Regression Model Checking with Berkson Measurement Errors\textsuperscript{1}

HIRA L. KOUL AND WEIXING SONG

MICHIGAN STATE UNIVERSITY

Abstract

This paper discusses asymptotically distribution free tests for fitting a parametric regression model to the regression function in the Berkson measurement error model. These tests are based on a martingale transform of a certain marked empirical processes of residuals. A simulation that is included shows a very desirable finite sample behavior of the proposed inference procedure.

1 Introduction

A classical problem in statistics is to use a covariate $X$ to explain the response $Y$. As is the practice this is often done in terms of the regression function $\mu(x) := E(Y|X = x)$, assuming it exists. Usually, in practice the predictor variable $X$ is assumed to be observable. But in many experiments, it is impossible to observe $X$. Instead, a surrogate $Z$ of $X$ can be measured. As an example, consider the herbicide study of Rudemo, et al. (1989) in which a nominal measured amount $Z$ of herbicide was applied to a plant but the actual amount absorbed by the plant $X$ is unobservable. As another example, from Wang (2004), an epidemiologist studies the severity of a lung disease, $Y$, among the residents in a city in relation to the amount of certain air pollutants, $X$. The amount of the air pollutants $Z$ can be measured at certain observation stations in the city, but the actual exposure of the residents to the pollutants, $X$, is unobservable and may vary randomly from the $Z$-values. In both cases, $X$ can be expressed as $Z$ plus a random error. There are many similar examples in agricultural or medical studies, see e.g., Fuller (1987), Carroll, Ruppert and Stefanski (1995), among others.

All these examples can be formalized into the so called Berkson model

\begin{equation}
Y = \mu(X) + \varepsilon, \quad X = Z + \eta,
\end{equation}

where the random errors $\eta$ and $\varepsilon$ and the observable control variable $Z$ are assumed to be mutually independent, with $E\varepsilon = 0 = E\eta$, and $0 < \sigma^2_{\varepsilon} := E(\varepsilon^2) < \infty$.

\textsuperscript{1}MSC: primary 62G08; secondary 62G10

Key words and phrases: Berkson Measurement Error; Marked Empirical Process; Brownian motion.
Let $\mathcal{M} := \{m_\theta(x) : x \in \mathbb{R}, \theta \in \Theta \subset \mathbb{R}^q\}$, $q \geq 1$, be a given class of real valued parametric functions. The problem of interest is to test the hypothesis that $\mu \in \mathcal{M}$, i.e., to test

$$H_0 : \mu(x) = m_{\theta_0}(x), \text{ for some } \theta_0 \in \Theta \text{ and for all } x,$$

against the alternatives

$$H_1 : H_0 \text{ is not true},$$

based on the $n$ i.i.d. observations $(Z_i, Y_i); 1 \leq i \leq n$, from the Berkson model (1.1).

In the case of no measurement error in $X$, several tests of $H_0$ are available in the literature. See Hart (1997) and references therein and Koul and Ni (2004), among others, for tests based on nonparametric regression function estimates. An and Cheng (1991), Stute (1997), Stute, Thies, and Zhu (1998), and Khmaladze and Koul (2004) base their tests on a certain marked empirical process of residuals. An advantage of these tests over those based on nonparametric regression function estimates is that they have nontrivial asymptotic power against $n^{-1/2}$-nonparametric alternatives. The latter two papers provide a martingale type transformation of the underlying marked empirical process whose asymptotic null distribution is free from the underlying model, error d.f. and the design d.f. The purpose of this paper is to develop similar tests for the above testing problem for the Berkson measurement errors model (1.1).

Let $\sigma^2(x) := E\{(Y - \mu(X))^2|X = x\}$ and $s_n(x)$ be a consistent estimator of $\sigma_\varepsilon(x)$. The marked empirical process used by STZ to construct tests of $H_0$ in the classical heteroscedastic regression model when $X_i$ are observable is

$$n^{-1/2} \sum_{i=1}^n \frac{(Y_i - \mu(X_i))}{s_n(X_i)} I(X_i \leq x), \quad x \in \mathbb{R}.$$ 

Clearly this process needs to be suitably modified in order for it to be useful in the current set up. Under the model assumptions, $\nu(z) := E(Y|Z = z) = E(\mu(X)|Z = z)$. Thus one may think of the new regression model $Y = \nu(Z) + \zeta$, where $E(\zeta|Z) = 0$, a.s., so that the error $\zeta$ is uncorrelated with $Z$. Let $\nu_\theta(z) := E(m_\theta(X)|Z = z)$.

Berkson (1950) pointed out that the ordinary least square estimators based on $(Z_i, Y_i); 1 \leq i \leq n$, are unbiased and consistent in parametric linear regression models without assuming the knowledge of the densities of $\varepsilon$, $X$ or $\eta$. But if the regression model is nonlinear in $X$ or if there are other parameters in the Berkson model that need to be estimated, then extra information about these densities should be supplied to ensure the identifiability. A standard assumption in the literature is to assume that $f_\eta$, the density of $\eta$, is known or unknown only up to an Euclidean parameter vector, cf., Carroll, et al. (1995), Huwang and Huang (2000), Wang (2004), among others. In this paper, we shall assume that $f_\eta$ is known.
Because of the independence of $Z$ and $\eta$, under (1.1),
\[
\nu(z) = \int \mu(x)f_\eta(x-z)dx, \quad \nu_\theta(z) = \int m_\theta(x)f_\eta(x-z)dx, \quad \theta \in \Theta, z \in \mathbb{R}.
\]
Since $f_\eta$ is known, $\nu_\theta$ is known up to the parameter $\theta$. Let
\[
\sigma^2_\nu(z) := E[(Y - \nu(Z))^2|Z = z], \quad \tau^2_\nu(z) := E[(\mu(X) - \nu(Z))^2|Z = z], z \in \mathbb{R}.
\]
Because $Z$ is uncorrelated with $Y - \nu(Z)$ and independent of $\varepsilon$, we obtain, with $\sigma^2_\varepsilon := \text{Var}(\varepsilon)$,
\[
(1.2) \quad \sigma^2_\nu(z) = \sigma^2_\varepsilon + \tau^2_\nu(z), \quad z \in \mathbb{R}.
\]
Extend the definition of $\nu$, $\nu_\theta$ and $\tau_\nu$ to $\mathbb{R}$ by assigning the value zero to these functions at $\pm \infty$. This convention will apply to analogs of these functions in the sequel. Note that then $\sigma^2_\nu(z) \geq \sigma^2_\varepsilon > 0$, for all $z \in \mathbb{R} := [-\infty, \infty]$.

Consider the problem of testing the simple hypothesis $H$: $\mu = \mu_0$, where $\mu_0$ is a known regression function. Set $\nu_0(z) := E(\mu_0(X)|Z = z)$ and write $\sigma_0(z)$, $\tau_0(z)$ for $\sigma_\nu(z)$, $\tau_\nu(z)$, respectively. Then,
\[
\tau^2_0(z) = \int [\mu_0(y) - \nu_0(z)]^2 f_\eta(y-z)dy.
\]
Observe that under $H$, $\tau^2_0(z)$ is known for all $z$, and that $E(Y - \nu_0(Z))^2 = E\tau^2_0(Z) + \sigma^2_\varepsilon$. Hence a consistent estimator of $\sigma^2_\nu$, under $H$, is given by
\[
s^2_{n0} := \left| n^{-1} \sum_{i=1}^n (Y_i - \nu_0(Z_i))^2 - n^{-1} \sum_{i=1}^n \tau^2_0(Z_i) \right|.
\]
This in turn gives a consistent estimator of $\sigma^2_0(z)$ to be
\[
\sigma^2_{n0}(z) := \tau^2_0(z) + s^2_{n0}, \quad z \in \mathbb{R}.
\]
The analog of the above marked empirical process suitable here for testing $H$ is
\[
V^0_n(z) := n^{-1/2} \sum_{i=1}^n \frac{Y_i - \nu_0(Z_i)}{\sigma_{n0}(Z_i)} I(Z_i \leq z), \quad z \in \mathbb{R}.
\]
Let $G$ denote the d.f. of $Z$ assumed to be continuous. In view of Theorem 2.1 below, under $H$, $V^0_n$ converges weakly to $B \circ G$ in $D(\mathbb{R})$ and uniform metric. This result, for example, implies that the test that rejects $H$ whenever
\[
(1.3) \quad \sup_{-\infty \leq z \leq \infty} |V^0_n(z)| > b_\alpha,
\]
is of the asymptotic size $\alpha$, where $b_\alpha$ is such that $P(\sup_{0 \leq t \leq 1} |B(t)| > b_\alpha) = \alpha$. Consistency of this test and its asymptotic power against $n^{-1/2}$-nonparametric alternatives is discussed in the subsection 2.1 below.
Now consider the more interesting problem of testing $H_0$. Let $P_\theta$ and $E_\theta$ denote the probability measure and expectation, respectively, under the model (1.1) when $\mu = m_\theta$, $\theta \in \Theta$. Let

\begin{equation}
\sigma^2_\theta(z) := E_\theta[(Y - \nu_\theta(Z))^2 | Z = z],
\end{equation}

\begin{equation}
\tau^2_\theta(z) := E[(m_\theta(X) - \nu_\theta(Z))^2 | Z = z] = \int [m_\theta(y) - \nu_\theta(z)]^2 f_\eta(y - z) \, dy, \quad \theta \in \Theta, \quad z \in \mathbb{R}.
\end{equation}

Arguing as for (1.2), we obtain that under (1.1) and $P_\theta$,

\begin{equation}
\sigma^2_\theta(z) = \sigma^2 + \tau^2_\theta(z) \geq \sigma^2 > 0, \quad z \in \mathbb{R}, \quad \theta \in \Theta.
\end{equation}

Under $H_0$, $\tau^2_{\theta_0}(z)$ is known except for $\theta_0$. Let $\theta_n$ be a $n^{1/2}$-consistent estimator of $\theta_0$ under $H_0$ and define

\begin{equation}
s^2_n := \left| n^{-1} \sum_{i=1}^{n} (Y_i - \nu_{\theta_n}(Z_i))^2 - n^{-1} \sum_{i=1}^{n} \tau^2_{\theta_n}(Z_i) \right|.
\end{equation}

This in turn suggests an estimator of $\sigma^2_{\theta_0}(z)$ to be

\begin{equation}
\sigma^2_{\theta_n}(z) := s^2_n + \tau^2_{\theta_n}(z), \quad z \in \mathbb{R}.
\end{equation}

The analog of the above process suitable for testing $H_0$ here is $\hat{V}_n(z) := V_n(z, \theta_n)$, $z \in \mathbb{R}$, where

\begin{equation}
V_n(z, \theta) := n^{-1/2} \sum_{i=1}^{n} \frac{Y_i - \nu_\theta(Z_i)}{\sigma_\theta(Z_i)} I(Z_i \leq z), \quad \theta \in \Theta, \quad z \in \mathbb{R} := [-\infty, \infty].
\end{equation}

However, tests based on this process are not generally asymptotically distribution free (ADF).

But, under some additional assumptions on the null model, the tests based on certain martingale transforms of the process $\hat{V}_n$ are shown to be ADF. This transformation is analogous to the one given in Stute, Thies, and Zhu (1998) (STZ) and is described in the subsection 2.2 below after describing a generic transformation and the needed additional assumptions. Subsection 2.1 discusses a test of the simple hypothesis $H$, its consistency against a fixed alternative and the asymptotic power against $n^{-1/2}$-nonparametric alternatives. Section 3 contains a simulation study. From this study one observes that the finite sample level approximates the nominal level well for larger sample sizes and the empirical power is high (above 0.9) for moderate to large sample sizes at the chosen alternatives. Section 4 contains some proofs.

2 Main Results

The first subsection discusses the tests of $H$ while those for $H_0$ are discussed in the subsection 2.2.
2.1 Tests of a simple hypothesis

To prove the claimed weak convergence of $V_n^0$ to a time transformed Brownian motion under $H$ and to discuss the consistency and asymptotic power of the test (1.3) we first give a general weak convergence result. Accordingly, let $(\xi_i, Z_i), 1 \leq i \leq n,$ be i.i.d. copies of the random vector $(\xi, Z),$ with $E(\xi | Z) = 0, 0 < E(\xi^2) < \infty.$ Let $\sigma^2(z) := E(\xi^2 | Z = z), L(z) := E\sigma^2(Z)I(Z \leq z), z \in \mathbb{R}$. Assume $L$ to be continuous. Consider the process

$$U_n(z) := n^{-1/2} \sum_{i=1}^{n} \xi_i I(Z_i \leq z), \quad z \in \mathbb{R}.$$ 

Observe that $EU_n(z) \equiv 0$ and $EU_n(y)U_n(z) = L(y \wedge z), y, z \in \mathbb{R}$. By the classical CLT, the finite dimensional distributions of the $U_n$-process converge weakly to that of $B \circ L$. Also, a direct calculation shows that

$$E\left(U_n(z) - U_n(z_1)\right)U_n(z_2) - U_n(z))\right)^2 = \frac{n-1}{n} [L(z) - L(z_1)] [L(z_2) - L(z)]$$

$$\leq \frac{1}{n} [L(z_2) - L(z_1)]^2, \quad -\infty \leq z_1 \leq z \leq z_2 \leq \infty.$$ 

Thus by Theorem 12.6 in Billingsley (1968),

$$(2.1) \quad U_n \Longrightarrow B \circ L, \quad \text{in } D(\mathbb{R}) \text{ and uniform metric.}$$

Now, recall $\sigma^2_0(z) = E\{(Y - \nu_0(Z))^2 | Z = z\}$ and let

$$S_n(z) := n^{-1/2} \sum_{i=1}^{n} \frac{Y_i - \nu_0(Z_i)}{\sigma_0(Z_i)} I(Z_i \leq z), \quad z \in \mathbb{R}.$$ 

Clearly this is like a $U_n$ process, with $\xi := (Y - \nu_0(Z))/\sigma_0(Z), \sigma^2(z) = 1,$ and $L = G$. In view of (1.2), $\sigma^2_0(z) \geq \sigma^2_\xi,$ for all $z \in \mathbb{R},$ so that $\xi$ is a well defined r.v. Under (1.1) and $H$, the assumptions for (2.1) are satisfied and we readily obtain that

$$(2.2) \quad S_n \Longrightarrow B \circ G, \quad \text{in } D(\mathbb{R}) \text{ and uniform metric.}$$

Next consider $V_n^0$ process. Let

$$\begin{align*}
\xi_i &:= \frac{Y_i - \nu_0(Z_i)}{\sigma_0(Z_i)}, \quad d_{ni} := \sigma_n(\xi_i)[\sigma_n(\xi_i) + \sigma_0(\xi_i)], \quad 1 \leq i \leq n, \\
W_n(z) &:= n^{-1/2} \sum_{i=1}^{n} \xi_i \left[ \frac{\sigma_0(Z_i)}{\sigma_0(\xi_i)} - 1 \right] I(Z_i \leq z), \quad z \in \mathbb{R}, \quad \Delta_n := \sigma^2_\xi - s_n^2.
\end{align*}$$

We obtain the decomposition $V_n^0(z) \equiv S_n(z) + W_n(z), \text{ where in } S_n, \xi_i \text{ are as in } (2.3)$. Using the elementary fact $(a/b) - 1 = (a^2 - b^2)/b(a+b)$ for any nonzero numbers $a, b$ and $\sigma^2_0(z) - \sigma^2_n(z) = \Delta_n,$
we obtain
\[ W_n(z) = \frac{n^{1/2}\Delta_n}{2} n^{-1} \sum_{i=1}^{n} \frac{\xi_i}{\sigma_0^2(Z_i)} I(Z_i \leq z) + n^{1/2}\Delta_n n^{-1} \sum_{i=1}^{n} \xi_i \left[ \frac{1}{d_{ni}} - \frac{1}{2\sigma_0^2(Z_i)} \right] I(Z_i \leq z) \]
\[ = \frac{n^{1/2}\Delta_n}{2} T_{n1}(z) + n^{1/2}\Delta_n T_{n2}(z), \text{ say.} \]

Assume that
\[ (2.4) \quad E\xi^4 + E(\mu_0(X) - \mu_0(Z))^4 < \infty. \]

Under (1.1), (2.4), and \( H \), \( n^{1/2}|\Delta_n| = O_p(1) \). Moreover,
\[ \sup_z |T_{n2}(z)| \leq n^{-1} \sum_{i=1}^{n} |\xi_i| \left| \frac{1}{d_{ni}} - \frac{1}{2\sigma_0^2(Z_i)} \right| \leq \max_{1 \leq i \leq n} \frac{2\sigma_0^2(Z_i) - d_{ni}}{d_{ni}\sigma_0^2(Z_i)} n^{-1} \sum_{i=1}^{n} |\xi_i| \]

Now note that from the definitions,
\[ d_{ni}\sigma_0^2(Z_i) = \sigma_0 \left\{ \left( s_{n0}^2 + \tau_0^2(Z_i) \right)^{1/2} + \left( \sigma_z^2 + \tau_0^2(Z_i) \right)^{1/2} \right\} \geq \sigma_z^4 > 0, \]
\[ |2\sigma_0^2(Z_i) - d_{ni}| = |\sigma_0^2(Z_i) - \sigma_0^2 n_0(Z_i) + \sigma_0(Z_i)(\sigma_0(Z_i) - \sigma_0 n_0(Z_i))| \leq 2|\Delta_n|, \quad \forall \ 1 \leq i \leq n. \]

Hence,
\[ \max_{1 \leq i \leq n} \frac{|2\sigma_0^2(Z_i) - d_{ni}|}{d_{ni}\sigma_0^2(Z_i)} \leq 2\sigma_z^{-4}|\Delta_n| = o_p(1). \]

These facts together with the fact that \( n^{-1} \sum_{i=1}^{n} |\xi_i| = O_p(1) \) yield
\[ (2.5) \quad \sup_z |T_{n2}(z)| = o_p(1). \]

Next consider, \( T_{n1}(z) \). Let \( L_0(z) := E\sigma_0^{-2}(Z)I(Z \leq z) \). Because \( \sigma_0^2(Z) \geq \sigma_z^2 > 0 \), for all \( z \in \mathbb{R} \), the measure induced by \( L_0 \) is a finite measure on \( \mathbb{R} \). Also, \( ET_{n1}(z) \equiv 0, \text{Var}(T_{n1}(z)) = n^{-1}L_0(z) \leq n^{-1}\sigma_z^{-2} \rightarrow 0 \), for all \( z \in \mathbb{R} \). This fact and a Glivenko-Cantelli type argument, where the interval is partitioned according to the measure induced by \( L_0(z) \), shows that \( \sup_{z \in \mathbb{R}} |T_{n1}(z)| = o_p(1) \). We summarize these observations in

**Theorem 2.1** Suppose the model (1.1) and \( H \) hold. In addition, assume that (2.4) holds. Then,
\[ (2.6) \quad \sup_{z \in \mathbb{R}} |V_0^0(z) - S_n(z)| = \sup_{z \in \mathbb{R}} |W_n(z)| = o_p(1), \]
\[ (2.7) \quad V_n^0 \Rightarrow B \circ G, \quad \text{in } D(\mathbb{R}) \text{ and uniform metric.} \]
To establish the consistency of the test (1.3) against a fixed alternative \( H_1: \mu = \mu_1 \), proceed as follows. Let \( E_1 \) denote the expectation under \( H_1 \). Assume \( \mu_1 \) satisfies \( \mu_1^2(X) < \infty \). Let \( \nu_1(z) := E(\mu_1(X)|Z = z) \) and suppose, additionally, that

\[
0 < \sigma^2_1 := E_1(Y - \nu_1(Z))^2 < \infty, \quad d := \sup_{z \in \mathbb{R}} \left| \frac{\nu_1(Z) - \nu_0(Z)}{\sigma_0(Z)} I(Z \leq z) \right| \neq 0.
\]

Write \( V^0_n = V^1_n + n^{1/2}\tilde{D}_n \), where

\[
V^1_n(z) := n^{-1/2} \sum_{i=1}^{n} \frac{Y_i - \nu_1(Z_i)}{\sigma_0(Z_i)} I(Z_i \leq z), \quad \tilde{D}_n(z) := n^{-1} \sum_{i=1}^{n} \frac{\nu_1(Z_i) - \nu_0(Z_i)}{\sigma_0(Z_i)} I(Z_i \leq z).
\]

Let \( D_n \) denote the average like \( \tilde{D}_n \) but with \( \sigma_0(Z_i)'s \) replaced by \( \sigma_0(Z_i)'s \). An argument similar to the one that yielded (2.5) shows that \( \sup_z |\tilde{D}_n(z) - D_n(z)| = o_p(1) \). By a Glivenko-Cantelli type argument, we also have \( \sup_z |D_n(z)| \to d \), a.s.

Hence, \( \sup_z |\tilde{D}_n(z)| \to d \), in probability.

Let

\[
S^1_n(z) := n^{-1/2} \sum_{i=1}^{n} \frac{Y_i - \nu_1(Z_i)}{\sigma_0(Z_i)} I(Z_i \leq z), \quad \psi(z) := E_1 \left( \frac{Y - \nu_1(Z)}{\sigma^2_1} \right)^2 I(Z \leq z), \quad z \in \mathbb{R}.
\]

Note that \( \psi(z) \leq \sigma^2_\psi \sigma^2_1 < \infty \), for all \( z \in \mathbb{R} \). By (2.1) applied to \( \xi_i = (Y_i - \nu_1(Z_i))/\sigma_0(Z_i) \), we have \( L = \psi, S^1_n = U_n \), and hence, \( S^1_n \Rightarrow B \circ \psi \) under \( H_1 \). Moreover, an argument like the one used in concluding (2.6) shows that \( \sup_z |V^1_n(z) - S^1_n(z)| = o_p(1) \), under \( H_1 \), so that we also have \( V^1_n \Rightarrow B \circ \psi \), under \( H_1 \). These facts and a routine argument show that the test (1.3) is consistent against \( H_1 \), under (2.4) and (2.8).

Moreover, the asymptotic power of this test against the local nonparametric \( n^{-1/2} \)-alternatives \( \mu(x) = \mu_0(x) + n^{-1/2} \delta(x) \), where \( \delta(x) \neq 0 \) and \( E\delta^2(X) < \infty \), can be shown to be \( P(\sup_{0 \leq t \leq 1} |B(t) + \int_0^{G^{-1}(t)} (\delta/\sigma_0)(G^{-1}(u)) du| > b_n) \). Similar facts can be established for any test based on a continuous function of \( V^0_n \).

### 2.2 Tests of \( H_0 \)

For the sake of completeness we shall first describe a generic transformation of a vector of functions that preserves Brownian motion and that is orthogonal to certain kinds of drifts in some Gaussian processes. This transformation has its roots in Khmaladze (1981, 1993), STZ and Khmaladze and Koul (2004). Then we state the needed assumptions on \( M \) and other underlying entities under which the application of this generic transformation to the process \( \hat{V}_n \) and its weak convergence to a time transformed Brownian motion is justified. These results in turn are then used to construct ADF tests for \( H_0 \) based on the process \( \hat{V}_n \).
To describe the generic transformation, let $U$ be a continuous r.v. with d.f. $K$ on $\mathbb{R}$ and $\ell$ be a vector of $q$ functions such that $E\|\ell(U)\|^2 < \infty$. Let

$$C_u := E\ell(U)\ell(U)\text{'}I(U \geq u), \quad u \in \mathbb{R}. $$

Suppose $C_u$ is positive definite for all $u \in \mathbb{R}$. Let $C_u^{-1}$ denote its inverse matrix. For a function $\gamma$ from $\mathbb{R}$ to $\mathbb{R}$, define the operators

$$T\gamma(u) := \int_{y \leq u} \gamma(y)\ell(y)C_y^{-1}dK(y)\ell(u), \quad K\gamma(u) := \gamma(u) - T\gamma(u), \quad u \in \mathbb{R},$$

We have the following

**Lemma 2.1** Let $\mathcal{L} := \{\gamma \in L_2(\mathbb{R}, K) : \int \gamma dK = 0\}$. Then the transformation $K$ is norm preserving from $L_2(\mathbb{R}, K)$ to $\mathcal{L}$:

$$\int K\gamma(u)\ell(u)\text{'}dK(u) = 0, \quad \int (K\gamma(u))^2 dK(u) = \int \gamma^2(u) dK(u).$$

Moreover, for any $\gamma_1, \gamma_2 \in L_2(\mathbb{R}, K)$,

$$E\{K\gamma_1(U)K\gamma_2(U)\} = E\{\gamma_1(U)\gamma_2(U)\}. $$

**Proof.** We have

$$\int T\gamma(u)\ell(u)\text{'}dK(u) = \iint_{y \leq u} \gamma(y)\ell(y)C_y^{-1}dK(y)\ell(u)\ell(u)\text{'}dK(u)$$

$$= \int \gamma(y)\ell(y)C_y^{-1}C_ydK(y), \quad \text{(by Fubini)}$$

$$= \int \gamma(y)\ell(y)\text{'}dK(y).$$

This proves the first claim of (2.9), because $K\gamma = \gamma - T\gamma$.

The second claim of (2.9) will follow from (2.10) which we prove next. By Fubini, the left hand side of (2.10) can be written as $\int \gamma^2 dK - A_1 - A_2 + B$, where,

$$A_1 := \iint_{y \leq u} \gamma_1(y)\gamma_2(u)\ell(y)C_y^{-1}\ell(u)dK(y)dK(u),$$

$$A_2 := \iint_{y \leq u} \gamma_1(u)\gamma_2(y)\ell(y)C_y^{-1}\ell(u)dK(y)dK(u),$$

$$B := \iint \gamma_1(y)\gamma_2(u)\ell(y)C_y^{-1}C_y\gamma\gamma C_u^{-1}\ell(u)dK(u)dK(y) = A_1 + A_2.$$ 

The last equality is obtained by splitting the range of integration into the two parts, $y \leq u$ and $y \geq u$, and realizing that the sum of these two integrals is $A_1 + A_2$ because $K$ is continuous. This proves (2.10).
Next, let $\zeta$ be a r.v. such that $E(\zeta|U) = 0$, $\tau^2(u) := E(\zeta^2|U = u) > 0$, for all $u$. Then from (2.10) we obtain that the process $W_\gamma(\zeta, U) := [\zeta/\tau(U)]K\gamma(U)$, as a process in $\gamma \in L_2(\mathbb{R}, K)$, is like $B_\gamma(K)$, where $B_\gamma$ is a Brownian motion in $\gamma$. Consequently, if $(\zeta_i, U_i)$, $1 \leq i \leq n$, are i.i.d. copies of $(\zeta, U)$, then by the classical CLT, the finite dimensional distributions of $n^{-1/2} \sum_{i=1}^n W_\gamma(\zeta_i, U_i)$, as $\gamma$ varies, will converge weakly to those of $B_\gamma(K)$. These are the basic observations that are useful in the construction of ADF tests for $H_0$ based on the process $\hat{V}_n$.

We now turn to the basic problem of testing $H_0$. Throughout, the true parameter $\theta_0$ under $H_0$ is assumed to be in the interior of $\Theta$. Consider the following assumptions.

(e) $Ee^4 + E(\mu_{\theta_0}(X) - \nu_{\theta_0}(Z))^4 < \infty$.

(m) For some positive continuous function $r(x)$ with $E r^4(X) < \infty$, the following holds:

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq \|\theta_1 - \theta_2\| r(x).$$

($\nu_1$) For every $z$, $\nu_{\theta}(z)$ is differentiable in $\theta$ in a neighborhood of $\theta_0$ with the vector of derivatives $\hat{\nu}_{\theta}(z)$, such that for every $0 < k < \infty$,

$$\sup_{1 \leq i \leq n, n^{1/2} |\theta - \theta_0| \leq k} n^{1/2} |\nu_{\theta_0}(Z_i) - \nu_{\theta_0}(Z_i) - (\theta - \theta_0)' \hat{\nu}_{\theta_0}(Z_i)| = o_p(1), \quad (P_{\theta_0}).$$

($\nu_2$) For some $q \times q$ square matrix $\hat{\nu}(z, \theta_0)$ and a nonnegative function $g(z, \theta_0)$, both measurable in the $z$-coordinate, the following holds:

$$E\|\hat{\nu}(z, \theta_0)\|^j g(z, \theta_0) < \infty, \quad E\|\hat{\nu}(z, \theta_0)\|\|\hat{\nu}(z, \theta_0)\|^j < \infty, \quad j = 0, 1$$

and for all $\delta > 0$, there exists an $\eta > 0$ such that $\|\theta - \theta_0\| < \eta$ implies

$$\|\hat{\nu}(z, \theta) - \hat{\nu}(z, \theta_0) - \hat{\nu}(z, \theta_0)(\theta - \theta_0)\| \leq \delta g(z, \theta_0)\|\theta - \theta_0\|, \quad \text{for } G\text{-almost all } z.$$

($\nu_3$) $E\|\hat{\nu}_{\theta_0}(Z)\|^2 < \infty$, and $L_{\theta_0}(z) := \int_{y \geq z} \sigma_{\theta_0}^{-2}(y)\hat{\nu}_{\theta_0}(y)\hat{\nu}_{\theta_0}(y)'dG(y)$ is positive definite for all $z \in \mathbb{R}$.

We note here that ($\nu_3$) and (1.5) imply that $E\|\hat{\nu}_{\theta_0}(Z)/\sigma_{\theta_0}(Z)\|^2 \leq \sigma_{\theta_0}^{-2}E\|\hat{\nu}_{\theta_0}(Z)\|^2 < \infty$ and the matrix $\Gamma_{\theta_0} := \int \sigma_{\theta_0}^{-2}(y)\hat{\nu}_{\theta_0}(y)\hat{\nu}_{\theta_0}(y)'dG(y)$ is positive definite.

The conditions ($\nu_1$)-($\nu_3$) are trivially satisfied by the model $m_{\theta}(x) = \theta'h(x)$, where $h = (h_1, \ldots, h_q)'$ is vector of real $q$-function such that $E\|h(X)\|^2 < \infty$, and the matrix $\int_{y \geq z} E(h(X)|Z = y)E(h(X)|Z = y)'dG(y)$ is positive definite for all $z \in \mathbb{R}$.

A set of somewhat stronger conditions in terms of the given model $M$ that imply ($\nu_1$)-($\nu_3$) are as follows.

(m1) The function $m_{\theta}(x)$ is differentiable in $\theta$ in a neighborhood of $\theta_0$, with the vector of differential
\( \hat{m}_{\theta_0} \) such that \( E\|\hat{m}_{\theta_0}(X)\|^2 < \infty \) and for every \( k < \infty \),

\[
\sup_{x \in \mathbb{R}, n^{1/2} \|\theta - \theta_0\| \leq k} n^{1/2} |m_\theta(x) - m_{\theta_0}(x) - (\theta - \theta_0) \dot{m}_{\theta_0}(x)| = o_p(1), \quad (P_{\theta_0}).
\]

(m2) For some \( q \times q \) square matrix \( \dot{m}_{\theta_0}(x) \) and a nonnegative function \( g_1(x, \theta_0) \), both measurable in the \( x \)-coordinate, the following holds:

\[
E\{\|\dot{m}_{\theta_0}(X)\|^2 E(g_1(X, \theta_0)|Z)\} < \infty, \quad E\{\|\dot{m}_{\theta_0}(X)\| E(\dot{m}_{\theta_0}(X)|Z)\} < \infty, \quad j = 0, 1,
\]

and for all \( \delta > 0 \), there exists a \( \eta > 0 \) such that \( \|\theta - \theta_0\| < \eta \) implies

\[
\|\dot{m}_{\theta}(x) - \dot{m}_{\theta_0}(x) - \dot{m}_{\theta_0}(x)(\theta - \theta_0)\| \leq \delta g_1(x, \theta_0)\|\theta - \theta_0\|, \quad \forall x.
\]

Because \( \nu_\theta(z) = \int m_\theta(y) f_\theta(y - z) dy \), and \( f_\theta(y - z)dy \equiv 1 \), (m1) readily implies \((\nu 1)\) with \( \dot{\nu}_{\theta_0}(z) \equiv E(\dot{m}_{\theta_0}(X)|Z = z) \). Also, by the Cauchy-Schwarz inequality applied to the conditional expectation, given \( Z \), \( E\|E(\dot{m}_{\theta_0}(X)|Z)\|^2 \leq E\|\dot{m}_{\theta_0}(X)\|^2 < \infty \), by (m1), so that the first condition of \((\nu 3)\) holds. Similarly, (m2) readily implies \((\nu 2)\) with \( \dot{\nu}_{\theta_0}(z) \equiv E(\dot{m}_{\theta_0}(X)|Z = z) \) and \( g(z, \theta_0) = E(g_1(X, \theta_0)|Z = z) \).

We are now ready to describe an asymptotically distribution free test of \( H_0 \). Apply the transformation \( K \) to \( \ell = \dot{\nu}_{\theta_0}/\gamma_{\theta_0}; U = Z; \xi = (Y - \nu_{\theta_0}(Z))/\gamma_{\theta_0}(Z) \), \( K = G \), and with \( \gamma(u) := I(u \leq z) \), \( z \in \mathbb{R} \). Note that now \( C_y = L_{\theta_0}(y) \). Denote the corresponding \( K\gamma \) by \( K_{\theta_0}(\xi, Z, G) \). In view of the above assumptions, this is well defined and

\[
K_{\theta_0}(\xi, Z, G) := \xi[I(Z \leq z) - \int_{y \leq z} \frac{\dot{\nu}_{\theta_0}(y)}{\gamma_{\theta_0}(y)} L_{\theta_0}^{-1}(y) \dot{\nu}_{\theta_0}(Z) I(Z \geq y) dG(y)].
\]

Let \( \xi_i := (Y_i - \nu_{\theta_0}(Z_i))/\gamma_{\theta_0}(Z_i) \) and define

\[
W_{\theta_0,G}(z) := n^{-1/2} \sum_{i=1}^n K_{\theta_0}(\xi_i, Z_i, G), \quad z \in \mathbb{R}.
\]

From the above discussion and the classical CLT we readily obtain that all finite dimensional distributions of \( W_{\theta_0,G}(z) \) converge weakly to those of \( B \circ G \). However, because this process depends on the parameters \( \theta_0 \) and \( G \), it is not useful for inference. The process that is useful for testing for \( H_0 \) is \( W_n(z) := W_{\theta_0,G_n}(z) \), \( z \in \mathbb{R} \), where \( G_n \) is the empirical of \( Z_i, 1 \leq i \leq n \). Let \( \xi_{ni} := (Y_i - \nu_{\theta_n}(Z_i))/\gamma_{\theta_n}(Z_i), 1 \leq i \leq n \). Recalling the definition (1.6) and of \( \dot{V}_n \), one can see that

\[
W_n(z) = n^{-1/2} \sum_{i=1}^n K_{\theta_n}(\xi_{ni}, Z_i, G_n)
= \dot{V}_n(z) - \int_{x \leq z} \frac{\dot{\nu}_{\theta_n}(x)}{\gamma_{\theta_n}(x)} L_{\theta_n}^{-1}(x) \int_{y \geq x} \frac{\dot{\nu}_{\theta_n}(y)}{\gamma_{\theta_n}(y)} \dot{V}_n(dy) dG_n(x), \quad z \in \mathbb{R}.
\]
Similar to the classical regression case discussed in STZ and Koul (2002), the matrices $L_{\theta_n}^{-1}(x)$ are unstable for $x$ in the right tail closer to infinity and often not uniformly continuous in the underlying parameter $\theta$, so, for a given sample size, the processes $W_n(z)$ may become very unstable for an arbitrarily large $z$. Hence, we need to assume that

(2.11) \quad \text{For some } z_0 < \infty, L_{\theta_{z_0}}(z_0) \text{ is non-singular.}

Consequently, we have to restrict $W_n(z)$ to the intervals $[-\infty, z_0]$. A practical choice of $z_0$ depends on the data. The following theorem gives the needed weak convergence result.

**Theorem 2.2** Under the above conditions and (2.11), and under $H_0$, $W_n \Rightarrow B \circ G$, in $D([-\infty, z_0])$ and uniform metric.

A proof of this theorem is given in the last section. As a consequence, the test that rejects $H_0$ whenever $\sup_{z \leq z_0} |W_n(z)| / 0.995 > b_\alpha$ will be of the asymptotic size $\alpha$. As in STZ, it is recommended take $z_0$ to be the 99% quantile of the $Z$-data in applications.

### 3 Simulations

This section contains a simulation study of the proposed model check for two cases: **Case 1**: $q = 2$ and $m_\theta$ is linear; **Case 2**: $q = 2$ and $m_\theta$ is nonlinear. In all cases, $\{Z_i\}_{i=1}^n$ are generated at random from the uniform distribution on $[-1, 1]$, $\{\varepsilon_i\}_{i=1}^n$ and $\{u_i\}_{i=1}^n$ are obtained as two independent random samples from $\mathcal{N}(0, (0, 1)^2)$. Then $(X_i, Y_i)$ are generated using the model $Y_i = m_\theta(X_i) + \varepsilon_i$, $X_i = Z_i + u_i$, $i = 1, 2, \cdots, n$. In each case we use the least square estimator of $\theta$, and the significance level is taken to be 0.05. The sample sizes $n$ chosen are 50, 100, 200, 300 and 500, each simulation being repeated 1000 times. The test statistic is $\hat{D}_n = \sup_{z \leq z_0} |W_n(z)| / 0.995$ with $z_0$ being the 99% quantile of the $Z$-data. Under $H_0$, $\sup_{z \leq z_0} |W_n(z)| / 0.995 \Rightarrow \sup_{0 \leq t \leq 1} |B(t)|$. Using the well known fact

$$P(\sup_{0 \leq t \leq 1} |B(t)| < b) = P(|B(1)| < b) + 2 \sum_{i=1}^{\infty} (-1)^i P((2i - 1)b < B(1) < (2i + 1)b),$$

the critical value corresponding to the significance level 0.05 is 2.24241. The empirical size and power are computed by using $\#\{\hat{D}_n > 2.24241\}/1000$.

**Case 1.** The data were simulated from the following four models.

- **Model 0**: $Y_i = 1 + 2X_i + \varepsilon_i$.
- **Model 1**: $Y_i = 1 + 2X_i + 1.4 \exp(-0.2X_i^2) + \varepsilon_i$.
- **Model 2**: $Y_i = 1 + 2X_i + 0.3X_i^2 + \varepsilon_i$.
- **Model 3**: $Y_i = 1 + 2X_i I(X_i \geq 0.2) + \varepsilon_i$. 

11
Data from Model 0 are used to study the empirical sizes, and from Models 1 to 3 are used to study
the empirical powers of the test.

From the simulation results in Table 1, one sees that the empirical level is close to the nominal
level of 0.05 with the increase in the sample size. It also shows that the test performs very well for
sample sizes larger than 200 at all the three alternatives.

<table>
<thead>
<tr>
<th>Model \ n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.017</td>
<td>0.031</td>
<td>0.031</td>
<td>0.043</td>
<td>0.048</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.258</td>
<td>0.704</td>
<td>0.975</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.446</td>
<td>0.827</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 1: Levels and powers of the test

Case 2. In this case, the data were simulated from the following four models.

Model 0: \( Y_i = 2X_i + \exp(X_i) + \varepsilon_i \),

Model 1: \( Y_i = 2X_i + \exp(X_i) + 0.5X_i^2[\exp(-0.2X_i) + \exp(1.2X_i^2)] + \varepsilon_i \),

Model 2: \( Y_i = 2X_i + \exp(X_i) + 0.5 + \varepsilon_i \),

Model 3: \( Y_i = 2X_i + \exp(X_i) + 0.5\exp(-0.2X_i) + \varepsilon_i \).

Data from Model 0 are used to study the empirical sizes, and from Models 1 to 3 are used to study
the empirical powers of the test. The results of the simulation study are shown in the Table 2. It
shows that the test performs very well.

<table>
<thead>
<tr>
<th>Model \ n</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.018</td>
<td>0.030</td>
<td>0.033</td>
<td>0.044</td>
<td>0.046</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.456</td>
<td>0.745</td>
<td>0.913</td>
<td>0.950</td>
<td>0.992</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.719</td>
<td>0.976</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.736</td>
<td>0.982</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 2: Levels and powers of the test

4 Some Proofs

Here we sketch a proof of Theorem 2.2. The transform \( W_n \) is an analog of the one given in STZ
with a major difference. In the no measurement error set up, STZ uses a part of the sample to
estimate the conditional variance and the other part is used to implement their test. But here
because of (1.5) we estimate this variance based the entire sample. This then means that some of
the arguments of the proofs though similar are necessarily different.

The proof consists of the following two steps.

(a) \[ \sup_{z \leq z_0} |W_n(z) - \mathcal{W}_{\theta_0, G}(z)| = o_p(1), \quad (P_{\theta_0}). \]

(b) \[ \mathcal{W}_{\theta_0, G} \Rightarrow B \circ G, \quad \text{in } D([-\infty, z_0]) \text{ and in the uniform metric.} \]

Step (b) has been proven in Section 2, so it suffices to prove step (a) only. To begin with rewrite
\( V_n(z) \) for \( V_n(z, \theta_0) \), and \( \nu_n, \nu_n, \nu_n, \tau_n, \sigma_n, L_n \) for \( \nu_{\theta_n}, \nu_{\theta_n}, \nu_{\theta_n}, \tau_{\theta_n}, \sigma_{\theta_n}, L_{\theta_n} \), respectively. Then, by
direct algebra, we obtain

\[ \mathcal{W}_{\theta_0, G}(z) = V_n(z) - \int_{x \leq z} \frac{\hat{\nu}_{\theta_0}(x)}{\sigma_{\theta_0}} L_{\theta_0}^{-1}(x) \int y \geq x \frac{\hat{\nu}_{\theta_0}(y)}{\sigma_{\theta_0}} V_n(dy) dG(x), \]
\[ \mathcal{W}_n(z) = \hat{V}_n(z) - \int_{x \leq z} \frac{\hat{\nu}_{\sigma_n}(x)}{\sigma_n} L_n^{-1}(x) \int y \geq x \frac{\hat{\nu}_n(y)}{\sigma_n} \hat{V}_n(dy) dG_n(x). \]

In view of the assumption

\[ ||n^{1/2}(\theta_n - \theta_0)|| = O_p(1), \]

we shall briefly prove that for some 0 < b < \infty,

\[ \sup_{1 \leq i \leq n} |\sigma_n^2(Z_i) - \sigma_{\theta_0}^2(Z_i)| = o_p(1). \]

To see this, note that \( \sigma_n^2(Z_i) - \sigma_{\theta_0}^2(Z_i) = \tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i) + s_n^2 - \sigma_n^2, \) where \( s_n^2 \) is defined in (1.5).

Since \( s_n^2 - \sigma_n^2 = o_p(1), \) so to prove (4.3) it suffice to show

\[ \sup_{1 \leq i \leq n} |\tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i)| = o_p(1). \]

Using (m), after some calculations, we can show that for some constant c,

\[ |\tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i)| \leq c||\theta_n - \theta_0||^2 \int r^2(y)f_n(y - Z_i)dy. \]
Also from (m), we know $Er^4(X) < \infty$, therefore, $\max_{1 \leq i \leq n} | \int r^2(y) f_\eta(y - Z_i) dy | = o_p(n^{1/2})$. Together with the fact (4.1), we see that (4.4), and hence (4.3), holds.

Now, consider the third term in the r.h.s. of (4.2). We shall show it is of the order $u_p(1)$. Using (1.6), the third term in the r.h.s. of (4.2) can be written as the sum

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma_n^2 - \sigma_n^2) \zeta_i I(Z_i \leq z) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{\tau^2_n(Z_i) - \tau^2_n(Z_i)}{(\sigma_n(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)} \right) I(Z_i \leq z).
$$

But the first term in (4.6) can be written as

$$
\sqrt{n}(\sigma^2_n - \sigma_n^2) \left[ \frac{1}{n} \sum_{i=1}^n \zeta_i I(Z_i \leq z) \left( \frac{1}{(\sigma_n(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)} - \frac{1}{2\sigma_n^2(Z_i)} \right) + \frac{1}{n} \sum_{i=1}^n \zeta_i I(Z_i \leq z) \right].
$$

By (4.3),

$$
\max_{1 \leq i \leq n} \left| \frac{1}{(\sigma_n(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)} - \frac{1}{2\sigma_n^2(Z_i)} \right| = o_p(1).
$$

By a Glivenko-Cantelli type argument, $\sup_{z \in \mathbb{R}} | n^{-1} \sum_{i=1}^n \zeta_i I(Z_i \leq z) / 2\sigma_n^2(Z_i) | = o_p(1)$. Also, the condition (e) ensures that $\sqrt{n}(\sigma^2_n - \sigma_n^2) = O_p(1)$. These facts together then imply that the first term in (4.6) is of the order $u_p(1)$.

From (4.5), the second term in (4.6) is bounded above by

$$
c\sqrt{n}||\theta_n - \theta_0||^2 \frac{1}{2n\sigma^2_n} \sum_{i=1}^n |\zeta_i| \int r^2(y) f_\eta(y - Z_i) dy
$$

which is of the order $u_p(1)$. Thus we have proved that the third term in the r.h.s. of (4.2) is of the order $u_p(1)$. The fourth term in the r.h.s. of (4.2) is bounded above by the l.h.s. of the assumption $(\nu_1)$ multiplied by $\sigma_n^{-1} \sqrt{n}(\theta_n - \theta_0)$, and hence $u_p(1)$. Similarly, $(\nu_3)$, (4.1) and (4.3) imply that the fifth term is $u_p(1)$.

These facts and a Glivenko-Cantelli type argument shows that

$$
\sup_{z \leq z_0} \left| \hat{V}_n(z) - V_n(z) - \sqrt{n}(\theta_n - \theta_0)' \frac{E}{\sigma_0(Z)} \hat{\theta}_0(Z) I(Z \leq z) \right| = o_p(1).
$$

For convenience, let $A_{n1}(z)$ denote the second term of $W_{\theta_0,G}(z)$ without the negative sign, $A_{n2}(z)$ denote the second term of $W_n(z)$ without the negative sign, also, let

$$
U_n(x) = \int_{y \geq x} \hat{\theta}_0(y) \sigma_0(y) V_n(dy), \quad \hat{U}_n(x) = \int_{y \geq x} \hat{\theta}_0(y) \hat{\theta}_0(dy).
$$

Then

$$
A_{n1}(z) = \int_{x \leq z} \frac{\hat{\theta}'_0(x)}{\sigma_0(x)} L_{\theta_0}^{-1}(x) U_n(x) G(dx), \quad A_{n2}(z) = \int_{x \leq z} \frac{\hat{\theta}'_0(x)}{\sigma_0(x)} L_{\theta_0}^{-1}(x) \hat{U}_n(x) G_n(dx).
$$

By the definition of $\hat{V}_n(y)$ and $V_n(y)$, $\hat{U}_n(x) - U_n(x)$ can be written as
This can be further decomposed into the sum of the following seven terms,

\[ B_{n1}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left( \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_n(Z_i)) - \frac{\hat{\theta}_0(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right). \]

\[ B_{n2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right] \]

\[ B_{n3}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right] \]

\[ B_{n4}(x) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \right] \]

\[ B_{n5}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right] \]

\[ B_{n6}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right] \]

\[ B_{n7}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \left[ \frac{\hat{\theta}_n(Z_i) - \nu_0(Z_i)}{\sigma^2_{\theta_0}(Z_i)} \cdot (Y_i - \nu_0(Z_i)) \right]. \]

Subtracting and adding \( \hat{\theta}_0(Z_i)(\theta_n - \theta_0) \) from \( \hat{\theta}_n(Z_i) - \hat{\theta}_0(Z_i) \), \( B_{n2}(x) \) can be written as

\[ B_{n2}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i I[Z_i \geq x] \frac{\hat{\theta}_n(Z_i) - \hat{\theta}_0(Z_i) - \nu_0(Z_i)(\theta_n - \theta_0)}{\sigma^2_{\theta_0}(Z_i)} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i I[Z_i \geq x] \frac{\hat{\theta}_0(Z_i)(\theta_n - \theta_0)}{\sigma^2_{\theta_0}(Z_i)} \]

(4.9)

By condition (\( \nu_2 \)), the first term in the r.h.s. of (4.9) is bounded above by

\[ \sqrt{n}\|\theta_n - \theta_0\| \cdot \varepsilon \gamma^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} |\zeta_i| K(Z_i, \theta_0). \]

This bound is of the order \( u_p(1) \) by the arbitrariness of \( \varepsilon \), (4.1) and (\( \nu_2 \)). Note that the empirical process

\[ \frac{1}{n} \sum_{i=1}^{n} \zeta_i I[Z_i \geq x] \frac{\hat{\theta}_0(Z_i)(Y_i - \nu_0(Z_i))}{\sigma^2_{\theta_0}(Z_i)} \]

is of order \( o_p(1) \) for any \( x \) and tight, then from (4.1), the second term in (4.9) is also of the order \( u_p(1) \). So \( B_{n2}(x) = u_p(1) \). Similarly one can show \( B_{n1}(x) = u_p(1) \).

Subtracting and adding \( \hat{\theta}_0(Z_i)(\theta_n - \theta_0) \), \( (\theta_n - \theta_0)\hat{\theta}_0(Z_i) \) from \( \hat{\theta}_n(Z_i) - \hat{\theta}_0(Z_i) \) and \( \hat{\theta}_n(Z_i) - \hat{\theta}_0(Z_i) \) respectively, \( B_{n4}(x) \) can be written as the following sum,
Therefore, we get (4.3).

This upper bound is of the order $1 - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \|\theta_n - \theta_0\| \frac{2\varepsilon\sigma^2 n^{-1}}{\sigma^2_{\theta_0}(Z_i)}$.

The absolute value of the above sum is bounded above by

$$\|\theta_n - \theta_0\| \frac{2\varepsilon\sigma^2 n^{-1}}{\sigma^2_{\theta_0}(Z_i)} \sum_{i=1}^{n} K(Z_i, \theta_0)$$

$$+ \sqrt{n}\|\theta_n - \theta_0\| \frac{2\varepsilon\sigma^2 n^{-1}}{\sigma^2_{\theta_0}(Z_i)} \sum_{i=1}^{n} K(Z_i, \theta_0) \|\hat{\nu}_{\theta_0}(Z_i)\|$$

$$+ \|\theta_n - \theta_0\| \sum_{i=1}^{n} \|\hat{\nu}_{\theta_0}(Z_i)\| \|\hat{\nu}_{\theta_0}(Z_i)\|$$

$$+ \sqrt{n}\|\theta_n - \theta_0\| \frac{2\varepsilon\sigma^2 n^{-1}}{\sigma^2_{\theta_0}(Z_i)} \sum_{i=1}^{n} \|\hat{\nu}_{\theta_0}(Z_i)\| \|\hat{\nu}_{\theta_0}(Z_i)\|.$$

This upper bound is of the order $u_p(1)$ by conditions $(\nu_1)$ and $(\nu_2)$. Hence $B_{n3}(x) = u_p(1)$. Same result is held for $B_{n3}(x)$.

Using the same method as in proving the $u_p(1)$ order of the third term in the r.h.s. of (4.2), we can show $B_{n5}(x)$ is of the order $u_p(1)$. $B_{n5}(x) = u_p(1)$ can be proved by using condition $(\nu_1)$ and (4.3).

Now consider $B_{n7}(x)$. Using condition $(\nu_1)$ and $(\nu_3)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \frac{\hat{\nu}_{\theta_0}(Z_i)}{\frac{1}{\sigma^2_{\theta_0}(Z_i)}} (\nu_{\theta_0}(Z_i) - \nu_n(Z_i))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \frac{\hat{\nu}_{\theta_0}(Z_i)}{\frac{1}{\sigma^2_{\theta_0}(Z_i)}} (\nu_{\theta_0}(Z_i) - \nu_n(Z_i) + (\theta_n - \theta_0) \hat{\nu}_{\theta_0}(Z_i))$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I[Z_i \geq x] \frac{\hat{\nu}_{\theta_0}(Z_i) \hat{\nu}_{\theta_0}'(Z_i)}{\frac{1}{\sigma^2_{\theta_0}(Z_i)}} (\theta_n - \theta_0)$$

$$= -L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) + u_p(1).$$

Therefore, we get

$$\sup_{x \leq \xi_0} \|\hat{U}(x) - U(x) + L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0)\| = u_p(1).$$

(4.10)
Now consider the asymptotic behavior of the matrix $L_n(x)$. Note that

$$
\left\| L_n(x) - L_{\theta_0}(x) \right\| = \left\| \int_{y \geq x} \frac{\nu'(y)\nu''(y)}{\sigma_n^2(y)} G_n(dy) - L_{\theta_0}(x) \right\|
$$

$$
= \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i'(Z_i)\nu_i''(Z_i)}{\sigma_n^2(Z_i)} I[Z_i \geq x] - L_{\theta_0}(x) \right\|
$$

$$
\leq \frac{1}{n} \sum_{i=1}^{n} \left\| \nu_i'(Z_i) - \nu_{\theta_0}(Z_i) \right\|^2 + \frac{2}{n} \sum_{i=1}^{n} \left\| \nu_{\theta_0}(Z_i) \right\| \left\| \nu_i'(Z_i) - \nu_{\theta_0}(Z_i) \right\|
$$

$$
+ \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\nu_{\theta_0}(Z_i)\nu_i''(Z_i)}{\sigma_n^2(Z_i)} \left( \frac{\sigma_n^2(Z_i)}{\sigma_{\theta_0}^2(Z_i)} - 1 \right) \right\| + \left\| \frac{1}{n} \sum_{i=1}^{n} \nu_{\theta_0}(Z_i)\nu_i'(Z_i) I[Z_i \geq x] - L_{\theta_0}(x) \right\|.
$$

Using conditions (ν2) and (4.3), we can show the first three terms are $u_p(1)$. The last term is also $u_p(1)$ by showing the tightness of $n^{-1} \sum_{i=1}^{n} \nu_{\theta_0}(Z_i)\nu_i'(Z_i) I[Z_i \leq x]/\sigma_{\theta_0}^2(Z_i)$. Consequently, we have

$$
(4.11) \quad \sup_{x \leq z_0} \left\| L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right\| = o_p(1).
$$

Finally, let’s prove the following result,

$$
(4.12) \quad \sup_{x \leq z_0} \left| A_{n1}(z) - A_{n2}(z) + \sqrt{n}(\theta_n - \theta_0) \right| \int_{x \leq z} \frac{\nu_{\theta_0}(x)}{\sigma_{\theta_0}(x)} G_n(dx) = o_p(1),
$$

where $A_{n1}(z)$ and $A_{n2}(z)$ are defined in (4.8). In fact, $A_{n2}(z)$ can be written as the sum

$$
\int_{x \leq z} \left( \frac{\nu_n'(x)}{\sigma_n(x)} - \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \right) \left( L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right) \left( \hat{U}_n(x) - U_n(x) + L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

$$
+ \int_{x \leq z} \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \left( L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right) \left( \hat{U}_n(x) - U_n(x) + L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

$$
+ \int_{x \leq z} \left( \frac{\nu_n'(x)}{\sigma_n(x)} - \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \right) L_{\theta_0}^{-1}(x) \left( \hat{U}_n(x) - U_n(x) + L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

$$
+ \int_{x \leq z} \left( \frac{\nu_n'(x)}{\sigma_n(x)} - \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \right) \left( L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right) \left( U_n(x) - L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

$$
+ \int_{x \leq z} \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \left( L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right) \left( U_n(x) - L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

$$
+ \int_{x \leq z} \left( \frac{\nu_n'(x)}{\sigma_n(x)} - \frac{\nu_{\theta_0}'(x)}{\sigma_{\theta_0}(x)} \right) \left( L_n^{-1}(x) - L_{\theta_0}^{-1}(x) \right) \left( U_n(x) - L_{\theta_0}(x) \sqrt{n}(\theta_n - \theta_0) \right) G_n(dx)
$$

Using conditions (ν2), (4.3), (4.11) and (4.10), we can show the first seven terms in the above sum are all the order of $u_p(1)$. Note that the last term is just $A_{n1}(z) - \sqrt{n}(\theta_n - \theta_0) \int_{x \leq z} \frac{\nu_{\theta_0}(x)}{\sigma_{\theta_0}(x)} G_n(dx)$, so (4.12) is proved. Using a Glivenko-Cantelli type argument, we can show that

$$
\int_{x \leq z} \frac{\nu_{\theta_0}(x)}{\sigma_{\theta_0}(x)} G_n(dx) = \int_{x \leq z} \frac{\nu_{\theta_0}(x)}{\sigma_{\theta_0}(x)} G(dx) + u_p(1).
$$

17
Thus we have

\[
(4.13) \sup_{z \leq z_0} \left| A_{n2}(z) - A_{n1}(z) + \sqrt{n} (\theta_n - \theta_0)' E \frac{\hat{\theta}_0(Z)}{\sigma_0(Z)} I[Z \leq z] \right| = o_p(1).
\]

Finally, from \( W_n(z) - W_{\theta_0,G}(z) = (\hat{V}_n(z) - V_n(z)) - (A_{n2}(z) - A_{n1}(z)) \), (4.7) and (4.13), the claim (a) is proved.

**References**


