hat{\sigma}^2_{3(k)}} & \displaystyle\sum_{k=j}^{n} \frac{\dot{\mu}_{\hat{\gamma}_n(k)} \dot{\mu}'_{\hat{\gamma}_n(k)}}{\hat{\sigma}^2_{3(k)}} \end{pmatrix}.$$

Let $s_0$ be the 99th percentile of $\hat{F}_S$ and $T_n ::= \sup_{s \le s_0} |\mathcal{W}_n(s)/0.995|$. For an $0 < \alpha < 1$, let $b_\alpha$ denote $(1-\alpha)$th percentile of the distribution of $\sup_{0 \le u \le 1} |B(u)|$. From Khmaladze and Koul (2004) we have $b_\alpha = 2.24241, 2.49771, 2.80705$, for $\alpha = 0.05, 0.025, 0.01$, respectively. In the following simulation, $T_n$ was computed 1,000 times for every sample size, and empirical size and power were computed by using $\#\{T_n \ge b_\alpha\}/1,000$.

**Simulation:** The data were generated from the following models:

$$\begin{aligned} &\text{Model 0: } Y_i = \beta X_i + \gamma T_i + \varepsilon_i, \\ &\text{Model 1: } Y_i = \beta X_i + \gamma T_i + \sin(T_i) + \varepsilon_i, \\ &\text{Model 2: } Y_i = \beta X_i + \gamma T_i + 0.1(T_i^2 - 4.03) + \varepsilon_i. \end{aligned} \tag{4.1}$$

Here the null hypothesis is $H_0 : g(t) = \gamma t$, $t \in \mathbb{R}$. Data from Model 0 were used to study the empirical level, while from Models 1 and 2 were used to study the empirical power of the test. In the simulation, $X = Z + \xi$, $T = S + \eta$, $\varepsilon \sim N(0, 1)$, $Z \sim N(1, 1)$, $\xi \sim N(0, 0.3^2)$, $S \sim N(1, 1)$, $\eta \sim N(0, 0.3^2)$ and $\beta_0 = 1, \gamma_0 = 2$. Under this set up, $m_\theta(z, s) = \beta z + \gamma s$, $\tau^2_\gamma(S) = 0.01\gamma^2$. Hence, $\tilde{\sigma}^2_3(s)$ did not depend on $s$. Also, in Model 2, $T^2 - 4.03$ is orthogonal to $T$. The estimators $\hat{\beta}_n$, $\hat{\gamma}_n$ were chosen to be the least square estimators based on the new regression model $Y = \beta Z + \gamma S + \zeta$. Then $\tilde{\sigma}^2_3(s)$ was simply the mean of squared residuals $Y_i - \hat{\beta}_n Z_i - \hat{\gamma}_n S_i$, not depending on $s$. Table 1 illustrates the simulation results.

To investigate the effects of the magnitude of measurement errors on level and power of the proposed test, we also conducted several additional simulations for different choices of $\sigma^2_\eta$ and $\sigma^2_\xi$. Our results also apply to the case in which $\Sigma^2_\xi = 0$ and $\sigma^2_\eta = 0$, that is, without measurement errors. Given all other distributional

Table 1. Simulation WITH measurement error.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| | Model 0 | 0.041 | 0.037 | 0.039 | 0.046 | 0.049 |
| 0.050 | Model 1 | 0.106 | 0.182 | 0.424 | 0.697 | 0.973 |
| | Model 2 | 0.210 | 0.462 | 0.781 | 0.915 | 0.991 |
| | Model 0 | 0.014 | 0.012 | 0.019 | 0.022 | 0.019 |
| 0.025 | Model 1 | 0.073 | 0.116 | 0.290 | 0.535 | 0.899 |
| | Model 2 | 0.136 | 0.342 | 0.684 | 0.866 | 0.985 |
| | Model 0 | 0.008 | 0.002 | 0.009 | 0.010 | 0.009 |
| 0.010 | Model 1 | 0.033 | 0.054 | 0.168 | 0.340 | 0.729 |
| | Model 2 | 0.074 | 0.201 | 0.559 | 0.770 | 0.968 |

Table 2. Simulation WITH measurement error, $\sigma_\eta^2 = 0.3^2, \sigma_\xi^2 = 0.5^2$.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| | Model 0 | 0.045 | 0.041 | 0.044 | 0.045 | 0.046 |
| 0.050 | Model 1 | 0.100 | 0.173 | 0.403 | 0.661 | 0.956 |
| | Model 2 | 0.200 | 0.425 | 0.742 | 0.886 | 0.986 |
| | Model 0 | 0.016 | 0.012 | 0.021 | 0.022 | 0.025 |
| 0.025 | Model 1 | 0.069 | 0.096 | 0.266 | 0.486 | 0.864 |
| | Model 2 | 0.129 | 0.296 | 0.628 | 0.821 | 0.980 |
| | Model 0 | 0.007 | 0.003 | 0.006 | 0.010 | 0.009 |
| 0.010 | Model 1 | 0.035 | 0.046 | 0.149 | 0.312 | 0.665 |
| | Model 2 | 0.064 | 0.179 | 0.499 | 0.719 | 0.949 |

Table 3. Simulation WITH measurement error, $\sigma_\eta^2 = 0.5^2, \sigma_\xi^2 = 0.5^2$.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| | Model 0 | 0.040 | 0.031 | 0.041 | 0.044 | 0.048 |
| 0.050 | Model 1 | 0.077 | 0.098 | 0.248 | 0.397 | 0.719 |
| | Model 2 | 0.162 | 0.332 | 0.632 | 0.804 | 0.961 |
| | Model 0 | 0.015 | 0.011 | 0.016 | 0.027 | 0.024 |
| 0.025 | Model 1 | 0.049 | 0.053 | 0.143 | 0.256 | 0.534 |
| | Model 2 | 0.103 | 0.226 | 0.514 | 0.712 | 0.930 |
| | Model 0 | 0.006 | 0.003 | 0.008 | 0.006 | 0.012 |
| 0.010 | Model 1 | 0.018 | 0.022 | 0.081 | 0.151 | 0.348 |
| | Model 2 | 0.047 | 0.128 | 0.375 | 0.580 | 0.871 |

assumptions unchanged, we also generated the data from the above model by setting $\xi = 0$ and $\eta = 0$. These simulation results are shown in Tables 2 to 4. From these tables we see that the level of the proposed test is robust against the variation in measurement errors, while power gets smaller, though not too drastically, as variances of measurement errors become larger.

We also conducted a simulation study when $X$ had two dimensions. Similar results were obtained, hence not reported here.

Table 4. Simulation WITHOUT measurement error.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| 0.050 | Model 0 | 0.041 | 0.040 | 0.045 | 0.045 | 0.049 |
| | Model 1 | 0.233 | 0.501 | 0.825 | 0.942 | 0.997 |
| | Model 2 | 0.315 | 0.608 | 0.915 | 0.982 | 0.999 |
| 0.025 | Model 0 | 0.017 | 0.012 | 0.020 | 0.025 | 0.022 |
| | Model 1 | 0.167 | 0.387 | 0.763 | 0.904 | 0.993 |
| | Model 2 | 0.218 | 0.493 | 0.867 | 0.971 | 0.999 |
| 0.010 | Model 0 | 0.005 | 0.004 | 0.005 | 0.015 | 0.007 |
| | Model 1 | 0.101 | 0.278 | 0.666 | 0.862 | 0.982 |
| | Model 2 | 0.123 | 0.371 | 0.786 | 0.936 | 0.999 |

To compare the performance of the $T_n$ test with the two tests studied in Liang (2006), we generated data from a model without measurement error (also used in Liang (2006)), $Y = 1.3X_1 + 0.45X_2 + 2.5T + \varepsilon$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ with $T \sim \text{Uniform}(0, 1)$, and $X$ from one of the following two cases:

**Case 1**: $(X_1, X_2) \sim N_2(\mathbf{0}, \text{diag}(0.3^2, 0.4^2))$; $(X_1, X_2)$ and $T$ are independent;
**Case 2**: $X_j = 0.4T + 0.6U_j$, $j = 1, 2$, and $U_1, U_2$, i.i.d. Uniform(0,1).

We also used the same alternatives as in Liang (2006):

$$g(t) = 2.5t + c[4.25 \exp(-3.25t) - 4 \exp(-6.5t) + 3 \exp(-9.75t)],$$

for $c = 0.2, 0.4, 0.6, 0.8$ and 1. In the simulation, $\sigma_\varepsilon$ was chosen to be 0.1, 0.25, and 0.5. The sample size $n = 100$ and nominal level 0.05 were considered for the purpose of illustration. Figure 1 presents empirical levels and powers of the three tests. The top panel is for Case 1 and the bottom panel for Case 2. In each plot, the solid line is for the $T_n$ test, the dashed line is for Liang's Cramér-von Mises type test and the dotted line for Liang's likelihood ratio test. From the figure, one sees that the likelihood ratio test is the most conservative while the levels of the $T_n$ and Cramér-von Mises type tests are both close to the nominal level 0.05. It is clear that the powers of these three tests increase as the value $c$ becomes larger. One can also see that the $T_n$ test is comparable to the Cramér-von Mises type test, and outperforms the likelihood ratio test at all configurations. Finally, $T_n$ test is relatively easy to compute.

**Remark 4.1.** *Robustness of the test.* In the Berkson model, it is usually assumed that the Berkson error density and/or variance are known. However, one may ask that if the test is somewhat robust against error misspecification. A satisfying answer to this question would require some theoretical arguments, such as finding out the influence function of the test procedure, but we believe this

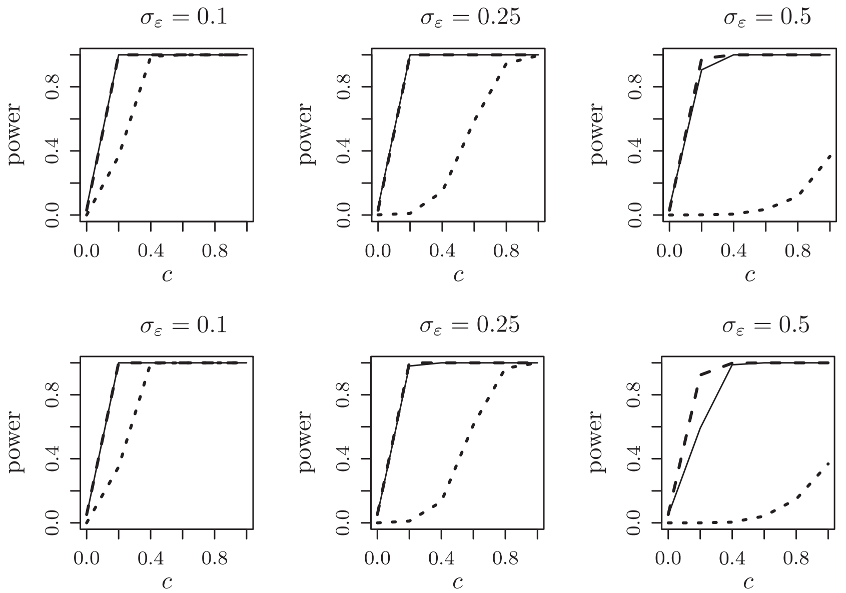


Figure 1. Power curves of three testing procedures.

is beyond the scope of the current paper. Instead, some simulation studies were conducted for the purpose of illustration.

We generated the data from models 0 to 2 in (4.1), except now $\xi$ and $\eta$ are independent Uniform$(-\sqrt{0.27}, \sqrt{0.27})$ r.v.'s; when computing the test statistic, we assumed that $\xi, \eta \sim N(0, 0.3^2)$. Note that these two distributions have the same variance. See Table 5 for the simulation results. Table 6 reports another simulation study in which the data were generated from models 0 to 2 with measurement errors being double exponential with mean 0, and variance $0.3^2$, but $N(0, 0.3^2)$ distribution was used in the test statistic. Again, note that these distributions have the same variance $0.3^2$.

The simulation results suggest that if the true and the misspecified measurement errors distributions were different but have the same variance, the proposed test is reasonably robust.

We also conducted some simulations when the distributions of measurement errors were completely misspecified, and when the distribution type was misspecified but the variance was correctly specified. The results were mixed. At present it is not clear that the test will work if $f_\eta$ is completely misspecified.

Table 5. Uniform distribution misspecified as Normal.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| | Model 0 | 0.040 | 0.042 | 0.044 | 0.046 | 0.040 |
| 0.050 | Model 1 | 0.111 | 0.246 | 0.503 | 0.733 | 0.970 |
| | Model 2 | 0.250 | 0.535 | 0.853 | 0.961 | 0.999 |
| | Model 0 | 0.017 | 0.018 | 0.021 | 0.023 | 0.016 |
| 0.025 | Model 1 | 0.059 | 0.154 | 0.358 | 0.580 | 0.904 |
| | Model 2 | 0.156 | 0.412 | 0.760 | 0.932 | 0.999 |
| | Model 0 | 0.008 | 0.006 | 0.007 | 0.015 | 0.007 |
| 0.010 | Model 1 | 0.028 | 0.090 | 0.232 | 0.409 | 0.755 |
| | Model 2 | 0.074 | 0.288 | 0.661 | 0.871 | 0.995 |

Table 6. Double exponential distribution misspecified as Normal.

| $\alpha$-level | Model\ $n$ | 50 | 100 | 200 | 300 | 500 |
|---|---|---|---|---|---|---|
| | Model 0 | 0.027 | 0.041 | 0.044 | 0.042 | 0.054 |
| 0.050 | Model 1 | 0.092 | 0.250 | 0.503 | 0.731 | 0.972 |
| | Model 2 | 0.269 | 0.496 | 0.862 | 0.967 | 0.999 |
| | Model 0 | 0.010 | 0.02 | 0.025 | 0.014 | 0.024 |
| 0.025 | Model 1 | 0.062 | 0.168 | 0.37 | 0.574 | 0.902 |
| | Model 2 | 0.153 | 0.396 | 0.771 | 0.933 | 0.993 |
| | Model 0 | 0.004 | 0.007 | 0.005 | 0.006 | 0.011 |
| 0.010 | Model 1 | 0.029 | 0.095 | 0.231 | 0.409 | 0.752 |
| | Model 2 | 0.080 | 0.292 | 0.664 | 0.882 | 0.988 |

## 5. Proofs

**Lemma 5.1.** *Suppose $U$ and $V$ are random variables with $E(U|V) = 0$, $0 \le E(U^2) < \infty$. Let $\sigma^2(v) = E(U^2|V = v)$, $L(v) = E\sigma^2(V)I(V \le v)$, $v \in \bar{\mathbb{R}}$. Let $(U_i, V_i)$, $1 \le i \le n$ be i.i.d. copies of $(U, V)$, and*

$$U_n(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i I(V_i \le v), \quad v \in \bar{\mathbb{R}} = [-\infty, \infty].$$

*Assume $L$ to be continuous. Then, $U_n \Rightarrow B \circ L$, in $D(\bar{\mathbb{R}})$ and uniform metric.*

The proof of this lemma uses Theorem 12.6 in Billingsley (1968). Details are similar to those appearing in STZ.

To state the next lemma, let $U$ be a continuous random vector of length $p$, $V$ be a continuous r.v. with d.f. $G$, and let $F(u, v)$ denote their joint d.f. Let $\ell(u, v)$ be a vector of $q$ functions with $E\|\ell(U, V)\|^2 < \infty$. Assume the matrix $C_v := E\ell(U, V)\ell(U, V)'I(V \ge v)$ is positive definite for all $v \in \mathbb{R}$. For a real-valued function $\psi \in L_2(\mathbb{R}, G)$, define the transforms

$$\mathcal{T}_\psi(u,v) := \int_{y\le v}\int \psi(y)\ell(x,y)' C_y^{-1} dF(x,y)\,\ell(u,v),$$
$$\mathcal{K}_\psi(u,v) := \psi(v) - \mathcal{T}_\psi(u,v).$$

The following lemma is an extension of Proposition 4.1 of Khmaladze and Koul (2004) and Lemma 9.1 of Koul (2006). Its proof is similar to that of these results, and hence is not presented.

**Lemma 5.2.** *For the entities defined above,*

$$E\mathcal{K}_\psi(U,V)\ell(U,V)' = 0, \qquad \forall\, \psi \in L_2(\mathbb{R}, F), \tag{5.1}$$
$$E\mathcal{K}_{\psi_1}(U,V)\mathcal{K}_{\psi_2}(U,V) = E\psi_1(V)\psi_2(V), \quad \forall\, \psi_1,\psi_2 \in L_2(\mathbb{R}, F). \tag{5.2}$$

**Remark 5.2** Let $\xi$ be a r.v. such that $E(\xi|U,V) = 0$, $E\xi^2 < \infty$, $\tau^2(u,v) := E(\xi^2|U=u, V=v) > 0$, for all $u,v$. Then the covariance function of the process

$$W_\psi(\xi,U,V) := \Big[\frac{\xi}{\tau(U,V)}\Big]\{\psi(V) - \mathcal{T}_\psi(U,V)\},$$

as a process in $\psi \in L_2(\mathbb{R}, G)$, is that of $B_\psi(G)$, where $B_\psi$ is a Brownian motion in $\psi$. Hence, if $(\xi_i, U_i, V_i)$, $1 \le i \le n$, are i.i.d. copies of $(\xi, U, V)$, then by the classical CLT, the finite dimensional distributions of $n^{-1/2}\sum_{i=1}^n W_\psi(\xi_i, U_i, V_i)$, as $\psi$ varies, converge weakly to those of $B_\psi(G)$.

To prove Theorem 2.1, recall that $\theta := (\beta_0', \gamma_0')'$, $e_i = \zeta_i/\sigma_{\zeta,\theta}(S_i)$, and let

$$\mathcal{W}_{\theta,F_{Z,S}}(s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\Big\{I(S_i \le s) - \int_{y\le s}\int \ell(x,y)' M_y^{-1} I(S_i \ge y)\, dF_{Z,S}(x,y)\,\ell(Z_i,S_i)\Big\}.$$

**Proof of Theorem 2.1.** The proof consists of the following two steps.

(a) For every $s_0 < \infty$, $\mathcal{W}_{\theta,F_{Z,S}} \Rightarrow B\circ F_S$, in $D([-\infty, s_0])$ and in uniform metric.

(b) $\sup_{s\le s_0}|\mathcal{W}_n(s) - \mathcal{W}_{\theta,F_{Z,S}}(s)| = o_p(1)$, $\quad (H_0)$.

**Proof of Part (a).** Applying Lemma 5.2 and Remark 5.2 to $\xi = e = \zeta/\sigma_{\zeta,\theta}$, $U = Z$, $V = S$, and to the family of indicator functions $\psi(v) = I(v \le x)$, $x \in \mathbb{R}$, we readily obtain that all finite dimensional distributions of $\mathcal{W}_{\theta,F_{Z,S}}$ converge weakly to those of $B\circ F_S$. Thus, the claim (a) would follow if we prove the tightness of $\mathcal{W}_{\theta,F_{Z,S}}$. Toward this, let

$$W_{3n}(s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i I(S_i \le s),$$

$$Q_n(s) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \int_{y \le s} \int \ell(x,y)' M_y^{-1} \ell(Z_i, S_i) I(S_i \ge y) dF_{Z,S}(x,y), \quad s \in \mathbb{R}.$$

Then we can rewrite $\mathcal{W}_{\theta,F_{Z,S}}(s) := W_{3n}(s) - Q_n(s)$. Lemma 5.1 applied to $U = e$, $V = S$, yields the tightness of $W_{3n}(s)$, $s \in \bar{\mathbb{R}}$, in uniform metric.

To prove the tightness of the $Q_n$ process, let $\varphi(s) := \int_{y \le s} \int \|\ell(z,y)' M_y^{-1}\| dF_{Z,S}(z,y), s \in \mathbb{R}$. Note that $\varphi$ is nondecreasing, non-negative and because of assumption (m), $\varphi(s) < \infty$ for all $s \in \mathbb{R}$. Moreover, $E(e|Z,S) = 0$, $E(e^2|Z,S) = 1$, and $\|M\|_\infty := \sup_{s\in\mathbb{R}} \|M_s\| \le E\|\ell(Z,S)\|^2 < \infty$, imply $E[Q_n(t) - Q_n(s)]^2 \le \|M\|_\infty [\varphi(t) - \varphi(s)]^2, \forall\, s \le t$. This bound, together with Theorem 15.6 of Billingsley (1968), imply that for every $s_0 < \infty$, $Q_n(s)$ is tight in uniform metric on $(-\infty, s_0]$. This completes the proof of part (a).

**Proof of Part (b).** Let $\ell_i := \ell(Z_i, S_i)$, $\hat{\ell}_i := \hat{\ell}(Z_i, S_i)$, and let

$$\widetilde{U}_n(y) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{e}_i\, \hat{\ell}_i\, I(S_i \ge y), \quad U_n(y) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i\, \ell_i\, I(S_i \ge y).$$

Then $\mathcal{W}_n(s)$ and $\mathcal{W}_{\theta,F_{Z,S}}(s)$ can be written as

$$\mathcal{W}_n(s) = \hat{W}_{3n}(s) - \int_{y \le s} \int \hat{\ell}(x,y)' \hat{M}_y^{-1} \widetilde{U}_n(y) d\hat{F}_{Z,S}(x,y), \tag{5.3}$$

$$\mathcal{W}_{\theta,F_{Z,S}}(s) = W_{3n}(s) - \int_{y \le s} \int \ell(x,y)' M_y^{-1} U_n(y) dF_{Z,S}(x,y). \tag{5.4}$$

Let $b_n := \hat{\beta}_n - \beta_0$. We can rewrite $\hat{W}_{3n}$ as the sum of six terms:

$$I_{n1}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i I(S_i \le s), \quad I_{n2}(s) = b_n' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i}{\sigma_{\zeta,\theta}(S_i)} I(S_i \le s),$$

$$I_{n3}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mu_{\hat{\gamma}_n}(S_i) - \mu_{\gamma_0}(S_i)}{\sigma_{\zeta,\theta}(S_i)} I(S_i \le s),$$

$$I_{n4}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \Big( \frac{\sigma_{\zeta,\theta}(S_i)}{\tilde{\sigma}_3(S_i)} - 1 \Big) I(S_i \le s),$$

$$I_{n5}(s) = b_n' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i}{\sigma_{\zeta,\theta}(S_i)} \Big( \frac{\sigma_{\zeta,\theta}(S_i)}{\tilde{\sigma}_3(S_i)} - 1 \Big) I(S_i \le s),$$

$$I_{n6}(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\mu_{\hat{\gamma}_n}(S_i) - \mu_{\gamma_0}(S_i)}{\sigma_{\zeta,\theta}(S_i)} \Big( \frac{\sigma_{\zeta,\theta}(S_i)}{\tilde{\sigma}_3(S_i)} - 1 \Big) I(S_i \le s).$$

The term $I_{n1}$ is simply $W_{3n}$. We can show that $\sup_{s\in\bar{\mathbb{R}}} |I_{nj}(s)| = o_p(1)$, $j = 4, 5, 6$. Because most of the arguments are similar, for the sake of brevity, we give details only for the case $j = 4$. First we show that

$$\max_{1\le i\le n} |\tilde\sigma_3^2(S_i) - \sigma^2_{\zeta,\theta}(S_i)| = o_p(1). \tag{5.5}$$

By definition, $\tilde\sigma_3^2(S_i) - \sigma^2_{\zeta,\theta}(S_i) = \hat\sigma^2_{3\varepsilon} + \hat\beta_n'\Sigma_\zeta\hat\beta_n + \tau^2_{\hat\gamma_n}(S_i) - \sigma^2_\varepsilon - \beta_0'\Sigma_\zeta\beta_0 - \tau^2_{\gamma_0}(S_i)$. Since $\hat\sigma^2_{3\varepsilon} - \sigma^2_\varepsilon = o_p(1)$, $\hat\beta'\Sigma_\zeta\hat\beta_n - \beta_0'\Sigma_\zeta\beta_0 = o_p(1)$, it suffices to show that

$$\max_{1\le i\le n} |\tau^2_{\hat\gamma_n}(S_i) - \tau^2_{\gamma_0}(S_i)| = o_p(1). \tag{5.6}$$

Note that for all $s\in\mathbb{R}$,

$$\begin{aligned}
|\tau^2_{\hat\gamma_n}(s) - \tau^2_{\gamma_0}(s)| \le{}& \Big|\int [g^2_{\hat\gamma}(s+y) - g^2_{\gamma_0}(s+y)] f_\eta(y)dy\Big| \\
&+ \Big|\Big[\int g_{\hat\gamma_n}(s+y)f_\eta(y)dy\Big]^2 - \Big[\int g_{\gamma_0}(s+y)f_\eta(y)dy\Big]^2\Big| \\
={}& A_n(s) + B_n(s), \quad \text{say.}
\end{aligned}$$

Let $\delta_n := \hat\gamma_n - \gamma_0$. By condition (g1), for all $s\in\mathbb{R}$,

$$\int \Big(g_{\hat\gamma_n}(s+y) - g_{\gamma_0}(s+y)\Big)^2 f_\eta(y)dy \le \|\delta_n\|^2 \int r^2(s+y)f_\eta(y)dy.$$

The assumption that $Er^4(T) < \infty$ implies $E(\int r^2(S+y)f_\eta(y)dy)^2 < \infty$. Hence, $\max_{1\le i\le n}\int r^2(S_i+y)f_\eta(y)dy = o_p(\sqrt{n})$ and, in view of (2.1),

$$\max_{1\le i\le n} n^{1/2}\int \Big(g_{\hat\gamma_n}(S_i+y) - g_{\gamma_0}(S_i+y)\Big)^2 f_\eta(y)dy = o_p(1).$$

This fact and a routine argument now shows that $\max_{1\le i\le n} A_n(S_i) = o_p(1)$, $\max_{1\le i\le n} B_n(S_i) = o_p(1)$, thereby completing the proof of (5.6), and hence that of (5.5).

Let $D_n := \sigma^2_\varepsilon - \hat\sigma^2_{3\varepsilon} + \beta_0'\Sigma_\zeta\beta - \hat\beta_n'\Sigma_\zeta\hat\beta_n'$. Then $I_{n4}(s)$ can be written as the sum of two terms:

$$\begin{aligned}
I_{n41}(s) &= D_n\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\Big[\frac{1}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i)+\sigma_{\zeta,\theta}(S_i))}\Big] I(S_i\le s),\\
I_{n42}(s) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\Big[\frac{\tau_0^2(S_i) - \tau^2_{\hat\gamma_n}(S_i)}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i)+\sigma_{\zeta,\theta}(S_i))}\Big] I(S_i\le s).
\end{aligned}$$

Subtracting and adding $1/2\sigma^2_{\zeta,\theta}(S_i)$, $I_{n41}$ can be written as the sum:

$$\sqrt{n}D_n\cdot\frac{1}{n}\sum_{i=1}^n e_i\Big[\frac{1}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i)+\sigma_{\zeta,\theta}(S_i))} - \frac{1}{2\sigma^2_{\zeta,\theta}(S_i)}\Big] I(S_i\le s)$$

$$+D_n \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{e_i}{2\sigma^2_{\zeta,\theta}(S_i)} I(S_i \le s).$$

By (5.5), and the fact that $\sigma^2_{\zeta,\theta} \ge \sigma^2_\varepsilon > 0$,

$$\max_{1\le i\le n} \Big| \frac{1}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i) + \sigma_{\zeta,\theta}(S_i))} - \frac{1}{2\sigma^2_{\zeta,\theta}(S_i)} \Big| = o_p(1). \tag{5.7}$$

By Lemma 5.1, we obtain

$$\sup_{s\in\bar{\mathbb{R}}} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{e_i}{2\sigma^2_{\zeta,\theta}(S_i)} I(S_i \le s) \Big| = O_p(1).$$

These facts, together with $\sqrt{n}(\sigma^2_\varepsilon - \hat\sigma^2_{3\varepsilon}) = O_p(1)$, $\sqrt{n}(\beta_0'\Sigma_\zeta\beta_0 - \hat\beta_n'\Sigma_\zeta\hat\beta_n') = O_p(1)$, and $\sum_{i=1}^n |e_i|/n = O_p(1)$, imply that $\sup_{s\in\bar{\mathbb{R}}} I_{n41}(s) = o_p(1)$.

Next, we sketch an argument for proving $\sup_{s\in\bar{\mathbb{R}}} I_{n42}(s) = o_p(1)$. For this purpose, we need a refined analysis of $\tau^2_{\hat\gamma_n}(S_i) - \tau^2_0(S_i)$. By definition,

$$\begin{aligned}
\tau^2_{\hat\gamma_n}(S_i) - \tau^2_0(S_i) = & \int [g_{\hat\gamma_n}(S_i + y) - g_{\gamma_0}(S_i + y)]^2 f_\eta(y)dy \\
& + 2\int g_{\gamma_0}(S_i + y)[g_{\hat\gamma_n}(S_i + y) - g_{\gamma_0}(S_i + y)] f_\eta(y)dy \\
& - \Big[ \int [g_{\hat\gamma_n}(S_i + y) - g_{\gamma_0}(S_i + y)] f_\eta(y)dy \Big]^2 \\
& - 2\int g_{\gamma_0}(S_i + y) f_\eta(y)dy \int [g_{\hat\gamma_n}(S_i + y) - g_{\gamma_0}(S_i + y)] f_\eta(y)dy.
\end{aligned}$$

Let $\Delta_{ni}(y) = g_{\hat\gamma_n}(S_i+y) - g_{\gamma_0}(S_i+y) - \delta_n'\dot g_{\gamma_0}(S_i+y)$. Subtracting and adding $\delta_n'\dot g_{\gamma_0}(S_i + \eta)$ to the difference $g_{\hat\gamma_n}(S_i + y) - g_{\gamma_0}(S_i + y)$ in the above integrals, and expanding various quadratics yields that $\tau^2_{\hat\gamma_n}(S_i) - \tau^2_0(S_i)$ is the sum of ten terms:

$$\begin{aligned}
& A_{i,1} := \int \Delta^2_{ni}(y) f_\eta(y)dy, \quad A_{i,2} = 2\delta_n' \int \dot g_{\gamma_0}(S_i + y)\Delta_{ni}(y) f_\eta(y)dy, \\
& A_{i,3} := \delta_n' \int \dot g_{\gamma_0}(S_i + y)\dot g_{\gamma_0}'(S_i + y) f_\eta(y)dy\, \delta_n, \\
& A_{i,4} := 2\int g_{\gamma_0}(S_i + y)\Delta_{ni}(y) f_\eta(y)dy, \\
& A_{i,5} := 2\delta_n' \int g_{\gamma_0}(S_i + y)\dot g_{\gamma_0}(S_i + y) f_\eta(y)dy, \quad A_{i,6} := -\Big[ \int \Delta_{ni}(y) f_\eta(y)dy \Big]^2, \\
& A_{i,7} := -2\delta_n' \int g_{\gamma_0}(S_i + y) f_\eta(y)dy \cdot \int \Delta_{ni}(y) f_\eta(y)dy,
\end{aligned}$$

$$A_{i,8} := -\delta_n' \int \dot g_{\gamma_0}(S_i+y) f_\eta(y)dy \cdot \int \dot g_{\gamma_0}'(S_i+\eta) f_\eta(y)dy\,\delta_n,$$
$$A_{i,9} := -2\int g_{\gamma_0}(S_i+y) f_\eta(y)dy \cdot \int \Delta_{ni}(y) f_\eta(y)dy,$$
$$A_{i,10} := -\delta_n' \int g_{\gamma_0}(S_i+y) f_\eta(y)dy \cdot \int \dot g_{\gamma_0}(S_i+\eta) f_\eta(y)dy.$$

From (g2), (g3), (2.1), for $j = 1, \dots, 10$, we can obtain that

$$\sup_{s\in\bar{\mathbb{R}}}\Big|\frac{1}{\sqrt n}\sum_{i=1}^n \frac{e_i A_{i,j}}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i)+\sigma_{\zeta,\theta}(S_i))} I(S_i\le s)\Big| = o_p(1). \tag{5.8}$$

These imply that $\sup_{s\in\bar{\mathbb{R}}} I_{n42}(s) = o_p(1)$. Here we present the proof of (5.8) for $j = 4$ only, as the proof for the other cases is similar.

The l.h.s. of (5.8) for $j = 4$ is bounded above by

$$2\sup_{1\le i\le n,\, y\in\mathbb{R}} n^{1/2}|\Delta_{ni}(y)|\cdot\frac{1}{n}\sum_{i=1}^n\frac{\Big|e_i\int g_{\gamma_0}(S_i+y)f_\eta(y)dy\Big|}{\tilde\sigma_3(S_i)(\tilde\sigma_3(S_i)+\sigma_{\zeta,\theta}(S_i))}.$$

By (g2) and (2.1), the first factor of this bound is $o_p(1)$. The square integrability of $e_i$, $g_{\gamma_0}(T_i)$, together with (5.7) and the Law of Large Numbers, imply that the second factor of the above bound is $O_p(1)$. Hence (5.8) holds for $j = 4$.

In summary, we obtain

$$\sup_{s\in\bar{\mathbb{R}}}\Big|\hat W_{3n}(s) - W_{3n}(s) + E\Big[\ell(Z,S)'I(S\le s)\Big]\sqrt n\binom{b_n}{\delta_n}\Big| = o_p(1). \tag{5.9}$$

Next, consider the difference $\widetilde U_n(y) - U_n(y)$. Let $\alpha_i := \mu_{\hat\gamma_n}(S_i) - \mu_{\gamma_0}(S_i)$, and $\dot\alpha_i := \dot\mu_{\hat\gamma_n}(S_i) - \dot\mu_{\gamma_0}(S_i)$. By replacing $1/\tilde\sigma_3^2(S_i)$ by $[\sigma^2_{\zeta,\theta}(S_i)/\tilde\sigma_3^2(S_i) - 1 + 1]/\sigma^2_{\zeta,\theta}(S_i)$, subtracting and adding $\beta_0$ from $\hat\beta_n$, $\mu_{\gamma_0}(S_i)$ from $\mu_{\hat\gamma_n}(S_i)$, and $\dot\mu_{\gamma_0}(S_i)$ from $\dot\mu_{\hat\gamma_n}(S_i)$, $\widetilde U_n(y) - U_n(y)$ can be rewritten as the sum of $D_{n1}(y)$, $D_{n2}(y)$ and a remainder term $R_n(y)$, where

$$D_{n1}(y) = -b_n'\frac{1}{\sqrt n}\sum_{i=1}^n\frac{Z_i}{\sigma_{\zeta,\theta}(S_i)}\,\ell_i\, I(S_i\ge y),$$
$$D_{n2}(y) = -\frac{1}{\sqrt n}\sum_{i=1}^n\frac{\alpha_i}{\sigma_{\zeta,\theta}(S_i)}\,\ell_i\, I(S_i\ge y).$$

By conditions (g2), (g3), (m), and (2.1), one can show $\sup_{y\in\bar{\mathbb{R}}}|R_n(y)| = o_p(1)$.

Subtract and add $\delta_n'\dot\mu_{\gamma_0}(S_i)$ to $\mu_{\hat\gamma_n}(S_i) - \mu_{\gamma_0}(S_i)$, to rewrite $-D_{n2}$ as the sum

$$\frac{1}{\sqrt n}\sum_{i=1}^n\frac{\mu_{\hat\gamma_n}(S_i)-\mu_{\gamma_0}(S_i)-\delta_n'\dot\mu_{\gamma_0}(S_i)}{\sigma^2_{\zeta,\theta}(S_i)}\,\ell_i\, I(S_i\ge y) + \frac{\delta_n'}{\sqrt n}\sum_{i=1}^n\frac{\dot\mu_{\gamma_0}(S_i)}{\sigma^2_{\zeta,\theta}(S_i)}\ell_i\, I(S_i\ge y).$$

In view of (g2), the first term of this sum is bounded from the above by

$$\sqrt{n}\max_{1\le i\le n}|\mu_{\hat\gamma_n}(S_i)-\mu_{\gamma_0}(S_i)-\delta_n'\dot\mu_{\gamma_0}(S_i)|\frac{1}{n\sigma_\varepsilon^2}\sum_{i=1}^n\|\ell_i\|=o_p(1).$$

Hence, with $\theta_n := \sqrt{n}(b_n', \delta_n')'$, $\sup_{y\le s_0}\left\|\widetilde{U}_n(y)-U_n(y)+\widetilde{M}_y\theta_n\right\|=o_p(1)$, where

$$\widetilde{M}_y=\begin{pmatrix}\displaystyle\frac{1}{n}\sum_{k=1}^n\frac{I(S_k\ge y)Z_kZ_k'}{\sigma^2_{\zeta,\theta}(S_k)} & \displaystyle\frac{1}{n}\sum_{k=1}^n\frac{I(S_k\ge y)Z_k\dot\mu'_{\gamma_0}(S_k)}{\sigma^2_{\zeta,\theta}(S_k)}\\[2ex] \displaystyle\frac{1}{n}\sum_{k=1}^n\frac{I(S_k\ge y)\dot\mu_{\gamma_0}(S_k)Z_k'}{\sigma^2_{\zeta,\theta}(S_k)} & \displaystyle\frac{1}{n}\sum_{k=1}^n\frac{I(S_k\ge y)\dot\mu_{\gamma_0}(S_k)\dot\mu'_{\gamma_0}(S_k)}{\sigma^2_{\zeta,\theta}(S_k)}\end{pmatrix}.$$

By a Glivenko-Cantelli argument, one can show that

$$\sup_{y\in\bar{\mathbb{R}}}\|\widetilde{M}_y-M_y\|=o_p(1). \tag{5.10}$$

This in turn implies that $\sup_{y\le s_0}\left\|\widetilde{U}_n(y)-U_n(y)+M_y\theta_n\right\|=o_p(1)$. Routine arguments, together with the conditions (e), (m) and (g3), lead to $\sup_{y\le s_0}\|\hat{M}_y^{-1}-M_y^{-1}\|=o_p(1)$, by the positive definiteness of $M_y$ for all $y\in\mathbb{R}$.

For convenience, let $\mathcal{P}_n(s)$ be the the second term on the right hand side of (5.3), and $\mathcal{P}_0(s)$ be the the second term on the right hand side of (5.4). Note that

$$\mathcal{P}_n(s)=\int_{y\le s}\int\frac{(x',\dot\mu'_{\hat\gamma_n}(y)-\dot\mu'_{\gamma_0}(y)+\dot\mu'_{\gamma_0}(y))}{\sigma_{\zeta,\theta}(y)}\Big[\frac{\sigma_{\zeta,\theta}(y)}{\tilde\sigma_3(y)}-1+1\Big]\cdot$$
$$[\hat{M}_y^{-1}-M_y^{-1}+M_y^{-1}]\cdot[\widetilde{U}_n(y)-U_n(y)+U_n(y)]d\hat{F}_{Z,S}(x,y),$$

which can be written as the sum of

$$B_{n1}(s)=\int_{y\le s}\int\frac{(x',\dot\mu'_{\gamma_0}(y))}{\sigma_{\zeta,\theta}(y)}M_y^{-1}U_n(y)d\hat{F}_{Z,S}(x,y),$$
$$B_{n2}(s)=\int_{y\le s}\int\frac{(x',\dot\mu'_{\gamma_0}(y))}{\sigma_{\zeta,\theta}(y)}M_y^{-1}[\widetilde{U}_n(y)-U_n(y)]d\hat{F}_{Z,S}(x,y),$$

and a remainder term $R_n(s)$, say. In view of (g3), (5.5), one verifies that

$$B_{n1}(s)=P_0(s)+u_p(1),\quad B_{n2}(s)=-E\Big[\frac{(Z',\dot\mu'_{\gamma_0}(S))}{\sigma_{\zeta,\theta}(S)}I(S\le s)\Big]\theta_n+u_p(1).$$

and $\sup_{s\le s_0}|R_n(s)|=o_p(1)$. The claim (a) follows from these results, (5.9), and the fact that

$$\mathcal{W}_n(s)-\mathcal{W}_{\theta_0,F_{Z,S}}(s)=[\hat{W}_{3n}(s)-W_{3n}(s)]-[P_n(s)-P_0(s)].$$

**Proof of Theorem 3.1.** Fix an $s_0 < \infty$. By a similar argument as in the null hypothesis case, we obtain

$$\mathcal{W}_n^a(s) \Longrightarrow B \circ \psi \quad \text{in } D([-\infty, s_0]) \text{ and uniform metric,} \tag{5.11}$$

where $\psi(s) = E(\tilde{\sigma}_a^2(S) I[S \le s]/\sigma_a^2(s))$, and $\tilde{\sigma}_a^2(s) = \sigma_\varepsilon^2 + \beta_a' \Sigma_\xi \beta_a + E_a([g_{\gamma_a}(T) - \mu_{\gamma_a}(S)]^2 | S = s)$.

Now, consider $\mathcal{R}_n^a(s)$. Write $\mathcal{R}_n^a(s) = \mathcal{R}_{n1}^a(s) - \mathcal{R}_{n2}^a(s)$, where

$$\begin{aligned} n^{-1/2}\mathcal{R}_{n1}^a(s) = & \frac{1}{n}\sum_{i=1}^n \frac{\beta_0' X_i + h(T_i) - \beta_a' X_i - g_{\gamma_a}(T_i)}{\sigma_a(S_i)} I(S_i \le s) \\ & + \frac{1}{n}\sum_{i=1}^n \frac{\beta_0' X_i + h(T_i) - \beta_a' X_i - g_{\gamma_a}(T_i)}{\sigma_a(S_i)} \Big[\frac{\sigma_a(S_i)}{\tilde{\sigma}_3(S_i)} - 1\Big] I(S_i \le s). \end{aligned}$$

A Glivenko-Cantelli type argument, together with (3.4), implies that

$$\sup_{s \in \bar{R}} \Big| n^{-1/2}\mathcal{R}_{n1}^a(s) - \mathcal{D}_1(s) \Big| = o_p(1). \tag{5.12}$$

Let

$$\hat{V}_n(y) := \frac{1}{n}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\tilde{\sigma}_3(S_i)} \hat{\ell}(Z_i, S_i) I(S_i \ge y).$$

Then

$$n^{-1/2}\mathcal{R}_{n2}^a(s) = \int_{y \le s} \int \hat{\ell}(x,y)' \hat{M}_y^{-1} \hat{V}_n(y) d\hat{F}_{Z,S}(x,y).$$

Subtracting and adding $\dot{\mu}_{\gamma_a}(S_i)$ from $\dot{\mu}_{\hat{\gamma}_a}(S_i)$, replacing $1/\hat{\sigma}_\zeta^2(S_i)$ with $(\sigma_a^2(S_i) / \hat{\sigma}_{3\zeta}^2(S_i) - 1 + 1)/\sigma_a^2(S_i)$, $\hat{V}_n(y)$ can be written as the sum of four terms:

$$\begin{aligned} \hat{V}_{n1}(y) &= \frac{1}{n}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\sigma_a(S_i)} \Big[\frac{\sigma_a^2(S_i)}{\tilde{\sigma}_3^2(S_i)} - 1\Big] \ell(Z_i, S_i) I(S_i \ge y), \\ \hat{V}_{n2}(y) &= \frac{1}{n}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\sigma_a(S_i)} \ell(Z_i, S_i) I(S_i \ge y), \\ \hat{V}_{n3}(y) &= \frac{1}{n}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\sigma_a^2(S_i)} \Big[\frac{\sigma_a^2(S_i)}{\tilde{\sigma}_3^2(S_i)} - 1\Big] \begin{pmatrix} 0 \\ \dot{\mu}_{\hat{\gamma}_n}(S_i) - \dot{\mu}_{\gamma_a}(S_i) \end{pmatrix} I(S_i \ge y), \\ \hat{V}_{n4}(y) &= \frac{1}{n}\sum_{i=1}^n \frac{Y_i - Y_i^a}{\sigma_a^2(S_i)} \begin{pmatrix} 0 \\ \dot{\mu}_{\hat{\gamma}_n}(S_i) - \dot{\mu}_{\gamma_a}(S_i) \end{pmatrix} I(S_i \ge y). \end{aligned}$$

Condition (g3), (3.4), and the additional assumption $E_a\Big[(Y - Y^a)/\sigma_a^2(S)\Big]^2 < \infty$, imply $\sup_{y \in \bar{\mathbb{R}}} \big|\hat{V}_{nj}(y)\big| = o_p(1)$, for $j = 1, 3, 4$. As for $\hat{V}_{n2}(y)$, a Glivenko-Cantelli type argument yields $\sup_{y\in\bar{R}} |\hat{V}_{n2}(y) - \rho(y)| = o_p(1)$. These facts in turn imply

$$\sup_{y\in\bar{R}} |\hat{V}_n(y) - \rho(y)| = o_p(1). \tag{5.13}$$

Using exactly the same argument as in the null case, one can verify that under the alternative $H_a$, $\sup_{y\le s_0} \left|\hat{M}_y^{-1} - A_y^{-1}\right| = o_p(1)$. Rewrite $n^{-1/2}\mathcal{R}_{n2}^a(s)$ as

$$\begin{aligned} n^{-1/2}\mathcal{R}_{n2}^a(s) &= \int_{y\le s}\int \frac{(x', \dot{\mu}'_{\hat{\gamma}_n}(y) - \dot{\mu}'_{\gamma_a}(y) + \dot{\mu}'_{\gamma_a}(y))}{\sigma_a(y)}\Big[\frac{\sigma_a(y)}{\tilde{\sigma}_3(y)} - 1 + 1\Big] \\ &\qquad \cdot[\hat{M}_y^{-1} - A_y^{-1} + A_y^{-1}]\cdot[\hat{V}_n(y) - \rho(y) + \rho(y)]d\hat{F}_{Z,S}(x,y) \\ &= \int_{y\le s}\int \frac{(x', \dot{\mu}'_{\gamma_a}(y))}{\sigma_a(y)} A_y^{-1}\rho(y)d\hat{F}_{Z,S}(x,y) + R_n(s). \end{aligned}$$

Under (g3), (3.4), one can show that $\sup_{s\le s_0} |R_n(s)| = o_p(1)$. Using a Glivenko-Cantelli type argument, one further concludes that

$$\int_{y\le s}\int \frac{(x', \dot{\mu}'_{\gamma_a}(y))}{\sigma_a(y)} A_y^{-1}\rho(y)d\hat{F}_{Z,S}(x,y) = \mathcal{D}_2(s) + u_p(1).$$

In fact, with $h_y = E(Z|S = y)$, we can write

$$\mathcal{D}_2(s) = \int_{y\le s} \frac{(h'_y, \dot{\mu}'_{\gamma_a}(y))}{\sigma_a(y)} A_y^{-1}\rho(y)dF_S(y).$$

So we have shown that

$$\sup_{s\le s_0} \left|n^{-1/2}\mathcal{R}_{n2}^a(s) - \mathcal{D}_2(s)\right| = o_p(1), \quad (H_a). \tag{5.14}$$

Then (5.12) and (5.14) jointly implies

$$\sup_{s\le s_0} \left|n^{-1/2}\mathcal{R}_n^a(s) - [\mathcal{D}_1(s) - \mathcal{D}_2(s)]\right| = o_p(1). \tag{5.15}$$

Finally, the consistency is derived by combining (5.11), (5.15), the inequality

$$\sup_{s\le s_0} \left|\mathcal{W}_n^a(s) + \mathcal{R}_n^a(s)\right| \ge \sup_{s\le s_0} \left|\mathcal{R}_n^a(s)\right| - \sup_{s\le s_0} \left|\mathcal{W}_n^a(s)\right|$$

and the condition $d = \sup_{s\le s_0} |\mathcal{D}_1(s) - \mathcal{D}_2(s)| > 0$.

**Proof of Theorem 3.2.** Details of the proof of this theorem are similar to that of Theorem 3.1, with obvious modifications.

**Acknowledgement**

The authors are grateful to two anonymous referees for providing a constructive criticism that helped to improve the presentation of the paper. Research of Hira L. Koul was supported in part by the NSF DMS grant 0704130.

**Appendix: $\sqrt{n}$-Consistency of the LSE**

The validity of Theorem 2.1, 3.1, and 3.2 require the $\sqrt{n}$-consistency of $\hat{\beta}_n$ and $\hat{\gamma}_n$ under all three hypotheses $H_0$, $H_a$ and $H_{\text{Loc}}$. Under some regularity conditions, we can show that the least squares procedure can provide such estimators.

The argument provided below is only for the case of $H_a$, but the adaption to the $H_0$ and $H_{\text{Loc}}$ cases is straightforward.

To be specific, let $h(\cdot)$ and $H_a$ be as in Section 3.1, and recall the definition (3.2). Assume

- C1: $L(\beta,\gamma) := E_a[Y-\beta' Z-\mu_\gamma(S)]^2$ exists for all $\beta,\gamma$ and takes unique minimum at $(\beta_a', \gamma_a')'$ which is an interior point of $\Theta$,
- C2: $\Sigma := E(ZZ')$ is positive definite,
- C3: the parameter space $\Gamma \in \mathbb{R}^q$ is convex and compact,
- C4: $Eh^2(T) < \infty$, $E\sup_\gamma |g_\gamma(T)|^2 < \infty$,
- C5: $E\sup_\gamma |\dot{g}_\gamma(T)|^2 < \infty$, $E\sup_\gamma \|\ddot{g}_{\gamma_a}(T)\|^2 < \infty$.

Note that the Lipschitz condition (g1) implies $E\sup_\gamma |g_\gamma(T)|^2 < \infty$. Note that the Lipschitz condition (g1) implies $E\sup_\gamma |g_\gamma(T)|^2 < \infty$.

The condition (C1) guarantees the validity of the least squares procedures while (C2) ensures the uniqueness of the least squares estimator for $\beta$, a common assumption even in simple linear regression models. Conditions (C3)$-$(C5) are the usual assumptions needed for proving consistency of the least squares estimators in nonlinear regression models.

Now, for a fixed $\gamma \in \Gamma$, the equation $\partial L(\beta,\gamma)/\partial\beta = -2E_a(Y - \beta' Z - \mu_\gamma(S))Z = 0$, yields

$$\beta(\gamma) := \Sigma^{-1}E_a Z(Y-\mu_\gamma(s)) = \Sigma^{-1}E_a[ZY] - \Sigma^{-1}E[Z\mu_\gamma(S)]. \tag{A.1}$$

Therefore, $L(\beta(\gamma),\gamma) \le L(\beta,\gamma)$, for all $\beta,\gamma$. Let $\tilde{\gamma}_a$ be a solution of $\partial L(\beta(\gamma),\gamma)/\partial\gamma = 0$ or, equivalently, a solution of

$$E_a[Y - b'Z + h(\gamma)'Z - \mu_\gamma(S)]\cdot[\dot{h}(\gamma)Z - \dot{\mu}_\gamma(S)] = 0, \tag{A.2}$$

where $b = \Sigma^{-1}E(YZ),\ h(\gamma) = \Sigma^{-1}E[Z\mu_\gamma(S)]$. Then we must have

$$L(\beta(\tilde{\gamma}_a),\tilde{\gamma}_a) \le L(\beta(\gamma),\gamma) \le L(\beta,\gamma),\ \forall\,\beta,\gamma.$$

Under (C1), (C2), $\tilde{\gamma}_a$ must be unique and $\beta_a = \beta(\tilde{\gamma}_a)$, $\gamma_a = \tilde{\gamma}_a$.

Now consider the empirical version of $L$: $L_n(\beta,\gamma) = n^{-1}\sum_{i=1}^n [Y_i - \beta' Z_i - \mu_\gamma(S_i)]^2$. For any fixed $\gamma \in \Gamma$, the equation $\partial L_n(\beta,\gamma)/\partial\beta = 0$ yields

$$\beta_n(\gamma) = \Sigma_n^{-1}[\overline{ZY} - \overline{Z\mu_\gamma(S)}], \tag{A.3}$$

where

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^n Z_i Z_i', \quad \overline{ZY} = \frac{1}{n}\sum_{i=1}^n Z_i Y_i, \quad \overline{Z\mu_\gamma(S)} = \frac{1}{n}\sum_{i=1}^n Z_i \mu_\gamma(S_i).$$

Therefore, $L_n(\beta_n(\gamma),\gamma) \le L_n(\beta,\gamma)$, for all $\beta, \gamma \in \Gamma$. Let $\tilde{\gamma}_n$ be a the solution of the equation $\partial L_n(\beta_n(\gamma),\gamma)/\partial\gamma = 0$, or

$$\frac{1}{n}\sum_{i=1}^n [Y_i - \beta_n'(\gamma) Z_i - \mu_\gamma(S_i)]\,[\dot{\beta}_n(\gamma) Z_i - \dot{\mu}_\gamma(S_i)] = 0,$$

where $\dot{\beta}_n(\gamma) := \Sigma_n^{-1}[\overline{ZY} - \overline{Z\dot{\mu}_\gamma(S)}]$. Then we must have $L_n(\beta_n(\tilde{\gamma}_n),\tilde{\gamma}_n) \le L_n(\beta_n(\gamma),\gamma) \le L(\beta,\gamma)$, $\forall\,\beta,\gamma$. In other words, $\beta_n(\tilde{\gamma}_n), \tilde{\gamma}_n$ is a minimizer of the nonlinear least square solution of (3.3). Denote these estimators simply by $\hat{\beta}_n$, $\hat{\gamma}_n$.

## A.1. Consistency of $\hat{\beta}_n$, $\hat{\gamma}_n$

Notice that

$$L_n(\beta_n(\gamma),\gamma) = \frac{1}{n}\sum_{i=1}^n [Y_i - \Sigma_n^{-1}\{(\overline{ZY}) - \overline{Z\mu_\gamma(S)})\}' Z_i - \mu_\gamma(S_i)]^2,$$
$$L(\beta(\gamma),\gamma) = E_a[Y - \Sigma^{-1}\{E_a(ZY) - E(Z\mu_\gamma(S))\}' Z - \mu_\gamma(S)]^2,$$

as functions of $\gamma$, are defined on a compact subset of $\mathbb{R}^q$. Then under some conditions, we can show that

$$L_n(\beta_n(\gamma),\gamma) \to L(\beta(\gamma),\gamma) \quad \text{uniformly for } \gamma. \tag{A.4}$$

For this purpose, we need the following lemma.

**Lemma A.1.** (Jennrich (1969)) *Let $g$ be a function on $\mathfrak{X} \times \Theta$, where $\mathfrak{X}$ is a Euclidean space and $\Theta$ is a compact subset of a Euclidean space. Let $g(x,\theta)$ be a continuous function of $\theta$ for each $x$ and a measurable function of $x$ for each $\theta$. Assume also that $g(x,\theta) \le h(x)$ for all $x$ and $\theta$, where $h$ is integrable with respect to a probability distribution function $F$ on $\mathfrak{X}$. If $X_1, X_2, \ldots$ is a random sample from $F$ then $n^{-1}\sum_{i=1}^n g(X_i,\theta) \to E(g(X,\theta))$, a.s. uniformly for all $\theta$ in $\Theta$.*

Expanding $L_n(\beta_n(\gamma),\gamma)$, one can see that to show (A.4), it suffices to show that, almost surely,

$$\overline{Y\mu_\gamma(S)} \to E_a(Y\mu_\gamma(S)), \quad \overline{Z\mu_\gamma(S)} \to E(Z\mu_\gamma(S)), \quad \overline{\mu_\gamma^2(S)} \to E(\mu_\gamma^2(S)),$$

uniformly in $\gamma \in \Gamma$. But these can be fulfilled by letting $h(\cdot) = \sup_\gamma |Y\mu_\gamma(S)|$, $\sup_\gamma |Z\mu_\gamma(S)|$, and $\sup_\gamma |\mu_\gamma^2(S)|$, and using (e), (c2), (C3), and Lemma A.1. Therefore, $\hat{\gamma}_n \to \gamma_a$ almost surely. For if there were a subsequence of $\hat{\gamma}_n$, say $\hat{\gamma}_{n_k}$, that converged to $\gamma_1$ almost surely, then

$$L(\beta(\gamma_1),\gamma_1) \leftarrow L_{n_k}(\beta_{n_k}(\hat{\gamma}_{n_k}),\hat{\gamma}_{n_k}) \le L_{n_k}(\beta_{n_k}(\gamma_a),\gamma_a) \to L(\beta(\gamma_a),\gamma_a)$$

and uniqueness of $\gamma_a$ imply the desired strong consistency. Finally, $\hat{\beta}_n \to \beta_a$ almost surely follows from the fact $\beta_n(\gamma) \to \beta(\gamma)$ uniformly for $\gamma \in \Gamma$, and the consistency of $\hat{\gamma}_n$ to $\gamma_a$.

## A.2. Convergence rates of $\hat{\beta}_n$, $\hat{\gamma}_n$

By Taylor expansion,

$$\frac{\partial L_n(\beta_n(\gamma),\gamma)}{\partial\gamma} = \frac{\partial L_n(\beta_n(\gamma),\gamma)}{\partial\gamma}\bigg|_{\gamma=\gamma_a} + \frac{\partial^2 L_n(\beta_n(\gamma),\gamma)}{\partial\gamma\partial\gamma'}\bigg|_{\gamma=\gamma^*}(\gamma-\gamma_a).$$

Evaluate both sides at $\gamma = \hat{\gamma}_n$, then

$$\begin{aligned} 0 &= \frac{\partial L_n(\beta_n(\gamma),\gamma)}{\partial\gamma}\bigg|_{\gamma=\gamma_a} + \frac{\partial^2 L_n(\beta_n(\gamma),\gamma)}{\partial\gamma\partial\gamma'}\bigg|_{\gamma=\gamma_n^*}(\hat{\gamma}_n-\gamma_a) \qquad (A.5)\\ &=: T_{n1} + T_{n2}(\hat{\gamma}_n-\gamma_a), \end{aligned}$$

where $\gamma_n^*$ lies between $\hat{\gamma}_n$ and $\gamma_a$. The convexity of $\Gamma$ implies $\gamma_n^* \in \Gamma$. By subtracting and adding $\beta(\gamma_a)$ and $\dot{\beta}(\gamma_a)$ from $\beta_n(\gamma_a)$ and $\dot{\beta}_n(\gamma_a)$, respectively, $T_{n1}$ can be written as the sum of four terms:

$$\begin{aligned} T_{n11} &= \frac{1}{n}\sum_{i=1}^n [Y_i - \beta'(\gamma_a)Z_i - \mu_{\gamma_a}(S_i)][\dot{\beta}(\gamma_a)Z_i - \dot{\mu}_{\gamma_a}(S_i)],\\ T_{n12} &= (\dot{\beta}_n(\gamma_a) - \dot{\beta}(\gamma_a))\frac{1}{n}\sum_{i=1}^n [Y_i - \beta'(\gamma_a)Z_i - \mu_{\gamma_a}(S_i)]Z_i,\\ T_{n13} &= (\beta_n(\gamma_a) - \beta(\gamma_a))'\frac{1}{n}\sum_{i=1}^n Z_i[\dot{\beta}(\gamma_a)Z_i - \dot{\mu}_{\gamma_a}(S_i)],\\ T_{n14} &= (\beta_n(\gamma_a) - \beta(\gamma_a))'\frac{1}{n}\sum_{i=1}^n Z_i(\dot{\beta}_n(\gamma_a) - \dot{\beta}(\gamma_a))Z_i. \end{aligned}$$

We first consider the asymptotic behavior of $\beta_n(\gamma_a)$. From (A.1) and (A.3), we obtain

$$\begin{aligned}\beta_n(\gamma_a)-\beta(\gamma_a) &= \Sigma_n^{-1}[\overline{ZY}-\overline{Z\mu_{\gamma_a}(S)}]-\Sigma^{-1}[E_a(ZY)-E(Z\mu_{\gamma_a}(S))] \qquad \text{(A.6)}\\ &= (\Sigma_n^{-1}-\Sigma^{-1})[\overline{ZY}-\overline{Z\mu_{\gamma_a}(S)}]\\ &\quad +\Sigma^{-1}[\overline{ZY}-\overline{Z\mu_{\gamma_a}(S)}-E_a(ZY)+E(Z\mu_{\gamma_a}(S))].\end{aligned}$$

By the LLN, $\overline{ZY}-\overline{Z\mu_{\gamma_a}(S)}\to E_a(ZY)-E(Z\mu_{\gamma_a}(S))$. Let

$$\Delta=E_a(ZY)-E(Z\mu_{\gamma_a}(S)),\quad \overline{ZZ'}=(B_{jk})_{p\times p},\quad E(ZZ')=(b_{jk})_{p\times p},$$

and $B^*_{jk}$ be the cofactor of $B_{jk}$ and $b^*_{jk}$ be the cofactor of $b_{jk}$. Then,

$$\begin{aligned}\sqrt{n}(\Sigma_n^{-1}-\Sigma^{-1})\Delta &= \sqrt{n}\left(\frac{(B^*_{jk})_{p\times p}}{|\overline{ZZ'}|}-\frac{(b^*_{jk})_{p\times p}}{|E(ZZ')|}\right)\Delta\\ &= -\sqrt{n}(|\overline{ZZ'}|-|E(ZZ')|)\frac{(B^*_{jk})_{p\times p}\Delta}{|\overline{ZZ'}|\cdot|E(ZZ')|}+\frac{\sqrt{n}(B^*_{jk}-b^*_{jk})_{p\times p}\Delta}{|E(ZZ')|}, \qquad \text{(A.7)}\end{aligned}$$

where $|A|$ denotes the determinant of the square matrix $A$. Using the fact that $\sqrt{n}(B^*_{jk}-b^*_{jk})=O_p(1)$, one can show both terms in (A.7) are $O_p(1)$. Also, it is easy to see that the second term in (A.6) has the same order. This then implies $\sqrt{n}(\beta_n(\gamma_a)-\beta(\gamma_a))=O_p(1)$. Therefore,

$$\begin{aligned}\sqrt{n}T_{n13} &= \frac{1}{n}\sum_{i=1}^{n}[\dot\beta(\gamma_a)Z_i-\dot\mu_{\gamma_a}(S_i)]Z_i'\sqrt{n}(\beta_n(\gamma_a)-\beta(\gamma_a))\\ &= E[\dot\beta(\gamma_a)ZZ'-\dot\mu_{\gamma_a}(S)Z']\sqrt{n}(\beta_n(\gamma_a)-\beta(\gamma_a))+o_p(1)=O_p(1).\end{aligned}$$

Similarly, by considering each row in the matrix $\dot\beta_n(\gamma_a)-\dot\beta(\gamma_a)$, we can show that, under (C4), $\sqrt{n}T_{n12}=O_p(1)$, $\sqrt{n}T_{n14}=o_p(1)$. Use that (A.2) implies $E_a[Y-\beta'(\gamma_a)Z-\mu_{\gamma_a}(S)][\dot\beta(\gamma_a)Z-\dot\mu_{\gamma_a}(S)]=0$ to show that $\sqrt{n}T_{n11}=O_p(1)$. Hence, $\sqrt{n}T_{n1}=O_p(1)$.

Now consider the matrix $T_{n2}$ in (A.5). Manipulating the derivatives of matrix, one obtains

$$\begin{aligned}T_{n2} &= \frac{1}{n}\sum_{i=1}^{n}[\dot\beta_n(\gamma_n^*)Z_i-\dot\mu_{\gamma_n^*}(S_i)][\dot\beta_n(\gamma_n^*)Z_i-\dot\mu_{\gamma_n^*}(S_i)]'\\ &\quad -\frac{1}{n}\sum_{i=1}^{n}[Y_i-\beta_n'(\gamma_n^*)Z_i-\mu_{\gamma_n^*}(S_i)][B_n(I_{q\times q}\otimes Z_i)-\ddot\mu_{\gamma_n^*}(S_i)],\end{aligned}$$

where $B_n=[\ddot\beta_{n1}(\gamma_n^*),\ddot\beta_{n2}(\gamma_n^*),\cdots,\ddot\beta_{nq}(\gamma_n^*)]_{q\times pq}$, and $\ddot\beta_{nj}(\gamma_n^*)$ is the derivative of the $j$-th row of $\dot\beta_n(\gamma)$ with respect to $\gamma$, then evaluated at $\gamma=\gamma_n^*$, $\otimes$ denotes the Kronecker product.

Since $\hat{\gamma}_n$ is strongly consistent for $\gamma$, so is $\gamma_n^*$. By (e), (C3), (C4), (C5), using Lemma A.1, we can show that $T_{n2}$ is $\Pi + o_p(1)$ asymptotically, where

$$\Pi := E[\dot{\beta}(\gamma_a)Z - \dot{\mu}_{\gamma_a}(S)][\dot{\beta}(\gamma_a)Z - \dot{\mu}_{\gamma_a}(S)]' \\ -E[Y - \beta(\gamma_a)Z - \mu_{\gamma_a}(S)][B(I_{q\times q} \otimes Z) - \ddot{\mu}_{\gamma_a}(S)],$$

with $B = [\ddot{\beta}_1(\gamma_a), \ddot{\beta}_2(\gamma_a), \cdots, \ddot{\beta}_q(\gamma_a)]_{q\times pq}$, and $\ddot{\beta}_j(\gamma_a)$ is the derivative of the $j$-th row of $\dot{\beta}(\gamma)$ with respect to $\gamma$, then evaluated at $\gamma = \gamma_a$. Finally, if $\Pi$ is nonsingular, we can get $\sqrt{n}(\hat{\gamma}_n - \gamma_a) = O_p(1)$. The claim $\sqrt{n}(\hat{\beta}_n - \beta_a) = O_p(1)$ can then be obtained by replacing $\gamma$ with $\hat{\gamma}_n$ in (A.3), and using a routine argument.

## References

An, H. Z. and Cheng, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Internat. Statist. Rev.* **59**, 287-307.

Billingsley, P. (1968). *Convergence of Probability Measure.* Wiley, New York.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman & Hall/CRC, Boca Raton.

Delaigle, A., Hall, P. and Qiu, P. H. (2006). Nonparametric methods for solving the Berkson errors-in-variables problem. *J. Roy. Statist. Soc. Ser. B* **68**, 201-220.

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

Hu, Y. and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* **76**, 195-216.

Huwang, L. and Huang, H. Y. S. (2000). On errors-in-variables in polynomial regression - Berkson case. *Statist. Sinica* **10**, 923-936.

Jennrich, R. J. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

Khmaladze, E. V. (1979). The use of $\omega^2$ tests for testing parametric hypotheses. *Theory Probab. Appl.* **24**, 283 - 301.

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

Koul, H. L. (2006). Model diagnostics via martingale transforms: a brief review. *Frontiers in Statistics*, 183-206, Imp. Coll. Press, London.

Koul, H. L. and Song, W. (2008). Model checking with Berkson measurement errors. *J. Statist. Plann. Inference* **138**, 1615-1628.

Liang, H. (2006). Checking linearity of non-parametric component in partially linear models with an application in systemic inflammatory response syndrome study. *Statist. Meth. Medical Res.* **15**, 273-284.

Lubin, J. H., Wang, Z. Y., Wang, L. D., Boice, Jr, J. D., Cui, H. X., Zhang, S. R., Conrath, S., Xia, Y., Shang, B., Cao, J. S. and Kleinerman, R. A.(2005). Adjusting lung cancer risks for temporal and spatial variations in radon concentration in dwellings in Gansu Province, China. *Radn Res.* **163**, 571-579.

Raaschou-Nielsen, O., Hertel, O., Thomsen, B. and Olsen, J. (2001). Air pollution from traffic at the residence of children with cancer. *Am. J. Epidem.* **153**, 433-443.

Reeves, G. K., Cox, D. R., Darby, S. C. and Whitley, E. (1998). Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statist. Med.* **17**, 2157-2177.

Stram, D. O., Huberman, M. and Wu, A. H. (2002). Is residual confounding a reasonable explanation for the apparent protective effects of beta-carotene found in epidemiological studies of lung cancer in smokers? *Am. J. Epidem.* **155**, 622-628.

Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26** 1916-1934.

Stute, W. and Zhu, L. X. (2002). Model checks for generalized linear models. *Scand. J. Statist* **29**, 535-545.

Stute, W. and Zhu, L. X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048-1083.

Thomas, D., Bowman, J., Jiang, L., Jiang, F. and Peters, J. (1999). Residential magnetic fields predicted from wiring configurations: II, Relationships to childhood leukemia. *Bioelectromagnetics* **20**, 414-422.

von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367-395.

Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *Ann. Statist.* **32**, 2559-2579.

Zhu, L. X. and Ng, Kai W. (2003). Checking the adequacy of a partial linear model. *Satist. Sinica* **13**, 763-781.

Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824-1027, U.S.A.

E-mail: koul@stt.msu.edu

Department of Statistics, Kansas State University, Manhattan, KS U.S.A.

E-mail: weixing@ksu.edu

(Received September 2008; accepted June 2009)