# SPECIFICATION TESTS FOR THE DISTRIBUTION OF ERRORS IN NONPARAMETRIC REGRESSION: A MARTINGALE APPROACH

**Juan Mora and Alicia Pérez-Alonso***

WP-AD 2008-11

Corresponding author: . J. Mora: Departamento del Análisis Económico, Universidad de Alicante, Apartado de Correos 99, 03080 Alicante, Spain. E-mail: juan@merlin.fae.ua.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Julio 2008
Depósito Legal: V-2995-2008



* J. Mora: Universidad de Alicante. A. Pérez-Alonso: Universidad Carlos III de Madrid.

# SPECIFICATION TESTS FOR THE DISTRIBUTION OF ERRORS IN NONPARAMETRIC REGRESSION: A MARTINGALE APPROACH

**Juan Mora and Alicia Pérez-Alonso**

**ABSTRACT**

We discuss how to test whether the distribution of regression errors belongs to a parametric family of continuous distribution functions, making no parametric assumption about the conditional mean or the conditional variance in the regression model. We propose using test statistics that are based on a martingale transform of the estimated empirical process. We prove that these statistics are asymptotically distribution-free, and two Monte Carlo experiments show that they work reasonably well in practice.

***Keywords:*** Specification Tests; Nonparametric Regression; Empirical Processes.

## 1. INTRODUCTION

Specification tests for the distribution of an observable random variable have a long tradition in Statistics. However, there are many situations in which the random variable of interest for the researcher is a non-observable regression error. For example, in Economics, the productivity of a firm is defined as the error term of a regression model whose dependent variable is firm profits; and, in Finance, the return of an asset over a period is usually defined as the error term of a dynamic regression model. In contexts such as these, knowing whether the distribution of the error term belongs to a specified parametric family or not may be crucial to achieve efficient estimation, to determine certain characteristics of interest (such as percentiles or number of modes) of the error term, or to design an efficient bootstrap procedure. This is the problem that we study in this paper.

Let $(X, Y)$ be a bivariate continuous random vector such that $E(Y^2)$ is finite, and denote $m(x) \equiv E(Y|X = x)$ and $\sigma^2(x) \equiv \text{Var}(Y|X = x)$. We can consider then the error term $\varepsilon \equiv \{Y - m(X)\}/\sigma(X)$, which is, by definition, a zero-mean unit-variance random variable. The objective of this paper is to describe how to test a parametric specification of the cumulative distribution function (c.d.f.) of $\varepsilon$, while making no parametric assumptions about the conditional mean function $m(\cdot)$ or the conditional variance function $\sigma^2(\cdot)$. Specifically, if $F_\varepsilon(\cdot)$ denotes the c.d.f. of $\varepsilon$ and $\mathcal{F} \equiv \{F(\cdot, \theta),\ \theta \in \Theta \subset \mathbb{R}^m\}$ denotes a parametric family of zero-mean unit-variance continuous c.d.f.'s, each of them known except for the parameter vector $\theta$, we propose a testing procedure to face the hypotheses

$$\begin{aligned} \text{H}_0 &: \exists\, \theta_0 \in \Theta \text{ such that } F_\varepsilon(\cdot) = F(\cdot, \theta_0), \qquad \text{vs.} \\ \text{H}_1 &: F_\varepsilon(\cdot) \notin \mathcal{F}, \end{aligned}$$

when independent and identically distributed (i.i.d.) observations $\{(X_i, Y_i)\}_{i=1}^n$, with the same distribution as $(X, Y)$, are available. The testing procedure that we propose here could also be used, with appropriate changes, if the family $\mathcal{F}$ reduces to one known c.d.f. (i.e. when there is no unknown parameter $\theta$), or if the error term that is to be analyzed is defined by removing only the conditional mean (i.e. when we consider the error term $Y - m(X)$). The specific test statistics that should be used in these more simple contexts are discussed below.

The testing problem that we study in this paper can also be considered as an extension of the classical goodness-of-fit problem. Suppose that a parametric specification for the c.d.f. of an observable continuous variable $Y$ is rejected using a traditional nonparametric goodness-of-fit statistic, such as the Kolmogorov-Smirnov one; one of the drawbacks of these statistics is that the rejection of the null hypothesis gives no intuition about the cause of the rejection. In this situation, it would be of interest to examine if the only reason why the null hypothesis has been rejected is because the parametric family fails to capture appropriately the behaviour in mean of $Y$; if we want to check whether this is the case, then we would have to analyze if the parametric specification is appropriate for $Y - m(X)$. If the null hypothesis were rejected again, we might be interested in going one step further and testing whether the parametric family fails to capture appropriately the behaviour in mean and variance of $Y$; thus, we would have to analyze if the parametric specification is appropriate for $\{Y - m(X)\}/\sigma(X)$, and this is precisely the testing problem that we consider here.

The test statistics that we propose in this paper can be motivated by studying the relationship between our problem and the classical goodness-of-fit problem. If the error term $\varepsilon$ were observable and parameter $\theta_0$ were known, our test would be the classical goodness-of-fit test. In our context, the unobservable errors must be replaced by residuals, which must be derived using nonparametric estimations of $m(\cdot)$ and $\sigma^2(\cdot)$ since no parametric form for these functions is assumed, and parameter $\theta_0$ must be replaced by an appropriate estimator, say $\widehat{\theta}$. Thus, any of the traditional nonparametric goodness-of-fit statistics could be used as a statistic for our test and computed using nonparametric residuals and the estimator $\widehat{\theta}$. However, it is well-known in the literature that the consequence of replacing errors by parametric residuals and parameters by estimators in goodness-of-fit tests is that the resulting statistics are no longer asymptotically distribution-free (see e.g. Durbin, 1973 or Loynes, 1980); furthermore, the asymptotic null distributions usually depend on unknown quantities and, hence, asymptotic critical values cannot be tabulated. In this paper we prove that this is also the case when nonparametric residuals are used, and we discuss how this problem can be circumvented in our testing problem. Specifically, by using the results derived in Akritas and Van Keilegom (2001), we derive the asymptotic behaviour of goodness-of-fit statistics based on nonparametric residuals and estimators; and then, following the methodology introduced in Khmaladze (1993), we derive the martingale-transformed test statistics that are appropriate in our context.

The rest of the paper is organized as follows. In Section 2 we introduce the empirical process on which our statistics are based and derive its asymptotic properties. In Section 3 we describe the martingale transformation that leads to asymptotically distribution-free test statistics. In Section 4 we report the results of a set of Monte Carlo experiments that illustrate the performance of the statistics with moderate sample sizes. Some concluding remarks are provided in Section 5. All proofs are relegated to an Appendix.

## 2. STATISTICS BASED ON THE ESTIMATED EMPIRICAL PROCESS

If we had observations of the error term $\{\varepsilon_i\}_{i=1}^n$ and parameter $\theta_0$ were known, we could use as a statistic for our test the asymptotic Kolmogorov-Smirnov statistic $K_n \equiv n^{1/2} \sup_{z\in\mathbb{R}} |F_n(z) - F(z, \theta_0)|$ or the Cramér-von Mises statistic $C_n \equiv \sum_{i=1}^n \{F_n(\varepsilon_i) - F(\varepsilon_i, \theta_0)\}^2$, where $F_n(\cdot)$ denotes the empirical c.d.f. based on $\{\varepsilon_i\}_{i=1}^n$. Both $K_n$ and $C_n$ are functionals of the so-called "empirical process" $\mathbf{V}_n(\cdot)$, defined for $z \in \mathbb{R}$ by

$$\mathbf{V}_n(z) \equiv n^{-1/2} \sum_{i=1}^n \{I(\varepsilon_i \leq z) - F(z, \theta_0)\},$$

where $I(\cdot)$ is the indicator function; hence, the asymptotic properties of $K_n$ and $C_n$ can be derived by studying the weak convergence of the empirical process $\mathbf{V}_n(\cdot)$. In our context, the test statistics must be constructed replacing errors by residuals and the unknown parameter by an estimator. Since no parametric assumption about the conditional mean $m(\cdot)$ or the conditional variance $\sigma^2(\cdot)$ is made, the residuals $\{\widehat{\varepsilon}_i\}_{i=1}^n$ must be constructed using nonparametric estimates of these functions. Specifically, we consider Nadaraya-Watson estimators, i.e. $\widehat{m}(x) \equiv \sum_{i=1}^n W_i(x, h_n) Y_i$ and $\widehat{\sigma}^2(x) = \sum_{i=1}^n W_i(x, h_n) Y_i^2 - \widehat{m}(x)^2$, where $W_i(x, h_n) \equiv K\{(x - X_i)/h_n\} / \sum_{j=1}^n K\{(x - X_j)/h_n\}$, $K(\cdot)$ is a known kernel function and $\{h_n\}$ is a sequence of positive smoothing values. With these estimates we construct the nonparametric residuals $\widehat{\varepsilon}_i \equiv \{Y_i - \widehat{m}(X_i)\}/\widehat{\sigma}(X_i)$. On the other hand, the unknown parameter must be replaced by an appropriate estimator $\widehat{\theta}$, usually also based on the residuals $\widehat{\varepsilon}_i$. Thus, the test statistics that can be used are

$$\widehat{K}_n \equiv n^{1/2} \sup_{z\in\mathbb{R}} \left|\widehat{F}_n(z) - F(z,\widehat{\theta})\right| \qquad \text{and} \qquad \widehat{C}_n \equiv \textstyle\sum_{i=1}^n \{\widehat{F}_n(\widehat{\varepsilon}_i) - F(\widehat{\varepsilon}_i,\widehat{\theta})\}^2,$$

where $\widehat{F}_n(\cdot)$ denotes the empirical c.d.f. based on $\{\widehat{\varepsilon}_i\}_{i=1}^n$. Both $\widehat{K}_n$ and $\widehat{C}_n$ are functionals of the process $\widehat{\mathbf{V}}_n(\cdot)$, defined for $z \in \mathbb{R}$ by

$$\widehat{\mathbf{V}}_n(z) = n^{-1/2} \textstyle\sum_{i=1}^n \{I(\widehat{\varepsilon}_i \leq z) - F(z,\widehat{\theta})\}.$$

Hence, the asymptotic behaviour of $\widehat{K}_n$ and $\widehat{C}_n$ is derived studying the asymptotic properties of the process $\widehat{\mathbf{V}}_n(\cdot)$, which will be hereafter referred to as the "estimated empirical process".

First of all we discuss the asymptotic relationship between the empirical process $\mathbf{V}_n(\cdot)$ and the estimated empirical process $\widehat{\mathbf{V}}_n(\cdot)$, since this relationship will be crucial to establishing the asymptotic behaviour of $\widehat{K}_n$ and $\widehat{C}_n$. The following assumptions will be required:

**Assumption 1:** The support of $X$, hereafter denoted $S_X$, is bounded, convex and has a non-empty interior.

**Assumption 2:** The c.d.f. of $X$, denoted $F_X(\cdot)$, admits a density function $f_X(\cdot)$ that is twice continuously differentiable and strictly positive in $S_X$.

**Assumption 3:** The conditional c.d.f. of $Y \mid X = x$, hereafter denoted $F(\cdot|x)$, admits a density function $f(\cdot|x)$. Additionally, both $F(y|x)$ and $f(y|x)$ are continuous in $(x,y)$, the partial derivatives $\frac{\partial}{\partial y} f(y|x)$, $\frac{\partial}{\partial x} F(y|x)$, $\frac{\partial^2}{\partial x^2} F(y|x)$ exist and are continuous in $(x,y)$, and $\sup_{x,y} |yf(y|x)| < \infty$, $\sup_{x,y} |y\frac{\partial}{\partial x} F(y|x)| < \infty$, $\sup_{x,y} |y^2 \frac{\partial}{\partial y} f(y|x)| < \infty$, $\sup_{x,y} |y^2 \frac{\partial^2}{\partial x^2} F(y|x)| < \infty$.

**Assumption 4:** The functions $m(\cdot)$ and $\sigma^2(\cdot)$ are twice continuously differentiable. Additionally, there exists $C > 0$ such that $\inf_{x\in S_X} \sigma^2(x) \geq C$.

**Assumption 5:** The kernel function $K(\cdot)$ is a symmetric and twice continuously differentiable probability density function with compact support and $\int uK(u)\mathrm{d}u = 0$.

**Assumption 6:** The smoothing value $h_n$ satisfies that $nh_n^4 = o(1)$, $nh_n^5/\log h_n^{-1} = O(1)$ and $\log h_n^{-1}/(nh_n^{3+2\delta}) = o(1)$ for some $\delta > 0$.

**Assumption 7:** The c.d.f $F(\cdot,\theta)$ admits a density function $f(\cdot,\theta)$ which is positive and uniformly continuous in $\mathbb{R}$. Additionally, $f(\cdot,\cdot)$ is twice differentiable with respect to both arguments, $F(\cdot,\cdot)$ has bounded derivative with respect to the second argument and $\sup_{z\in\mathbb{R}}|zf(z,\theta)| < \infty$ for every $\theta \in \Theta$.

**Assumption 8:** For a certain metric $d(\cdot,\cdot)$, there is a unique value $\theta_0$ in $\Theta$ satisfying that $F(\cdot,\theta_0) \equiv \arg\inf_{F\in\mathcal{F}} d(F_\varepsilon, F)$.

**Assumption 9:** The estimator $\widehat{\theta}$ satisfies that $\widehat{\theta} - \theta_0 = o_p(1)$. Additionally, if $\mathrm{H}_0$ holds then $n^{1/2}(\widehat{\theta} - \theta_0) = n^{-1/2}\sum_{i=1}^n \psi(X_i, \varepsilon_i, \theta_0) + o_p(1)$, where the function $\psi(\cdot,\cdot,\cdot)$ is twice continuously differentiable with respect to the second argument and is such that $\sup_{z\in\mathbb{R}}|\frac{\partial^2}{\partial z^2}\psi(x,z,\theta_0)| < \infty$, $E\{\psi(X,\varepsilon,\theta_0)\} = 0$ and $\Omega \equiv E\{\psi(X,\varepsilon,\theta_0)\psi(X,\varepsilon,\theta_0)'\}$ is finite.

Assumptions 1-6, which are similar to those introduced in Akritas and Van Keilegom (2001), guarantee that the nonparametric estimators of the conditional mean and variance behave properly. Assumption 7 allows us to use mean-value arguments to analyze the effect of introducing the estimator $\widehat{\theta}$. Assumption 8 ensures that the true parameter $\theta_0$ is identified under $\mathrm{H}_0$. The metric that is introduced in this assumption depends on the procedure that is used to estimate $\theta_0$; note that a natural estimation procedure in this context would be residual-based maximum-likelihood, but other procedures such as minimum distance or method of moments might be preferred for robustness or computational reasons. Assumption 9 implies that, under $\mathrm{H}_0$, the estimator $\widehat{\theta}$ is root-$n$-consistent; note also that the asymptotic expansion that is assumed for $n^{1/2}(\widehat{\theta}-\theta_0)$ is satisfied, under suitable smoothness assumptions, by most estimators.

Our first proposition states an "oscillation-like" result between the empirical process and the estimated empirical process in our context.

**Proposition 1:** *If* $\mathrm{H}_0$ *holds and assumptions 1-9 are satisfied then*

$$\sup_{z\in\mathbb{R}}\left|\widehat{\mathbf{V}}_n(z)-\{\mathbf{V}_n(z)+\mathbf{A}_{1n}(z)+\mathbf{A}_{2n}(z)-\mathbf{A}_{3n}(z)\}\right|=o_p(1),$$

*where*

$$\mathbf{A}_{1n}(z)\equiv f(z,\theta_0)n^{-1/2}\textstyle\sum_{i=1}^n\{(\varphi_1(X_i,Y_i)+\beta_{1n}\},$$

$$\mathbf{A}_{2n}(z)\equiv zf(z,\theta_0)n^{-1/2}\textstyle\sum_{i=1}^n\{\varphi_2(X_i,Y_i)+\beta_{2n}\},$$

$$\mathbf{A}_{3n}(z)\equiv F_\theta(z,\theta_0)'n^{1/2}(\widehat{\theta}-\theta_0),$$

$F_\theta(z,\theta)\equiv\frac{\partial}{\partial\theta}F(z,\theta)$, $\varphi_1(x,y)\equiv-\sigma(x)^{-1}\int\{I(y\leq v)-F(v|x)\}\mathrm{d}v$, $\varphi_2(x,y)\equiv-\sigma(x)^{-2}\int\{v-m(x)\}\{I(y\leq v)-F(v|x)\}\mathrm{d}v$ *and, for* $j=1,2$, $\beta_{jn}\equiv\frac{1}{2}h_n^2\{\int u^2K(u)\mathrm{d}u\}\{\int\varphi_j^*(x,x)f_X(x)\mathrm{d}x\}$, $\varphi_j^*(u,x)\equiv\frac{\partial^2}{\partial u^2}E[\varphi_j(u,Y)|X=x]$.

Note that processes $\mathbf{A}_{1n}(\cdot)$ and $\mathbf{A}_{2n}(\cdot)$ arise as a consequence of the nonparametric estimation of the conditional mean and variance, respectively, whereas $\mathbf{A}_{3n}(\cdot)$ reflects the effect of estimating $\theta_0$. The following theorem states the asymptotic behaviour of $\widehat{K}_n$ and $\widehat{C}_n$.

**Theorem 1:** *Suppose that assumptions 1-9 hold. Then:*

*a) If* $\mathrm{H}_0$ *holds then*

$$\widehat{K}_n\xrightarrow{d}\sup_{t\in\mathbb{R}}|D(t)|\qquad\text{and}\qquad\widehat{C}_n\xrightarrow{d}\int\{D(t)\}^2\mathrm{d}t,$$

*where* $D(\cdot)$ *is a zero-mean Gaussian process on* $\mathbb{R}$ *with covariance structure*

$$\mathrm{Cov}\{D(s), D(t)\} = F(\min(s,t),\theta_0) - F(s,\theta_0)F(t,\theta_0) + H(s,t,\theta_0),$$

*and*

$$\begin{aligned} H(s,t,\theta_0) \equiv\ & f(s,\theta_0)[E\{I(\varepsilon \le t)\varepsilon\} + \tfrac{s}{2}E\{I(\varepsilon \le t)(\varepsilon^2 - 1)\}] \\ & + f(t,\theta_0)[E\{I(\varepsilon \le s)\varepsilon\} + \tfrac{t}{2}E\{I(\varepsilon \le s)(\varepsilon^2 - 1)\}] \\ & + f(s,\theta_0)f(t,\theta_0)[1 + \tfrac{s+t}{2}E(\varepsilon^3) + \tfrac{st}{4}\{E(\varepsilon^4) - 1\}] \\ & - F_\theta(s,\theta_0)' E\{I(\varepsilon \le t)\psi(X,\varepsilon,\theta_0)\} \\ & - F_\theta(t,\theta_0)' E\{I(\varepsilon \le s)\psi(X,\varepsilon,\theta_0)\} \\ & - f(s,\theta_0)F_\theta(t,\theta_0)'[E\{\psi(X,\varepsilon,\theta_0)\varepsilon\} + \tfrac{s}{2}E\{\psi(X,\varepsilon,\theta_0)(\varepsilon^2 - 1)\}] \\ & - f(t,\theta_0)F_\theta(s,\theta_0)'[E\{\psi(X,\varepsilon,\theta_0)\varepsilon\} + \tfrac{t}{2}E\{\psi(X,\varepsilon,\theta_0)(\varepsilon^2 - 1)\}] \\ & + F_\theta(s,\theta_0)'\Omega F_\theta(t,\theta_0). \end{aligned}$$

*b) If* $\mathrm{H}_1$ *holds then,* $\forall\ c \in \mathbb{R}$*,*

$$P(\widehat{K}_n > c) \to 1 \qquad \textit{and} \qquad P(\widehat{C}_n > c) \to 1.$$

Since the covariance structure of the limiting process depends on the underlying distribution of the errors and the true parameter, it is not possible to obtain asymptotic critical values valid for any situation. To overcome this problem, it would be possible to approximate critical values by bootstrap methods. This is the approach that is followed in Neumeyer et al. (2005) in a closely related context. However, following Khmaladze (1993) and Bai (2003), it is also possible to propose test statistics based on a martingale transform of the estimated process; this is the alternative approach that we explore in the next section. The advantage of this alternative approach is that it usually leads to test statistics with better power properties (see e.g. Koul and Sakhanenko 2005 and Mora and Neumeyer 2008) with much less computational effort, since no resampling is required.

## 3. STATISTICS BASED ON A MARTINGALE-TRANSFORMED PROCESS

As Proposition 1 states, three new processes appear in the relationship between the estimated empirical process $\widehat{\mathbf{V}}_n(\cdot)$ and the true empirical process $\mathbf{V}_n(\cdot)$. These three additional processes stem from the estimation of the conditional mean, the conditional variance and the unknown parameter. If we follow the methodology described in Bai (2003), this relationship leads us to consider the martingale-transformed process

$$\mathbf{W}_n(z) \equiv n^{1/2}\{\widehat{F}_n(z) - \int_{-\infty}^{z} q(u,\theta_0)'C(u,\theta_0)^{-1}\overline{d}_n(u,\theta_0)f(u,\theta_0)\mathrm{d}u\}, \qquad (1)$$

where

$$q(u,\theta) \equiv (1, f_u(u,\theta)/f(u,\theta), 1 + uf_u(u,\theta)/f(u,\theta), f_\theta(u,\theta)'/f(u,\theta))',$$
$$C(u,\theta) \equiv \int_u^{+\infty} q(\tau,\theta)q(\tau,\theta)'f(\tau,\theta)\mathrm{d}\tau,$$
$$\overline{d}_n(u,\theta) \equiv \int_u^{+\infty} q(\tau,\theta)\mathrm{d}\widehat{F}_n(\tau) = n^{-1}\sum_{i=1}^n I(\widehat{\varepsilon}_i \geq u)q(\widehat{\varepsilon}_i,\theta),$$

and $f_u(u,\theta) \equiv \frac{\partial}{\partial u}f(u,\theta)$, $f_\theta(u,\theta) \equiv \frac{\partial}{\partial \theta}f(u,\theta)$. Since process $\mathbf{W}_n(\cdot)$ depends on the unknown parameter $\theta_0$, we cannot use it to construct test statistics; obviously, the natural solution is to replace again $\theta_0$ by $\widehat{\theta}$. Thus, we consider the estimated martingale-transformed process $\widehat{\mathbf{W}}_n(\cdot)$, defined in the same way as $\mathbf{W}_n(\cdot)$ but replacing $\theta_0$ by $\widehat{\theta}$. With this estimated process we can derive a Kolmogorov-Smirnov or a Cramér-von Mises statistic as above. However, in this case, the supremum (in the Kolmogorov-Smirnov case) and the integral (in the Cramér-von Mises case) are not taken with respect to $\mathbb{R}$, because the asymptotic equivalence between $\mathbf{W}_n(\cdot)$ and $\widehat{\mathbf{W}}_n(\cdot)$ is only proved at intervals $(-\infty, z_0]$, with $z_0 \in \mathbb{R}$ (see Theorem 4 in Bai, 2003). Thus, the statistics that we consider are $\overline{K}_{n,z_0} \equiv F(z_0,\widehat{\theta})^{-1/2}\sup_{z\in(-\infty,z_0]}\left|\widehat{\mathbf{W}}_n(z)\right|$ and $\overline{C}_{n,z_0} \equiv F(z_0,\widehat{\theta})^{-2}n^{-1}\sum_{i=1}^n I(\widehat{\varepsilon}_i \leq$

$z_0)\widehat{\mathbf{W}}_n(\widehat{\varepsilon}_i)^2$, where $z_0$ is any large enough fixed real number; note that the factor that depends on $F(z_0, \widehat{\theta})$ is introduced in order to obtain an asymptotic distribution that does not depend on $z_0$.

The asymptotic behaviour of the martingale-transformed statistics are derived studying the convergence of $\widehat{\mathbf{W}}_n(\cdot)$. Given $\theta \in \Theta$ and $M > 0$, denote $N_n(\theta, M) \equiv \{v \in \Theta; ||v - \theta|| \leq Mn^{-1/2}\}$ and $q_\theta(u, \theta) \equiv \frac{\partial}{\partial \theta} q(u, \theta)$. The following assumptions, which ensure that the martingale transformation behaves properly, are required.

**Assumption 10:** $C(u, \theta)$ is a non-singular matrix for every $u \in [-\infty, +\infty)$ and for every $\theta \in \Theta$.

**Assumption 11:** There exists $M_0$ such that

$$\sup_{v \in N_n(\theta_0, M_0)} \int_{-\infty}^{+\infty} ||q_\theta(u, v)||^2 f(u, \theta_0) \mathrm{d}u = O(1).$$

Assumption 10 ensures that the martingale transformation can be performed. This assumption, which is not satisfied in some cases, might be relaxed at the cost of some more technical complexity; in this case, generalized inverse matrices would have to be used (see Tsigroshvili 1998).

**Theorem 2:** *Suppose that assumptions 1-11 hold.*

*a) If* $\mathrm{H}_0$ *holds and* $F(z_0, \theta_0) \in (0, 1)$, *then:*

$$\overline{K}_{n,z_0} \xrightarrow{d} \sup_{t \in [0,1]} |\mathbf{W}(t)| \qquad \text{and} \qquad \overline{C}_{n,z_0} \xrightarrow{d} \int_{[0,1]} \{\mathbf{W}(t)\}^2 \mathrm{d}t,$$

*where* $\mathbf{W}(\cdot)$ *is a Brownian motion.*

*b) If* $\mathrm{H}_1$ *holds,* $E(\varepsilon^3) < \infty$, $\int u^4 f_u(u, \theta_0) \mathrm{d}u < \infty$ *and* $f(\cdot, \theta)$ *satisfies the Fréchet-Cramér-Rao regularity conditions (see e.g. Rohatgi and Saleh 2001,* *p. 391) then there exists* $z_* \in \mathbb{R}$ *such that, if* $z_0 \geq z_*$, $\forall\ c \in \mathbb{R}$*:*

$$P\left(\overline{K}_{n,z_0} > c\right) \to 1 \qquad \textit{and} \qquad P\left(\overline{C}_{n,z_0} > c\right) \to 1.$$

It follows from this theorem that a consistent asymptotically valid testing procedure with significance level $\alpha$ is to reject $\mathrm{H}_0$ if $\overline{K}_{n,z_0} > k_\alpha$, or to reject $\mathrm{H}_0$ if $\overline{C}_{n,z_0} > c_\alpha$, for a large enough $z_0$, where $k_\alpha$ and $c_a$ denote appropriate critical values derived from the c.d.f.'s of $\sup_{t\in[0,1]} |\mathbf{W}(t)|$ and $\int_{[0,1]}\{\mathbf{W}(t)\}^2 \mathrm{d}t$. Specifically, the critical values with the usual significance levels are $k_{0.10} = 1.96$, $k_{0.05} = 2.24$, $k_{0.01} = 2.81$ for $\overline{K}_{n,z_0}$ (see Shorack and Wellner 1986, p.34), and $c_{0.10} = 1.196$, $c_{0.05} = 1.656$, $c_{0.01} = 2.787$ for $\overline{C}_{n,z_0}$ (see Rothman and Woodroofe 1972). Also note that we only include here results under a fixed alternative; it would also be possible to derive results under local alternatives, and this might be the starting point for a power-based comparison between $\overline{K}_{n,z_0}$ or $\overline{C}_{n,z_0}$ and bootstrap-based tests based on $\widehat{K}_n$ or $\widehat{C}_n$. However, the asymptotic results that would obtained in this way depend on various unknown quantities and do not lead to any clear conclusion (see Mora and Neumeyer 2008).

The statistics $\overline{K}_{n,z_0}$ and $\overline{C}_{n,z_0}$ are designed to test whether the c.d.f. of the error term $\varepsilon = \{Y - m(X)\}/\sigma(X)$ belongs to a parametrically specified family of zero-mean unit-variance continuous c.d.f.'s. If we were interested in testing whether the c.d.f. of the error term $\varepsilon = \{Y - m(X)\}/\sigma(X)$ is a known zero-mean unit-variance c.d.f. $F_0(\cdot)$, then the test statistics should be based on the process $\mathbf{W}_n(\cdot)$ defined as in (1), but considering

$$q(u) \equiv (1, \frac{f_{0,u}(u)}{f_0(u)}, 1 + \frac{u f_{0,u}(u)}{f_0(u)})', \tag{2}$$

where $f_0(\cdot)$ and $f_{0,u}(\cdot)$ denote the first and second derivative of $F_0(\cdot)$.

If we were interested in testing whether the c.d.f. of the error term $Y - m(X)$ belongs to a parametrically specified family of zero-mean continuous c.d.f.'s, then the test statistics should be based on the process $\widehat{\mathbf{W}}_n(\cdot)$ as defined above, but considering $q(u, \theta) \equiv (1,\ f_u(u, \theta)/f(u, \theta),\ f_\theta(u, \theta)'/f(u, \theta))'$. Finally, if we were interested in testing whether the c.d.f. of the error term $Y - m(X)$ is a known zero-mean c.d.f. $F_0(\cdot)$, then the test statistics that we would use should be based again on the process $\mathbf{W}_n(\cdot)$ defined as in (1), but considering $q(u) \equiv (1,\ f_{0,u}(u)/f_0(u))'$.

## 4. SIMULATIONS

In order to check the behaviour of the statistics, we perform two sets of Monte Carlo experiments. In all experiments, independent and identically distributed $\{(X_i, Y_i)\}_{i=1}^n$ are generated as follows: $X_i$ has uniform distribution on $(0, 1)$ and $Y_i = 1 + X_i + \varepsilon_i$, where $X_i$ and $\varepsilon_i$ are independent, and the distribution of $\varepsilon_i$ changes across experiments. In the first set of experiments we test the null hypothesis that the error term $\varepsilon \equiv \{Y - m(X)\}/\sigma(X)$ is standard normal, when in fact it follows a standardized Student's t distribution with $1/\delta$ degrees of freedom, and we consider $\delta = 0$, 1/12, 1/9, 1/7, 1/5 and 1/3; thus, in our first set of experiments $\text{H}_0$ is true if and only if $\delta = 0$, and the other values of $\delta$ allow us to examine the ability of the testing procedure to detect deviations from the null hypothesis caused by thick tails. In the second set of experiments we test the null hypothesis that the error term $\varepsilon$ is distributed as a standardized Student's t distribution with $\theta$ (unknown) degrees of freedom, when in fact $\varepsilon = [U - E(U)]/\text{Var}(U)^{1/2}$ and $U$ is a skewed Student's $\text{t}_5$ distribution (see Fernandez and Steel, 1998) with density function $f_\gamma(x) = 2(\gamma + 1/\gamma)^{-1}[f(\gamma x)I(x < 0) + f(x/\gamma)I(x \geq 0)]$, where $f(\cdot)$ is the Student's $\text{t}_5$ density, and we consider $\gamma = 1$, 1.25, 1.50, 1.75 and 2; thus, in our second set of experiments $H_0$ is true if and only if $\gamma = 1$, and the other values of $\gamma$ allow us to examine the ability of the testing procedure to detect deviations from the null hypothesis caused by asymmetries. Note that the error $U$ can be generated from a uniform random variable on $(0,1)$, say $Z$, by considering $U = \gamma^{-1}Q_5((\gamma^2 + 1)Z/2)I(Z < (\gamma^2+1)^{-1}) + \gamma Q_5([(1-\gamma^{-2}) + (1+\gamma^{-2})Z]/2)I(Z \geq (\gamma^2+1)^{-1})$, where $Q_5(\cdot)$ is the inverse of the c.d.f. of a Student's $t_5$ distribution; also note that $E(U) = 4\sqrt{5}(\gamma - 1/\gamma)/(3\pi)$ and $\text{Var}(U) = [80 + (\gamma^2 + 1/\gamma^2 - 1)(15\pi^2 - 80)]/(9\pi^2)$.

In the first set of experiments, since there is no $\theta$ parameter under the null hypothesis, the test statistics are based on the process $\mathbf{W}_n(\cdot)$ defined as in (1), but now with the function $q(\cdot)$ that appears in (2), which in this specific case proves to be $(1,\, -u,\, 1-u^2)'$. In the second set of experiments parameter $\theta$ is estimated by the method of moments using the fourth order moment and assuming that the null hypothesis is true, i.e. $\widehat{\theta} = (4\widehat{m}_4 - 6)/(\widehat{m}_4 - 3)$, where $\widehat{m}_4 \equiv n^{-1}\sum_{i=1}^{n}\widehat{\varepsilon}_i^4$; in these experiments the test statistics are based on the process $\widehat{\mathbf{W}}_n(\cdot)$ and, in this case, since $f(u,\theta)$ is the density of a standardized Student's $t_\theta$ density, then $f_u(u,\theta)/f(u,\theta) = -u(\theta+1)(u^2+\theta-2)^{-1}$ and

$$\frac{f_\theta(u,\theta)}{f(u,\theta)} = \frac{1}{2}[\psi(\frac{\theta+1}{2}) - \psi(\frac{\theta}{2}) - \frac{\theta - 2 - \theta u^2}{(\theta-2)(\theta-2+u^2)} - \ln(1 + \frac{u^2}{\theta-2})],$$

where $\psi(\cdot)$ is the digamma function. The computation of the statistics requires the use of Nadaraya-Watson estimates of the conditional mean and variance functions. We have used the standard normal density function as a kernel function $K(\cdot)$, and various smoothing values to analyze how this selection influences the results; specifically, we consider $h^{(j)} = C^{(j)}\widehat{\sigma}_X n^{-1/5}$, for $j = 1, ..., 4$, where $\widehat{\sigma}_X$ is the sample standard deviation of $\{X_i\}_{i=1}^n$ and $C^{(j)} = j/2$. The integrals within the martingale-transformed process have been approximated numerically. Finally, we consider $z_0 = 1.645$ in experiment 1 and $z_0 = 2.015$ in experiment 2; thus, when $H_0$ is true approximately the top 5% residuals are discarded.

In Tables 1 and 2, we report the proportion of rejections of the null hypothesis for $n = 100$ and $n = 500$ with various significance levels; these results are based on 1000 replications. We only report the results for the Cramer-von Mises type statistics, since the results that are obtained with the Kolmogorov-Smirnov type statistic are quite similar. The results that we obtain show that the statistic works reasonably well for these sample sizes, though the empirical sizes are always below the nominal ones. In addition, the performance of the statistics does not seem to be very sensitive to the choice of the smoothing value, especially in the second experiment.

## 5. CONCLUDING REMARKS

In this paper we discuss how to test if the distribution of errors from a nonparametric regression model belongs to a parametric family of continuous distribution functions. We propose using test statistics that are based on a martingale transform of the estimated empirical process. These test statistics are asymptotically distribution-free, and our Monte Carlo results suggest that they work reasonably well in practice.

The present research could be extended in several directions. First of all, it would be interesting to extend our results to the case of testing symmetry of the error distribution in a nonparametric regression model. We should take into account that the null hypothesis is no longer a parametric one; thus, the martingale transformation cannot be estimated parametrically and the usual problems with nonparametric convergence rates may arise. Related references in this context are

Dette el al. (2002) and Neumeyer and Dette (2007), where a simple symmetric wild bootstrap is proposed to mimic the distribution of the statistic. Comparison of level and power properties for finite samples with the new martingale approach would be interesting.

A generalization to models with higher dimensional covariates would be desirable, but it is not straightforward to extend the results of Akritas and Van Keilegom (2001) due to the so-called curse of dimensionality. However, these results should suffice to derive asymptotic properties in additive models. To that end, we should first provide an "oscillation like result" for the empirical distribution function of residuals in additive regression models. Thus, the proof for our martingale transformed process could be generalized to this sort of models, because we do not use the specific form of the model here.

In addition to this, it would be also interesting to extend the results we have already obtained to dynamic models. The main point here is to extend Theorem 1 of Akritas and Van Keilegom (2001), which proposes a consistent estimator of the distribution of the error term $\varepsilon$ based on nonparametric regression residuals when $\varepsilon$ is independent of $X$ in an i.i.d. context, to a context with dependent data. Recently, Hansen (2008) provides uniform convergence rates for kernel estimators of density functions and regression functions when the observations come from a stationary $\beta$-mixing sequence. Furthermore, Franke et al. (2002) prove the consistency of bootstrap kernel estimators in a nonparametric model of nonlinear autoregression when $\{Y_i\}_{i=1}^n$ is a strictly stationary and ergodic process. Using their results, a bootstrap version of the test could be proposed in a context of dependence.

## APPENDIX: PROOFS

**Proof of Proposition 1:** Assume that $H_0$ holds and let $\widehat{\theta}$ be an appropriate estimator of $\theta_0$. If we add and subtract $F(z,\theta_0)$ to $\widehat{\mathbf{V}}_n(\cdot)$, we obtain

$$\begin{aligned} \widehat{\mathbf{V}}_n(z) &= n^{-1/2}\sum_{i=1}^{n}[I(\widehat{\varepsilon}_i \leq z) - F(z,\theta_0)] - n^{1/2}[F(z,\widehat{\theta}) - F(z,\theta_0)] \\ &= (I) - (II). \end{aligned} \tag{3}$$

By Taylor expansion, the second term admits the approximation

$$(II) = F_\theta(z,\theta_0)' n^{1/2}(\widehat{\theta} - \theta_0) + F_{\theta\theta}(z,\overline{\theta})' n^{1/2}(\widehat{\theta} - \theta_0)^2/2, \tag{4}$$

where $F_{\theta\theta}$ denotes the second partial derivative of $F(\cdot,\cdot)$ with respect to the second argument and $\overline{\theta}$ denotes a mean value between $\widehat{\theta}$ and $\theta_0$. Apply assumption 9 to show that the last term is $O_p(n^{-1/2})$.

From Theorem 1 in Akritas and Van Keilegom (2001), we obtain the following expansion of the empirical c.d.f. based on the estimated residuals $\widehat{\varepsilon}_i$:

$$\begin{aligned} \widehat{F}_n(z) &= n^{-1}\sum_{i=1}^{n} I(\widehat{\varepsilon}_i \leq z) \\ &= n^{-1}\sum_{i=1}^{n} I(\varepsilon_i \leq z) + n^{-1}\sum_{i=1}^{n}\varphi(X_i,Y_i,z) + \beta_n(z) + R_n(z), \end{aligned} \tag{5}$$

where $\varphi(x,y,z) = -f(z,\theta_0)\sigma^{-1}(x)\int[I(y \leq v) - F(v|x)](1 + z\frac{v-m(x)}{\sigma(x)})\mathrm{d}v$, $\beta_n(z) = \frac{1}{2}h_n^2\{\int u^2K(u)\mathrm{d}u\}\{\int \varphi^*(x,x,z)f_X(x)dx\}$, $\varphi^*(u,x,z) \equiv \frac{\partial^2}{\partial u^2}E[\varphi(u,Y,z)|X = x]$, and $\sup_{z\in\mathbb{R}}|R_n(z)| = o_p(n^{-1/2}) + o_p(h_n^2) = o_p(n^{-1/2})$. Note that

$$\begin{aligned} \varphi(x,y,z) &= f(z,\theta_0)\varphi_{1n}(x,y) + zf(z,\theta_0)\varphi_{2n}(x,y), \\ \beta_n(z) &= f(z,\theta_0)\beta_{1n} + zf(z,\theta_0)\beta_{2n} \end{aligned}$$

where $\varphi_{1n}(\cdot,\cdot)$, $\varphi_{2n}(\cdot,\cdot)$, $\beta_{1n}$ and $\beta_{2n}$ are as defined above. The proposition follows immediately by appealing to (4) and (5) in (3). ■

**Proof of Theorem 1:** First we prove the theorem for $\widehat{K}_n$. Note that, under $\mathrm{H}_0$, $\widehat{K}_n = \sup_{z\in\mathbb{R}} \left|\widehat{\mathbf{D}}_n(z)\right| + o(1)$, where we define

$$\widehat{\mathbf{D}}_n(z) \equiv n^{-1/2}\sum_{i=1}^{n}\{I(\widehat{\varepsilon}_i \leq z) - F(z,\widehat{\theta}) - \beta_n(z)\}, \tag{6}$$

and $\beta_n(\cdot)$ is defined above. To derive the asymptotic distribution of $\widehat{K}_n$, it suffices to prove that $\widehat{\mathbf{D}}_n(\cdot)$ converges weakly to $\mathbf{D}(\cdot)$, and then apply the continuous mapping theorem. From Proposition 1 and (6), it follows that $\widehat{\mathbf{D}}_n(\cdot)$ has the same asymptotic behaviour as $\mathbf{D}_n(z) \equiv n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \leq z) - F(z,\theta_0) + \varphi(X_i, Y_i, z)] - F_\theta(z,\theta_0)' n^{1/2}(\widehat{\theta} - \theta_0)$, where the function $\varphi(\cdot,\cdot,\cdot)$ is defined above.

To analyze the process $\mathbf{D}_n(\cdot)$, we follow a similar approach to that used in the proof of Theorem 3.1 in Dette and Neumeyer (2007), though now an additional term turns up due to the estimation of parameter $\theta_0$. We can rewrite $\varphi(\cdot,\cdot,\cdot)$ as follows:

$$\begin{aligned}
\varphi(x,y,z) &= -\tfrac{f(z,\theta_0)}{\sigma(x)}(1 - \tfrac{zm(x)}{\sigma(x)})\{\int_y^\infty (1 - F(v|x))dv - \int_{-\infty}^y F(v|x)\mathrm{d}v\} \\
&\quad -\tfrac{zf(z,\theta_0)}{\sigma^2(x)}\{\int_y^\infty v(1 - F(v|x))dv - \int_{-\infty}^y vF(v|x)\mathrm{d}v\} \\
&= -\tfrac{f(z,\theta_0)}{\sigma(x)}(1 - \tfrac{zm(x)}{\sigma(x)})(m(x) - y) - \tfrac{zf(z,\theta_0)}{2\sigma^2(x)}(\sigma^2(x) + m^2(x) - y^2).
\end{aligned}$$

For $y = m(x) + \sigma(x)\varepsilon$, we have

$$\varphi(x,y,z) = \varphi(x, m(x) + \sigma(x)\varepsilon, z) = f(z,\theta_0)(\varepsilon + \frac{z}{2}(\varepsilon^2 - 1)). \tag{7}$$

We also have for the bias part

$$\begin{aligned}
\beta_n(z) &= -h_n^2\{\int k(u)u^2\mathrm{d}u\} \times \{f(z,\theta_0)\int \frac{1}{\sigma^2(x)}[(m''\sigma f_X)(x) \\
&\quad +2(m'\sigma f_X')(x) - 2(\sigma' m' f_X)(x)]\mathrm{d}x + zf(z,\theta_0)\int \frac{1}{\sigma^2(x)}[2(\sigma'\sigma f_X')(x) \\
&\quad +(\sigma''\sigma f_X)(x) - (m'(x))^2 f_X(x) - 3(\sigma'(x))^2 f_X(x)]\mathrm{d}x\}/2,
\end{aligned}$$

where we use the prime and the double prime to denote the first and second order derivatives of the corresponding function, respectively. Observe that the bias can be omitted if $nh_n^4 = o(1)$. By assumption 9 and replacing (7) in $\mathbf{D}_n(z)$, we obtain

$$\begin{aligned}\mathbf{D}_n(z) &= n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \le z) - F(z,\theta_0) + f(z,\theta_0)(\varepsilon_i + \frac{z}{2}(\varepsilon_i^2 - 1)) \\ &\quad -F_\theta(z,\theta_0)'\psi(X_i,\varepsilon_i,\theta_0)] + o_p(1) \\ &= \widetilde{\mathbf{D}}_n(z) + o_p(1),\end{aligned}$$

where the last line defines the process $\widetilde{\mathbf{D}}_n(\cdot)$. Obviously, under our assumptions, $E[\widetilde{\mathbf{D}}_n(z)] = 0$. For $s,t \in \mathbb{R}$, straightforward calculation of the covariances yields that $\mathrm{Cov}\{\widetilde{\mathbf{D}}_n(s), \widetilde{\mathbf{D}}_n(t)\} = F(\min(s,t),\theta_0) - F(s,\theta_0)F(t,\theta_0) + H(s,t,\theta_0)$, where $H(\cdot,\cdot,\cdot)$ is defined in Theorem 1. Hence, the covariance function of $\widetilde{\mathbf{D}}_n(\cdot)$ converges to that of $\mathbf{D}(\cdot)$.

To prove weak convergence of process $\mathbf{D}_n(\cdot)$, it suffices to prove weak convergence of $\widetilde{\mathbf{D}}_n(\cdot)$. Let $\ell^\infty(\mathcal{G})$ denote the space of all bounded functions from a set $\mathcal{G}$ to $\mathbb{R}$ equipped with the supremum norm $\|v\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |v(g)|$, and define $\mathcal{G} = \{\delta_z(\cdot),\, z \in \mathbb{R}\}$ as the collection of functions of the form

$$\delta_z(\varepsilon) = I(\varepsilon \le z) + f(z,\theta_0)(\varepsilon + \frac{z}{2}(\varepsilon^2 - 1)) - F_\theta(z,\theta_0)'\psi(X,\varepsilon,\theta_0). \qquad (8)$$

With this notation, observe that

$$\widetilde{\mathbf{D}}_n(z) = n^{-1/2}\sum_{i=1}^{n}(\delta(\varepsilon_i) - E[\delta(\varepsilon_i)])$$

is a $\mathcal{G}$-indexed empirical process in $\ell^\infty(\mathcal{G})$. Proving weak convergence of $\widetilde{\mathbf{D}}_n(\cdot)$ in $\ell^\infty(\mathcal{G})$ entails that the class $\mathcal{G}$ is Donsker. Following Theorem 2.6.8 of van der Vaart and Wellner (1996, p.142), we have to check that $\mathcal{G}$ is pointwise separable, is a Vapnik-Červonenkis class of sets, or simply a VC-class and has an envelope function $\Delta(\cdot)$ with weak second moment[1]. Using the remark in the proof of the aforementioned theorem, the latter condition on the envelope can be promoted to the stronger condition that the envelope has a finite second moment. Pointwise separability of $\mathcal{G}$ follows from p. 116 in van der Vaart and Wellner (1996). More precisely, define the class $\mathcal{G}_1 = \{\delta_z(\cdot),\ z \in \mathbb{Q}\}$, which is a countable dense subset of $\mathcal{G}$ (dense in terms of pointwise convergence). For every sequence $z_m \in \mathbb{Q}$ with $z_m \searrow z$ as $m \longrightarrow \infty$, which means that $z_m$ decreasingly approaches $z$ as $m \longrightarrow \infty$, and $\delta_z(\cdot) \in \mathcal{G}$, we consider the sequence $\delta_{z_m}(\cdot) \in \mathcal{G}_1$. First, for each $\varepsilon \in \mathbb{R}$, the sequence $\delta_{z_m}(\cdot)$ fulfils that $\delta_{z_m}(\varepsilon) \longrightarrow \delta_z(\varepsilon)$ pointwise as $m \longrightarrow \infty$, since $\delta_z(\cdot)$ is right continuous for every $\varepsilon \in \mathbb{R}$. Second, $\delta_{z_m}(\cdot) \longrightarrow \delta_z(\cdot)$ in $L_2(P)$-norm, where $P$ is the probability measure corresponding to the distribution of $\varepsilon$,

$$
\begin{aligned}
&\|\delta_{z_m}(\varepsilon) - \delta_z(\varepsilon)\|_{P,2}^2 \equiv \int |\delta_{z_m}(\varepsilon) - \delta_z(\varepsilon)|^2 f(v,\theta_0)\mathrm{d}v \leq \\
&3[F(z_m,\theta_0) - F(z,\theta_0) + (f(z_m,\theta_0) - f(z,\theta_0))^2 E(\varepsilon^2) \\
&+(z_m f(z_m,\theta_0) - z f(z,\theta_0))^2 E(\varepsilon^2 - 1)^2/4] \\
&+(F_\theta(z_m,\theta_0) - F_\theta(z,\theta_0))'\Omega(F_\theta(z_m,\theta_0) - F_\theta(z,\theta_0)) \\
&-2(F_\theta(z_m,\theta_0) - F_\theta(z,\theta_0))'E\{(I(\varepsilon \leq z_m) - I(\varepsilon \leq z))\psi(X,\varepsilon,\theta_0)\} \\
&-2(f(z_m,\theta_0) - f(z,\theta_0))(F_\theta(z_m,\theta_0) - F_\theta(z,\theta_0))'E\{\psi(X,\varepsilon,\theta_0)\varepsilon\} \\
&-2(z_m f(z_m,\theta_0) - z f(z,\theta_0))(F_\theta(z_m,\theta_0) - F_\theta(z,\theta_0))'E\{\psi(X,\varepsilon,\theta_0)(\varepsilon^2 - 1)\} \\
&\longrightarrow 0 \qquad \text{as} \qquad m \longrightarrow \infty.
\end{aligned}
$$

[1] Consider an arbitrary collection $X_n = \{x_1, ..., x_n\}$ of $n$ points in a set $\mathcal{X}$ and a collection $\mathcal{C}$ of subsets of $\mathcal{X}$. We say that $\mathcal{C}$ *picks out* a certain subset $A$ of $X_n$ if $A = C \cap X_n$ for some $C \in \mathcal{C}$. Additionally, we say that $\mathcal{C}$ *shatters* $X_n$ if all of the $2^n$ subsets of $X_n$ are *picked out* by the sets in $\mathcal{C}$. The VC-index $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest $n$ for which no set $X_n \subset \mathcal{X}$ is shattered by $\mathcal{C}$. We say that $\mathcal{C}$ is a VC-class if $V(\mathcal{C})$ is finite. Finally, a collection $\mathcal{G}$ is a VC-class of functions if the collection of all subgraphs $\{(x,t),\ g(x) < t\}$, where $g$ ranges over $\mathcal{G}$, forms a VC-class of sets in $\mathcal{X} \times \mathbb{R}$. See van der Vaart and Wellner (1996, chapter 2.6) for further details.

For $z \in \mathbb{R}$, we may rewrite (8) as $\delta_z(\varepsilon) = g_1(\varepsilon) + g_2(\varepsilon)$, where $g_1(\varepsilon) = I(\varepsilon \leq z)$ and $g_2(\varepsilon) = f(z, \theta_0)(\varepsilon + \frac{z}{2}(\varepsilon^2 - 1)) - F_\theta(z, \theta_0)'\psi(X, \varepsilon, \theta_0)$. Let us now define the class of all indicator functions of the form $\mathcal{C}_1 = \{\varepsilon \longmapsto I(\varepsilon \leq d),\ d \in \mathbb{R}\}$ such that $g_1(\cdot) \in \mathcal{C}_1$. Consider any two point sets $\{\varepsilon_1,\ \varepsilon_2\} \subset \mathbb{R}$ and assume, without loss of generality, that $\varepsilon_1 < \varepsilon_2$. It is easy to verify that $\mathcal{C}_1$ can *pick out* the null set and the sets $\{\varepsilon_1\}$ and $\{\varepsilon_1,\ \varepsilon_2\}$ but cannot *pick out* $\{\varepsilon_2\}$. Thus, the VC-index $V(\mathcal{C}_1)$ of the class $\mathcal{C}_1$ is equal to 2; and hence $\mathcal{C}_1$ is a VC-class. Note that $\psi(\cdot,\cdot,\cdot) = (\psi_1(\cdot,\cdot,\cdot), ..., \psi_m(\cdot,\cdot,\cdot))$. We define the class of functions $\mathcal{C}_2 = \{\varepsilon \longmapsto a\varepsilon + b(\varepsilon^2 - 1) + c_1\psi_1(X, \varepsilon, \theta_0) + ... + c_m\psi_m(X, \varepsilon, \theta_0) |\ a, b, c_1, ..., c_m \in \mathbb{R}\}$ such that $g_2(\cdot) \in \mathcal{C}_2$. By Lemma 2.6.15 of van der Vaart and Wellner (1996) and assumption 9, for fixed $X \in \mathbb{R}$, the class of functions $\mathcal{C}_2$ is a VC-class with $V(\mathcal{C}_2) \leq dim(\mathcal{C}_2) + 2$. Finally, by Lemma 2.6.18 of van der Vaart and Wellner (1996), the sum of VC-classes builds out a new VC-class. This yields the VC property of $\mathcal{G}$. Recall that an envelope function of a class $\mathcal{G}$ is any function $x \mapsto \Delta(x)$ such that $|\delta_z(x)| \leq \Delta(x)$ for every $x$ and $\delta_z(\cdot)$. Using that $f(\cdot, \theta)$ is bounded away from zero, $\sup_{\varepsilon \in \mathbb{R}} |\varepsilon f(\varepsilon, \theta)| < \infty$ and that $F(\cdot, \cdot)$ has bounded derivative with respect to the second argument, it follows that $\mathcal{G}$ has an envelope function of the form

$$\Delta(\varepsilon) = 1 + \alpha_1\varepsilon + \alpha_2(\varepsilon^2 - 1) - \alpha_3'\psi(X, \varepsilon, \theta_0),$$

where $\alpha = (1,\ \alpha_1,\ \alpha_2,\ \alpha_3')'$ is a $(3 + m) \times 1$ vector of constants. Finally, note that our assumption 9 readily implies that this envelope has a finite second moment, which completes the proof of part **a**.

On the other hand, under our assumptions, $\sup_{z \in \mathbb{R}} |\widehat{F}_n(z) - F_\varepsilon(z)| = o_p(1)$. Also, by applying the mean-value theorem, $F(z, \widehat{\theta}) = F(z, \theta_0) + F_\theta(z, \theta^{**})(\widehat{\theta} - \theta_0)$ for

$\theta^{**}$ a mean value between $\widehat{\theta}$ and $\theta_0$. From assumption 8, under $\mathrm{H}_0$, $\theta_0$ is the true parameter, while under $\mathrm{H}_1$, $\theta_0$ corresponds to the best approximation $F(\cdot,\theta_0)$ of $F_\varepsilon$ in the class $\mathcal{F}$ with respect to the relevant metric $d(\cdot,\cdot)$. From assumption 9, under $\mathrm{H}_0$, the last term is $O_p(n^{-1/2})$, while under the alternative the last term is $o_p(1)$. Thus, irrespective of whether $\mathrm{H}_0$ holds true or not, $\sup_{z\in\mathbb{R}}|F(z,\widehat{\theta})-F(z,\theta_0)| = o_p(1)$. Therefore $\sup_{z\in\mathbb{R}}|\widehat{F}_n(z)-F(z,\widehat{\theta})| \xrightarrow{p} \sup_{z\in\mathbb{R}}|F_\varepsilon(z)-F(z,\theta_0)|$. Under $\mathrm{H}_1$, $\sup_{z\in\mathbb{R}}|F_\varepsilon(z)-F(z,\theta_0)|>0$ and this concludes the proof of part **b**.

For the second test statistic observe that $\widehat{C}_n=\int\{\widehat{F}_n(v)-F(v,\widehat{\theta})\}^2\mathrm{d}\widehat{F}_n(v)$. As before, the asymptotic distribution of this statistic can be obtained from Proposition 1 and the uniform convergence of $\widehat{F}_n(\cdot)$. ■

The following two propositions are required in the proof of Theorem 2.

**Proposition A1:** *Suppose that assumptions 1-11 hold. Then,*

$$\int_{-\infty}^{+\infty}||q(u,\widehat{\theta})-q(u,\theta_0)||^2 f(u,\theta_0)\mathrm{d}u=o_p(1).$$

**Proof:** Under assumption 7, $q(\cdot,\cdot)$ is continuously differentiable with respect to $\theta$. Thus, by a Taylor expansion we obtain $q(u,\widehat{\theta})=q(u,\theta_0)+q_\theta(u,\theta^*)(\widehat{\theta}-\theta_0)/2$, where $\theta^*$ lies between $\widehat{\theta}$ and $\theta_0$. Observe that

$$\int_{-\infty}^{+\infty}||q(u,\widehat{\theta})-q(u,\theta_0)||^2 f(u,\theta_0)\mathrm{d}u \le \tfrac{1}{4}||\widehat{\theta}-\theta_0||^2\int_{-\infty}^{+\infty}||q_\theta(u,\theta^*)||^2 f(u,\theta_0)\mathrm{d}u$$
$$\le \tfrac{1}{4}||\widehat{\theta}-\theta_0||^2\sup_{v\in N(\theta_0,M_0)}\int_{-\infty}^{+\infty}||q_\theta(u,v)||^2 f(u,\theta_0)\mathrm{d}u=o_p(1),$$

where the first inequality follows using $||q(u,\widehat{\theta})-q(u,\theta_0)||^2\le||q_\theta(u,\theta^*)||^2||\widehat{\theta}-\theta_0||^2/4$, and the last equality follows using assumptions 9 and 11. More precisely, using assumption 9, it is straightforward to show that $(\widehat{\theta}-\theta_0)=O_p(n^{-1/2})$ under $\mathrm{H}_0$. Then, $||\widehat{\theta}-\theta_0||^2=O_p(n^{-1})$, and we get $\frac{1}{4}O_p(n^{-1})O(1)=o_p(1)$. Under the alternative hypothesis $\mathrm{H}_1$, from assumption 9, $\widehat{\theta}-\theta_0=o_p(1)$, and we get $\frac{1}{4}o_p(1)O(1)=o_p(1)$. ■

**Proposition A2:** *Suppose that assumptions 1-11 hold. Then,*

$$\sup_{z\in\mathbb{R}} || n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)\{q(\varepsilon_i,\widehat{\theta}) - q(\varepsilon_i,\theta_0)\} - \int_z^{+\infty}\{q(u,\widehat{\theta}) - q(u,\theta_0)\}f(u,\theta_0)\mathrm{d}u]|| = o_p(1).$$

**Proof:** As above, under assumption 7, $q(\cdot,\cdot)$ is continuously differentiable with respect to $\theta$. Thus, by a Taylor expansion we obtain $q(u,\widehat{\theta}) = q(u,\theta_0) + q_\theta(u,\theta^*)(\widehat{\theta}-\theta_0)/2$, where $\theta^*$ lies between $\widehat{\theta}$ and $\theta_0$. Therefore,

$$n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)\{q(\varepsilon_i,\widehat{\theta}) - q(\varepsilon_i,\theta_0)\} - \int_z^{+\infty}\{q(u,\widehat{\theta}) - q(u,\theta_0)\}f(u,\theta_0)\mathrm{d}u]$$
$$= n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)\{q(\varepsilon_i,\widehat{\theta}) - q(\varepsilon_i,\theta_0)\} - E(I(\varepsilon \geq z)\{q(\varepsilon,\widehat{\theta}) - q(\varepsilon,\theta_0)\})]$$
$$= n^{-1}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)q_\theta(\varepsilon_i,\theta^*) - E(I(\varepsilon \geq z)q_\theta(\varepsilon,\theta^*)]n^{1/2}(\widehat{\theta}-\theta_0)/2.$$

By assumption 9, it is straightforward to show that, under $\mathrm{H}_0$, $n^{1/2}(\widehat{\theta}-\theta_0)$ is $O_p(1)$ and the remaining term is $o_p(1)$ using some uniform strong law of large numbers. On the other hand, under $\mathrm{H}_1$, $\widehat{\theta}-\theta_0 = o_p(1)$ and the remaining term is $O_p(1)$ using some Central Limit Theorem. The result holds given that $O_p(1)o_p(1) = o_p(1)$. ■

**Proof of Theorem 2:** In the following reasoning we assume that the null hypothesis holds. Let $t = F(z,\theta_0)$, then $z = F^{-1}(t,\theta_0)$. Interchanging the variables, we shall first show that $\overline{\mathbf{W}}_n(\cdot) \equiv \mathbf{W}_n(F^{-1}(\cdot,\theta_0))$ converges weakly to a standard Brownian motion. Let $D[0,b]$ $(b>0)$ denote the space of cadlag functions on $[0,b]$ endowed with the Skorohod metric. Furthermore, define the linear mapping $\Gamma : D[0,1] \to D[0,1]$ as follows

$$\Gamma(\alpha(\cdot))(t) \equiv \int_0^t q(F^{-1}(s,\theta_0),\theta_0)'C(F^{-1}(s,\theta_0),\theta_0)^{-1}[\int_s^1 q(F^{-1}(r,\theta_0),\theta_0)\mathrm{d}\alpha(r)]\mathrm{d}s.$$

Let $Q(t) = (Q_1(t),Q_2(t),Q_3(t),Q_4(t))' = (t, f(F^{-1}(t,\theta_0)), f(F^{-1}(t,\theta_0))F^{-1}(t,\theta_0), F_\theta(F^{-1}(t,\theta_0))')'$; so $q(F^{-1}(\cdot,\theta_0))$ is the derivative of $Q(\cdot)$. Then it follows that

$$\Gamma(Q_l(\cdot)) = Q_l(\cdot), \quad \text{for} \quad l = 1,2,3,4. \tag{9}$$

From $C(F^{-1}(s,\theta_0))^{-1}C(F^{-1}(s,\theta_0)) = \mathbf{I}_4$, we have $C(F^{-1}(s,\theta_0))^{-1}\{\int_s^1 q(r)\mathrm{d}Q_1(r)\}$ $= (1,0,0,0)'$. Thus $\Gamma(Q_1(\cdot))(t) = \int_0^t q(s)'(1,0,0,0)'\mathrm{d}s = Q_1(t)$. A parallel analysis establishes similar results for the remaining components of $Q(\cdot)$.

Let $\widehat{t} = F(F^{-1}(t),\widehat{\theta})$. Thus $\widehat{\mathbf{V}}_n(t) = n^{1/2}[\widehat{F}_n(t)-t]+n^{1/2}[t-\widehat{t}]$. Note that $\widehat{\mathbf{V}}_n(\cdot)$ can be rewritten as follows

$$\widehat{\mathbf{V}}_n(\cdot) = n^{1/2}[\widehat{F}_n(F^{-1}(\cdot,\theta_0)) - Q_1(\cdot)] + n^{1/2}[Q_1(\cdot) - F(F^{-1}(Q_1(\cdot),\theta_0),\widehat{\theta})].$$

Using the linearity of $\Gamma(\cdot)$, (6) and (7), routine calculations yield that $\overline{\mathbf{W}}_n(\cdot) = \widehat{\mathbf{V}}_n(\cdot) - \Gamma(\widehat{\mathbf{V}}_n(\cdot))$. Using Proposition 1, the linearity of $\Gamma(\cdot)$ and (6), it follows that

$$\Gamma(\widehat{\mathbf{V}}_n(z)) = \Gamma(\mathbf{V}_n(z)) + n^{-1/2}\sum_{i=1}^n [f(z,\theta_0)(\varphi_{1n}(X_i,Y_i)+\beta_{1n})$$
$$+zf(z,\theta_0)(\varphi_{2n}(X_i,Y_i)+\beta_{2n})] - F_\theta(z,\theta_0)'n^{1/2}(\widehat{\theta}-\theta_0) + o_p(1).$$

Notice that the bias term $\beta_n(\cdot) = f(z,\theta_0)\beta_{1n} + zf(z,\theta_0)\beta_{2n}$ can be omitted if $nh_n^4 = o(1)$. Using Proposition 1 again, we have $\overline{\mathbf{W}}_n(\cdot) = \mathbf{V}_n(\cdot) - \Gamma(\mathbf{V}_n(\cdot)) + o_p(1) + o(1)$. Thus, as $\mathbf{V}_n(\cdot)$ converges weakly to a standard Brownian bridge $B(\cdot)$ on $[0,1]$, $\overline{\mathbf{W}}_n(\cdot)$ converges weakly to $B(\cdot) - \Gamma(B(\cdot))$, which is a standard Brownian motion on $[0,1]$ (see Khamaladze, 1981 or Bai, 2003, p. 543).

Let us now define $\widetilde{\overline{\mathbf{W}}}_n(\cdot) \equiv \widehat{\mathbf{W}}_n(F^{-1}(\cdot,\theta_0))$. Observe that propositions A1 and A2 imply that assumption D1 of Bai (2003) holds. Hence, to prove that $\widetilde{\overline{\mathbf{W}}}_n(\cdot) = \overline{\mathbf{W}}_n(\cdot) + o_p(1)$, we follow exactly the lines of the proof of Theorem 4 of Bai (2003). Introduce $\overline{C}(u) = \int_u^{+\infty} q(\tau,\widehat{\theta})q(\tau,\widehat{\theta})'f(\tau,\theta)\mathrm{d}\tau$. Thus, for every

$t_0 \in (0,1)$, it follows that

$$\sup_{t\in[0,t_0]} |\widetilde{\overline{\mathbf{W}}}_n(t) - \overline{\mathbf{W}}_n(t)| \ \le$$
$$\sup_{z\in(-\infty,F^{-1}(t_0,\theta_0))} n^{1/2} | \int_{-\infty}^{z} q(u,\widehat{\theta})' C(u,\widehat{\theta})^{-1} \overline{d}_n(u,\widehat{\theta})\{f(u,\widehat{\theta}) - f(u,\theta_0)\}\mathrm{d}u| \ +$$
$$\sup_{z\in(-\infty,F^{-1}(t_0,\theta_0))} n^{1/2} | \int_{-\infty}^{z} q(u,\widehat{\theta})' \{C(u,\widehat{\theta})^{-1} - \overline{C}(u)^{-1}\} \overline{d}_n(u,\widehat{\theta}) f(u,\theta_0)\mathrm{d}u| \ +$$
$$\sup_{z\in(-\infty,F^{-1}(t_0,\theta_0))} n^{1/2} | \int_{-\infty}^{z} [q(u,\widehat{\theta})' \overline{C}(u)^{-1} \overline{d}_n(u,\widehat{\theta})$$
$$-q(u,\theta_0)' C(u,\theta_0)^{-1} \overline{d}_n(u,\theta_0)] f(u,\theta_0)\mathrm{d}u| \ \equiv (I) + (II) + (III).$$

We next show that $(I)$, $(II)$ and $(III)$ are all small under $\mathrm{H}_0$. We first prove that $(III)$ is $o_p(1)$. For ease of notation we write $\widehat{q} \equiv q(F^{-1}(\cdot,\theta_0),\widehat{\theta})$, $\widehat{C} \equiv C(F^{-1}(\cdot,\theta_0),\widehat{\theta})$, $\overline{C} \equiv \overline{C}(F^{-1}(\cdot,\theta_0))$, $q \equiv q(F^{-1}(\cdot,\theta_0),\theta_0)$, $C \equiv C(F^{-1}(\cdot,\theta_0),\theta_0)$. From $\overline{C}^{-1}\overline{C} = \mathbf{I}_4$, we have $\overline{C}^{-1} \int_s^1 q(F^{-1}(r,\theta_0),\widehat{\theta})\mathrm{d}r = (1,0,0,0)'$, since the first column of $\overline{C}$ is $\int_s^1 q(F^{-1}(r,\theta_0),\widehat{\theta})\mathrm{d}r$. Hence, $\int_0^t \widehat{q}' \overline{C}^{-1} \int_s^1 \{q(F^{-1}(r,\theta_0),\widehat{\theta})\mathrm{d}r = t$. Analogously, from (9), we have $\int_0^t q' C^{-1} \int_s^1 \{q(F^{-1}(r,\theta_0),\theta_0)\mathrm{d}r = t$ for $l = 1$. Thus, $(III)$ can also be expressed as follows

$$(III) = \sup_{t\in[0,t_0)} | \int_0^t [\widehat{q}' \overline{C}^{-1} \int_s^1 q(F^{-1}(r,\theta_0),\widehat{\theta}) \mathrm{d}\widehat{\mathbf{V}}_n(r)$$
$$-q' C^{-1} \int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)]\mathrm{d}s| \ \le$$
$$\sup_{t\in[0,t_0)} \int_0^t |\widehat{q}' \overline{C}^{-1} \int_s^1 \{q(F^{-1}(r,\theta_0),\widehat{\theta}) - q(F^{-1}(r,\theta_0),\theta_0)\} \mathrm{d}\widehat{\mathbf{V}}_n(r)|\mathrm{d}s + \tag{10}$$
$$\sup_{t\in[0,t_0)} \int_0^t |\widehat{q}' \{\overline{C}^{-1} - C^{-1}\} \int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)|\mathrm{d}s +$$
$$\sup_{t\in[0,t_0)} \int_0^t |\{\widehat{q} - q\}' C^{-1} \int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)|\mathrm{d}s.$$

Now observe that using Proposition 1, assumption 6 and $z = F^{-1}(t,\theta_0)$, for the estimated empirical process we can write

$$\widehat{\mathbf{V}}_n(t) = \mathbf{V}_n(t) + g(t)' n^{-1/2} \Sigma_n + o_p(1) + o(1). \tag{11}$$

where $g(t) \equiv (Q_2(t), Q_3(t), Q_4(t))'$ and $\Sigma_n = (\sum_{i=1}^n \varphi_{1n}(X_i,Y_i),\ \sum_{i=1}^n \varphi_{2n}(X_i,Y_i),\ -(\widehat{\theta} - \theta_0))'$. We next write equation (11) in its differential form

$$\mathrm{d}\widehat{\mathbf{V}}_n(t) = \mathrm{d}\mathbf{V}_n(t) + \dot{g}(t)'\mathrm{d}t\ \Sigma_n + o_p(1), \tag{12}$$

where $\dot{g}(\cdot)$ denotes the derivative of $g(\cdot)$. Therefore, for all $s$ in $(0, t_0)$, applying Cauchy-Schwarz inequality and (12), we derive that

$$\begin{gathered}
|\widehat{q}'\overline{C}^{-1} \int_s^1 \{q(F^{-1}(r,\theta_0),\widehat{\theta}) - q(F^{-1}(r,\theta_0),\theta_0)\} \mathrm{d}\widehat{\mathbf{V}}_n(r)| \;\leq \\
||\widehat{q}||\; ||\overline{C}^{-1}||\; (||\int_s^1 \{q(F^{-1}(r,\theta_0),\widehat{\theta}) - q(F^{-1}(r,\theta_0),\theta_0)\} \mathrm{d}\mathbf{V}_n(r)|| + \\
\{\int_s^1 ||q(F^{-1}(r,\theta_0),\widehat{\theta}) - q(F^{-1}(r,\theta_0),\theta_0)||^2 \mathrm{d}r\}^{1/2} \;\times \\
[\{\int_s^1 ||\dot{g}(r)||^2 \mathrm{d}r\}^{1/2} O_p(1) + o_p(1)]) = ||\widehat{q}|| O_p(1) o_p(1),
\end{gathered} \tag{13}$$

where the last equality follows from propositions A2 and A1. Similarly, reasoning as in (13),

$$\begin{gathered}
|\widehat{q}'\{\overline{C}^{-1} - C^{-1}\} \int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)| \;\leq \\
||\widehat{q}||\; ||\overline{C}^{-1} - C^{-1}||\; ||\int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)|| \;\leq \\
||\widehat{q}||\; ||\overline{C}^{-1} - C^{-1}||\; \{||\int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\mathbf{V}_n(r)|| + O_p(1)\} \\
= ||\widehat{q}|| o_p(1) O_p(1).
\end{gathered} \tag{14}$$

To see the last equality, note that $\int_s^1 q(F^{-1}(r,\theta_0),\theta_0)\mathrm{d}\mathbf{V}_n(r) = O_p(1)$ by the functional central limit theorem and, $||\overline{C}^{-1} - C^{-1}|| = o_p(1)$, uniformly in $s$, using the same argument as in Bai (2003, p. 548). Finally,

$$\begin{gathered}
|\{\widehat{q} - q\}' C^{-1} \int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)| \;\leq \\
||\widehat{q} - q||\; ||C^{-1}||\; ||\int_s^1 q(F^{-1}(r,\theta_0),\theta_0) \mathrm{d}\widehat{\mathbf{V}}_n(r)|| = ||\widehat{q} - q|| O_p(1) O_p(1).
\end{gathered} \tag{15}$$

Therefore, from (10), (13), (14) and (15), we have

$$\begin{aligned}
(III) \;&\leq\; o_p(1)(\int_0^1 ||\widehat{q}||^2 \mathrm{d}s)^{1/2} + O_p(1)(\int_0^1 ||\widehat{q} - q||^2 \mathrm{d}s)^{1/2} \\
&=\; o_p(1)O_p(1) + O_p(1)o_p(1) = o_p(1),
\end{aligned}$$

where the last equality follows from propositions A1.

To analyze $(I)$ and $(II)$, observe that, under assumptions 7 and 9, $f(\cdot, \widehat{\theta}) = f(\cdot, \theta_0) + o_p(1)$ (this follows applying a Taylor expansion). Hence, $\widehat{C} = \overline{C} + o_p(1)$.

Using the same arguments as above it is straightforward to show that $(I) = o_p(1)$ and $(II) = o_p(1)$. Thus, under $\mathrm{H}_0$, $\widetilde{\widetilde{\mathbf{W}}}_n(\cdot)$ also converges weakly to a Brownian motion $\mathbf{W}^{(1)}(\cdot)$ in the space $D[0, t_0]$; hence, the martingale-transformed statistic $\overline{K}_{n,z_0} \equiv F(z_0, \widehat{\theta})^{-1/2} \sup_{t\in[0,F(z_0,\theta_0)]} \left|\widetilde{\widetilde{\mathbf{W}}}_n(t)\right|$ converges in distribution to $F(z_0, \theta_0)^{-1/2} \sup_{t\in[0,F(z_0,\theta_0)]} \left|\mathbf{W}^{(1)}(t)\right| = \sup_{t\in[0,1]} |\mathbf{W}(t)|$, where we denote $\mathbf{W}(t) \equiv F(z_0, \theta_0)^{-1/2}\mathbf{W}^{(1)}(F(z_0, \theta_0)t)$, which is a Brownian motion in the space $D[0, 1]$. Similarly, from Proposition 1 and the uniform convergence of $\widehat{F}_n(\cdot)$, $\overline{C}_{n,z_0} \equiv F(z_0, \widehat{\theta})^{-2}n^{-1} \sum_{i=1}^{n} I(\widehat{\varepsilon}_i \leq z_0)\widehat{\mathbf{W}}_n(\widehat{\varepsilon}_i)^2 = F(z_0, \widehat{\theta})^{-2} \int I(z \leq z_0)\{\widehat{\mathbf{W}}_n(z)\}^2 \mathrm{d}F_\varepsilon(z) + o_p(1)$, and hence $\overline{C}_{n,z_0}$ converges in distribution to $F(z_0, \theta_0)^{-2} \int_0^{F(z_0,\theta_0)} \{\mathbf{W}^{(1)}(t)\}^2 \mathrm{d}t = \int_{[0,1]} \{\mathbf{W}(t)\}^2 \mathrm{d}t$. This completes the proof of part **a**.

On the other hand, under $\mathrm{H}_1$, the assertion can be deduced from the probability limit of $n^{-1/2}\widehat{\mathbf{W}}_n(z)$, which is

$$\mathbf{\Upsilon}(z) \equiv F_\varepsilon(z) - \int_{-\infty}^{z} q(u, \theta_0)' C(u, \theta_0)^{-1} \{\int_{u}^{+\infty} q(\tau, \theta_0) \mathrm{d}F_\varepsilon(\tau)\} f(u, \theta_0) \mathrm{d}u.$$

Let us first assume that $\mathbf{\Upsilon}(z) = 0$ for every $z \in \mathbb{R}$; if this is the case, then $\partial\mathbf{\Upsilon}(z)/\partial z = 0$, and this amounts to saying that

$$f_\varepsilon(z) - q(z, \theta_0)'\Pi(z)f(z, \theta_0) = 0, \tag{16}$$

where we define $\Pi(z) = C(z, \theta_0)^{-1}\{\int_z^{+\infty} q(\tau, \theta_0)f_\varepsilon(\tau)\mathrm{d}\tau\}$, and $f_\varepsilon(\cdot)$ denotes the density function of $\varepsilon$. Let us show that $\Pi(z)$ is constant: by the fundamental theorem of calculus and the rules from matrix derivation, we have

$$\begin{gathered}\tfrac{\partial}{\partial z}\Pi(z) = C(z, \theta_0)^{-1} \tfrac{\partial}{\partial z} C(z, \theta_0)C(z, \theta_0)^{-1}\{\textstyle\int_z^{+\infty} q(\tau, \theta_0)f_\varepsilon(\tau)\mathrm{d}\tau\} \\ -C(z, \theta_0)^{-1}q(z, \theta_0)f_\varepsilon(z) = -C(z, \theta_0)^{-1}q(z, \theta_0) \tfrac{\partial}{\partial z}\mathbf{\Upsilon}(z) = 0,\end{gathered}$$

where the second equality follows using $\partial C(z, \theta_0)/\partial z = -q(z, \theta_0)q(z, \theta_0)'f(z, \theta_0)$, and the last equality follows from $\partial\mathbf{\Upsilon}(z)/\partial z = 0$. Thus it follows that $\Pi(z) =$

$(\Pi_1, \Pi_2, \Pi_3, \Pi_4)'$ is constant, where $\Pi_4 = (\Pi_{41}, ..., \Pi_{4m})'$. From (16), it follows that

$$f_\varepsilon(z) = (\Pi_1 + \Pi_3) f(z, \theta_0) + \Pi_2 f_z(z, \theta_0) + \Pi_3 z f_z(z, \theta_0) + f_\theta(z, \theta)' \Pi_4. \qquad (17)$$

If we integrate the two terms in (17), and also these two terms premultiplied by $z$, $z^2$ and $z^3$, we derive a system of four linear equations in four unknowns. Under our assumptions, which ensure that the integration and differentiation operators can be exchanged, the only solution to this system is $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4 = 0$; this implies that $f_\varepsilon(z) = f(z, \theta_0)$. Thus, we have proved that if $\mathbf{\Upsilon}(z) = 0$ for every $z \in \mathbb{R}$, then $\mathrm{H}_0$ holds; therefore, under $\mathrm{H}_1$, there exists $z_* \in R$ such that $\mathbf{\Upsilon}(z_*) \neq 0$, and if $z_0 \geq z_*$ then $n^{-1/2}\overline{K}_{n,z_0}$ converges in probability to $F(z_0, \theta_0)^{-1/2} \sup_{z \in (-\infty, z_0]} |\mathbf{\Upsilon}(z)| > 0$, and $n^{-1/2}\overline{C}_{n,z_0}$ converges in probability to $F(z_0, \theta_0)^{-2} \int_{-\infty}^{z_0} \mathbf{\Upsilon}(z)^2 \mathrm{d}F_\varepsilon(z) > 0$. The result in part **b** follows from here. ■

## REFERENCES

Akritas, M. and van Keilegom, I., 2001. Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics* 28, 549-567.

Bai, J., 2003. Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85, 531-549.

Dette, H., Neumeyer, N. and Kusi-Appiah, S., 2002. Testing symmetry in non-parametric regression models. *Journal of Nonparametric Statistics* 14, 477-494.

Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics* 1, 279-290.

Fernández, C. and Steel, M.F.J., 1998. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93, 359-371.

Franke, J., Kreiss, J-P., and Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* 8, 1-37.

Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, forthcoming.

Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability and its Applications* 26, 240-257.

Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovations martingales. *Annals of Statistics* 21, 798-829.

Koul, H.L. and Sakhanenko, L., 2005. Goodness-of-fit testing in regression: A finite sample comparison of bootstrap methodology and Khmaladze transformation. *Statistics and Probability Letters* 74, 290-302.

Loynes, R.M., 1980. The empirical distribution function of residuals from generalised regression. *Annals of Statistics* 8, 285-298.

Mora, J. and Neumeyer, N., 2008. The two-sample problem with regression errors: An empirical process approach. Mimeo..

Neumeyer, N. and Dette, H., 2007. Testing for symmetric error distribution in nonparametric regression models. *Statistica Sinica* 17, 775-795.

Neumeyer, N., Dette, H. and Nagel, E.R., 2005. Bootstrap tests for the error distribution in linear and nonparametric regression models. *Australian and New Zealand Journal of Statistics* 48, 129-156.

Rohatgi, V.K. and Saleh, A.K.M.E., 2001. *An Introduction to Probability and Statistics* (2nd Ed.). New York: Wiley-Interscience.

Rothman, E.D. and Woodroofe, M., 1972. A Cramér-von Mises type statistic for testing symmetry. *Annals of Mathematical Statistics* 43, 2035-2038.

Shorack, G. and Wellner, J.A., 1986. *Empirical Processes with Applications to Statistics*. New York: Wiley.

Tsigroshvili, A., 1998. Some notes on goodness-of-fit and innovations martingales. *Proceedings of A. Razmadze Mathematical Institute* 117, 89-102.

Van der Vaart, A.W. and Wellner, J.A., 1996. *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

**TABLE 1: Proportion of Rejections of $H_0$**

| $\delta$ | $h^{(1)}$ | $h^{(2)}$ | $h^{(3)}$ | $h^{(4)}$ | $h^{(1)}$ | $h^{(2)}$ | $h^{(3)}$ | $h^{(4)}$ |
|---|---|---|---|---|---|---|---|---|
| | $n = 100$ | | | | $n = 500$ | | | |
| | | | | $\alpha = 0.010$ | | | | |
| 0 | 0.005 | 0.004 | 0.003 | 0.003 | 0.007 | 0.006 | 0.008 | 0.006 |
| 1/12 | 0.015 | 0.032 | 0.016 | 0.026 | 0.095 | 0.126 | 0.149 | 0.162 |
| 1/9 | 0.049 | 0.037 | 0.027 | 0.034 | 0.254 | 0.307 | 0.339 | 0.357 |
| 1/7 | 0.072 | 0.069 | 0.064 | 0.064 | 0.419 | 0.491 | 0.521 | 0.535 |
| 1/5 | 0.144 | 0.132 | 0.130 | 0.151 | 0.712 | 0.769 | 0.792 | 0.803 |
| 1/3 | 0.369 | 0.376 | 0.371 | 0.376 | 0.988 | 0.994 | 0.995 | 0.996 |
| | | | | $\alpha = 0.050$ | | | | |
| 0 | 0.015 | 0.018 | 0.013 | 0.015 | 0.045 | 0.037 | 0.038 | 0.040 |
| 1/12 | 0.044 | 0.060 | 0.049 | 0.059 | 0.177 | 0.232 | 0.258 | 0.280 |
| 1/9 | 0.079 | 0.090 | 0.074 | 0.075 | 0.378 | 0.455 | 0.486 | 0.499 |
| 1/7 | 0.126 | 0.127 | 0.111 | 0.105 | 0.555 | 0.618 | 0.650 | 0.663 |
| 1/5 | 0.216 | 0.202 | 0.201 | 0.238 | 0.821 | 0.865 | 0.884 | 0.892 |
| 1/3 | 0.484 | 0.494 | 0.493 | 0.481 | 0.996 | 0.998 | 0.998 | 0.998 |
| | | | | $\alpha = 0.100$ | | | | |
| 0 | 0.041 | 0.053 | 0.041 | 0.044 | 0.083 | 0.076 | 0.077 | 0.079 |
| 1/12 | 0.074 | 0.086 | 0.075 | 0.081 | 0.237 | 0.308 | 0.344 | 0.359 |
| 1/9 | 0.110 | 0.122 | 0.109 | 0.114 | 0.460 | 0.527 | 0.556 | 0.570 |
| 1/7 | 0.174 | 0.165 | 0.161 | 0.150 | 0.629 | 0.690 | 0.719 | 0.736 |
| 1/5 | 0.269 | 0.266 | 0.250 | 0.288 | 0.873 | 0.902 | 0.909 | 0.917 |
| 1/3 | 0.569 | 0.578 | 0.568 | 0.554 | 0.998 | 1.000 | 1.000 | 1.000 |

**TABLE 2: Proportion of Rejections of $H_0$**

| $\gamma$ | $h^{(1)}$ | $h^{(2)}$ | $h^{(3)}$ | $h^{(4)}$ | $h^{(1)}$ | $h^{(2)}$ | $h^{(3)}$ | $h^{(4)}$ |
|---|---|---|---|---|---|---|---|---|
| | $n = 100$ | | | | $n = 500$ | | | |
| | $\alpha = 0.010$ | | | | | | | |
| 1 | 0.005 | 0.007 | 0.006 | 0.006 | 0.009 | 0.009 | 0.007 | 0.007 |
| 1.25 | 0.026 | 0.028 | 0.031 | 0.035 | 0.174 | 0.172 | 0.160 | 0.171 |
| 1.5 | 0.077 | 0.074 | 0.078 | 0.086 | 0.305 | 0.307 | 0.321 | 0.322 |
| 1.75 | 0.127 | 0.136 | 0.133 | 0.147 | 0.684 | 0.702 | 0.704 | 0.717 |
| 2 | 0.189 | 0.198 | 0.201 | 0.211 | 0.902 | 0.924 | 0.921 | 0.913 |
| | $\alpha = 0.050$ | | | | | | | |
| 1 | 0.034 | 0.037 | 0.041 | 0.041 | 0.046 | 0.049 | 0.048 | 0.052 |
| 1.25 | 0.064 | 0.061 | 0.062 | 0.064 | 0.273 | 0.287 | 0.271 | 0.274 |
| 1.5 | 0.110 | 0.101 | 0.103 | 0.118 | 0.418 | 0.425 | 0.416 | 0.405 |
| 1.75 | 0.174 | 0.189 | 0.182 | 0.187 | 0.773 | 0.821 | 0.818 | 0.832 |
| 2 | 0.292 | 0.310 | 0.302 | 0.305 | 0.957 | 0.975 | 0.969 | 0.964 |
| | $\alpha = 0.100$ | | | | | | | |
| 1 | 0.067 | 0.074 | 0.071 | 0.069 | 0.094 | 0.096 | 0.099 | 0.104 |
| 1.25 | 0.102 | 0.124 | 0.120 | 0.109 | 0.344 | 0.361 | 0.342 | 0.350 |
| 1.5 | 0.162 | 0.177 | 0.174 | 0.181 | 0.518 | 0.542 | 0.521 | 0.522 |
| 1.75 | 0.262 | 0.283 | 0.271 | 0.285 | 0.924 | 0.942 | 0.927 | 0.926 |
| 2 | 0.356 | 0.374 | 0.361 | 0.367 | 1.000 | 1.000 | 0.999 | 0.999 |