European University Institute

Max Weber Programme

EUI Working Papers

MWP 2007/34

Specification tests for the distribution of errors in nonparametric regression: a martingale approach

Juan Mora Alicia Pérez-Alonso

EUROPEAN UNIVERSITY INSTITUTE

MAX WEBER PROGRAMME

Specification tests for the distribution of errors in nonparametric regression: a martingale approach

JUAN MORA   ALICIA PEREZ-ALONSO

EUI Working Paper MWP No. 2007/34

This text may be downloaded for personal research purposes only. Any additional reproduction for other purposes, whether in hard copy or electronically, requires the consent of the author(s), editor(s). If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper or other series, the year, and the publisher.

The author(s)/editor(s) should inform the Max Weber Programme of the EUI if the paper is to be published elsewhere, and should also assume responsibility for any consequent obligation(s).

ISSN 1830-7728

© 2007 Juan Mora Alicia Pérez-Alonso

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

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 the resulting test statistics are asymptotically distribution-free, and a set of Monte Carlo experiments shows that they work reasonably well in practice.

Keywords

Empirical Process; Nonparametric residual; Martingale Transform; Monte Carlo simulation.

Specification tests for the distribution of errors in nonparametric regression: a martingale approach

Juan Mora* and Alicia Pérez-Alonso†

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

*J. Mora (corresponding author): Departamento de Fundamentos del Análisis Económico, Universidad de Alicante. *E-mail:* juan@merlin.fae.ua.es.

†A. Pérez-Alonso: Department of Economics and Max Weber Programme, European University Institute. *E-mail*: alicia.perez-alonso@eui.eu.

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 us describe the specific framework that we consider. 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 $\sigma^2(x) \equiv \mathrm{Var}(Y|X = x)$. 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 ε, while making no parametric assumptions about the conditional mean function $m(\cdot)$ or the conditional variance function $\sigma^2(\cdot)$. Specifically, if $F(\cdot)$ denotes the c.d.f. of ε 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 θ, we propose a testing procedure to face the hypotheses

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

assuming that independent and identically distributed 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 θ), 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 ε were observable and parameter θ_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 θ_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 θ_0 were known, we could use as a statistic for our test the asymptotic Kolmogorov-Smirnov statistic K_n or the Cramér-von Mises statistic C_n, which are defined by

$$\begin{aligned} K_n &\equiv n^{1/2} \sup_{z \in \mathbb{R}} \left| F_n(z) - F(z, \theta_0) \right|, \\ C_n &\equiv \sum_{i=1}^{n} \{ F_n(\varepsilon_i) - F(\varepsilon_i, \theta_0) \}^2, \end{aligned}$$

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.

$$\begin{aligned} \widehat{m}(x) &= \sum_{i=1}^{n} W_i(x, h_n) Y_i, \\ \widehat{\sigma}^2(x) &= \sum_{i=1}^{n} W_i(x, h_n) Y_i^2 - \widehat{m}(x)^2, \end{aligned}$$

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 $\widehat{\theta}$; we discuss below the asymptotic properties that $\widehat{\theta}$ must satisfy. Using this estimator and the nonparametric residuals, we can define now the statistics

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

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

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

This process will be 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 $\sup_{x,y} |y f(y|x)| < \infty$, $\sup_{x,y} |y \frac{\partial}{\partial x} F(y|x)| < \infty$, $\sup_{x,y} |y^2 \frac{\partial}{\partial y} f(y|x)| < \infty$, $\sup_{x,y} |y^2 \frac{\partial^2}{\partial x^2} F(y|x)| < \infty$.

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: If H_0 holds, then there exists a function $\psi(\cdot, \cdot, \cdot)$ such that $n^{1/2}(\widehat{\theta} - \theta_0) = n^{-1/2} \sum_{i=1}^n \psi(X_i, \varepsilon_i, \theta_0) + o_p(1)$. Additionally, $E\{\psi(X, \varepsilon, \theta_0)\} = 0$, $\Omega \equiv E\{\psi(X, \varepsilon, \theta_0)\psi(X, \varepsilon, \theta_0)'\}$ is finite, $\psi(\cdot, \cdot, \cdot)$ is twice continuously differentiable with respect to the second argument and $\sup_{z\in\mathbb{R}} |\frac{\partial^2}{\partial z^2}\psi(x, z, \theta)| < \infty$.

Assumption 9: If H_1 holds, then there exists $\theta_* \in \mathbb{R}^m$ such that $n^{1/2}(\widehat{\theta} - \theta_*) = O_p(1)$.

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 parametric estimator $\widehat{\theta}$. Assumptions 8-9 ensure that the parametric estimator behaves properly both under H_0 and H_1.

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

Proposition 1: *If* H_0 *holds and assumptions 1-8 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

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

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

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

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

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 θ_0. The following theorem states the asymptotic behaviour of $\widehat{K}_n$ and $\widehat{C}_n$.

Theorem 1: *Suppose that assumptions 1-7 hold. Then:*

a) If H_0 *holds and assumption 8 is satisfied then:*

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

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 H_1 *holds and assumption 9 is satisfied 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, in the next section we propose to consider test statistics that are based on a martingale transform of the estimated empirical process, in the spirit of Khmaladze (1993), Bai (2003) and Khmaladze and Koul (2004).

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

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

where

$$\begin{aligned}
&q(u) \equiv (1, f_u(u,\theta_0)/f(u,\theta_0), 1 + u f_u(u,\theta_0)/f(u,\theta_0), f_\theta(u,\theta_0)'/f(u,\theta_0))', \\
&C(u) \equiv \textstyle\int_u^{+\infty} q(\tau)q(\tau)' f(\tau,\theta_0)\mathrm{d}\tau, \\
&\overline{d}_n(u) \equiv \textstyle\int_u^{+\infty} q(\tau)\mathrm{d}\widehat{F}_n(\tau) = n^{-1}\sum_{i=1}^{n} I(\widehat{\varepsilon}_i \geq u) q(\widehat{\varepsilon}_i),
\end{aligned}$$

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 θ_0, we cannot use it to construct test statistics; obviously, the natural solution would be to replace again θ_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 θ_0 by $\widehat{\theta}$. With this estimated process we can construct the Kolmogorov-Smirnov and Cramér-von-Mises martingale-transformed statistics

$$\begin{aligned}
\overline{K}_n &\equiv \sup_{z \in \mathbb{R}} \left| \widehat{\mathbf{W}}_n(z) \right|, \\
\overline{C}_n &\equiv \sum_{i=1}^{n} \widehat{\mathbf{W}}_n(\widehat{\varepsilon}_i)^2.
\end{aligned}$$

The asymptotic behaviour of these statistics can be derived studying the weak convergence of $\widehat{\mathbf{W}}_n(\cdot)$. Throughout, for fixed $M_0 > 0$, let $N(\theta_0, M_0) = \{\theta \in \Theta; ||\theta - \theta_0|| \leq M_0 n^{-1/2}\}$ denote the neighborhood of θ_0. Analogously, for fixed $M_1 > 0$, let $N(\theta_*, M_1)$ denote the neighborhood of θ_*. The following additional assumptions, which ensure that the martingale transformation can be performed and behaves properly, are required.

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

Assumption 11: If H_0 holds, then $\sup\limits_{v\in N(\theta_0,M_0)} \int\limits_{-\infty}^{+\infty} ||q_\theta(u)_{|\theta=v}||^2 f(u,\theta_0)\mathrm{d}u = O_p(1)$, where $q_\theta(\cdot)_{|\theta=v}$ denotes the derivative of $q(\cdot)$ with respect to θ evaluated at $\theta = v$.

Assumption 12: If H_1 holds, then $\sup\limits_{v\in N(\theta_*,M_1)} \int\limits_{-\infty}^{+\infty} ||q_\theta(u)_{|\theta=v}||^2 f(u,\theta_*)\mathrm{d}u = O_p(1)$.

Theorem 2: *Suppose that assumptions 1-7 and 10 hold. Then, for every $\epsilon \in (0,1)$, in the space $D[0,1-\epsilon]$,*

a) If H_0 holds and assumptions 8 and 11 are satisfied then:

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

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

b) If H_1 holds and assumptions 9 and 12 are satisfied then, $\forall\ c \in \mathbb{R}$:

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

As Bai (2003) points out in page 540, $\overline{K}_n = (1-\epsilon)^{-1/2}\overline{K}_{n\epsilon} \xrightarrow{d} \sup_{t\in[0,1]} |\mathbf{W}(t)|$, because $(1-\epsilon)^{-1/2}\sup_{t\in[0,1-\epsilon]} |\mathbf{W}(t)|$ and $\sup_{t\in[0,1]} |\mathbf{W}(t)|$ have the same distribution. Therefore, the same critical values can be used for $\overline{K}_{n\epsilon}$ and $\overline{K}_n$ after appropriate rescaling. It follows from this theorem that a consistent asymptotically valid testing procedure with significance level α is to reject H_0 if $\overline{K}_n > k_\alpha$, or to reject H_0 if $\overline{C}_n > c_\alpha$, where k_α 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 for $\overline{K}_n$ with the most usual significance levels are $k_{0.10} = 1.96$, $k_{0.05} = 2.24$, $k_{0.01} = 2.81$ (see e.g. Shorack and Wellner, 1986, p.34), and the critical values for $\overline{C}_n$ with the most usual significance levels are $c_{0.10} = 1.196$, $c_{0.05} = 1.656$, $c_{0.01} = 2.787$ (see e.g. Rothman and Woodroofe, 1972).

The statistics $\overline{K}_n$ and $\overline{C}_n$ 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 $Y - m(X)$ belongs to a parametrically specified family of zero-mean continuous c.d.f.'s, then the statistics that we would use are defined in the same way as $\overline{K}_n$ and $\overline{C}_n$, but considering $q(u) \equiv (1,\ f_u(u,\theta_0)/f(u,\theta_0),\ f_\theta(u,\theta_0)'/f(u,\theta_0))'$. 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 statistics that we would use are $\sup_{z\in\mathbb{R}} |\mathbf{W}_n(z)|$ and $\sum_{i=1}^n \mathbf{W}_n(\widehat{\varepsilon}_i)^2$, where $\mathbf{W}_n(\cdot)$ is defined as above but now considering $q(u) \equiv (1,\ f_{0,u}(u)/f_0(u),\ 1 + uf_{0,u}(u)/f_0(u))'$, where $f_0(\cdot)$ and $f_{0,u}(\cdot)$ denote the first and second derivative of $F_0(\cdot)$. 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 statistics that we would use are again $\sup_{z\in\mathbb{R}} |\mathbf{W}_n(z)|$ and $\sum_{i=1}^n \mathbf{W}_n(\widehat{\varepsilon}_i)^2$, but now $\mathbf{W}_n(\cdot)$ is defined as above 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 a set of Monte Carlo experiments. In each experiment, 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 ε_i are independent, and ε_i has a standardized Student's t distribution with $1/\delta$ degrees of freedom. The value of δ varies from one experiment to another; specifically, we consider $\delta = 0,\ 1/12,\ 1/9,\ 1/7,\ 1/5$ and $1/3$ (when $\delta = 0$, the distribution of ε_i is generated from a standard normal distribution). Using the generated data set $\{(X_i, Y_i)\}_{i=1}^n$ as observations, we test the null hypothesis that the distribution of the error term $\{Y - m(X)\}/\sigma(X)$ is standard normal. Observe that, according to the data generation mechanism, the null hypothesis is true if and only if $\delta = 0$; thus the experiment with $\delta = 0$ allows us to examine the empirical size of the test, and the experiments with $\delta > 0$ allow us to examine the ability of the testing procedure to detect deviations from the null hypothesis caused by thick tails.

The test is performed using the statistics described at the end of the previous section, i.e. the Kolmogorov-Smirnov type statistic $\sup_{z\in\mathbb{R}} |\mathbf{W}_n(z)|$ and the Cramér-von Mises type statistic $\sum_{i=1}^{n} \mathbf{W}_n(\widehat{\varepsilon}_i)^2$, where $\mathbf{W}_n(\cdot)$ is defined as above. Note that in the specific test that we are considering in this set of experiments, the function $q(\cdot)$ that appears in the definition of $\mathbf{W}_n(\cdot)$ proves to be $q(u) \equiv (1, -u, 1 - u^2)'$. 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 the selection of the smoothing value 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. We only discuss the results for the Cramér-von Mises type statistic, since the results that are obtained with the Kolmogorov-Smirnov type statistic are quite similar. In Table 1, 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. The results that we obtain show that the statistic works reasonably well for these sample sizes, and its performance is not very sensitive to the choice of the smoothing value.

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 symmetry tests. Under a nonlinear regression model, conditional symmetry is equivalent to the symmetry of the error term about zero. This is the null hypothesis we are interested in. Symmetry and conditional symmetry play an important role in many situations. The following examples may illustrate the relevance of constructing consistent tests of symmetry and conditional symmetry. Conditional symmetry is part of the stochastic restrictions on unobservable errors used in semiparametric modelling (Powell, 1994). Adaptive estimation relies on the assumption of conditional symmetry (Bickel, 1982; Newey, 1988). In macroeconomics, the symmetry of innovations also plays an important role (Campbell and Henstchel, 1992). In Finance, knowing whether returns or risks exhibit symmetry may help in the choice of an adequate risk measure for portfolio risk management (Gouriérox, Laurent and Scaillet, 2000). Knowledge of the properties of the error term in a regression model has efficiency implications for bootstrapping (Davidson and Flachaire, 2001).

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 proposed a consistent estimator of the distribution of the error term based on nonparametric regression residuals for iid observations, to a context with dependent observations. This would allow us to apply a martingale transform to the nonparametric-oscillation like results derived.

APPENDIX: PROOFS

Proof of Proposition 1: Assume that H_0 holds and let $\widehat{\theta}$ be an appropriate estimator of θ_0. If we add and substract $F(z, \theta_0)$ to $\widehat{\mathbf{V}}_n(\cdot)$, we obtain

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

By Taylor expansion, the second term admits the approximation

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

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 θ_0. Apply assumption 8 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 $\widehat{\varepsilon}_i$:

$$\begin{aligned} \widehat{F}_n(z) &= n^{-1} \sum_{i=1}^{n} I(\widehat{\varepsilon}_i \leq z) \\ &= n^{-1} \sum_{i=1}^{n} I(\varepsilon_i \leq z) + n^{-1} \sum_{i=1}^{n} \varphi(X_i, Y_i, z) + \beta_n(z) + R_n(z), \quad (3) \end{aligned}$$

where $\varphi(x,y,z) = -f(z,\theta_0)\sigma^{-1}(x)\int[I(y \le v) - F(v|x,\theta_0)](1 + z\frac{v-m(x)}{\sigma(x)})\mathrm{d}v$, $\beta_n(z) = \frac{1}{2}h_n^2\{\int u^2K(u)\mathrm{d}u\}E\{\varphi_{xx}(X,Y,z)\}$, $\varphi_{xx}(x,y,z) = \frac{\partial^2}{\partial x^2}\varphi(x,y,z)$ 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 $\varphi_{1n}(\cdot,\cdot)$, $\varphi_{2n}(\cdot,\cdot)$, β_{1n} and β_{2n} are as defined above. The proposition follows immediately by appealing to (2) and (3) in (1). ■

Proof of Theorem 1: First we prove the theorem for $\widehat{K}_n$. Note that, under H_0, $\widehat{K}_n = \sup_{z\in\mathbb{R}}\left|\widehat{\mathbf{D}}_n(z)\right| + o_p(1)$, where we define

$$\widehat{\mathbf{D}}_n(z) \equiv n^{-1/2}\sum_{i=1}^{n}\{I(\widehat{\varepsilon}_i \le z) - F(z,\widehat{\theta}) - \beta_n(z)\}, \tag{4}$$

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 (4), it follows that $\widehat{\mathbf{D}}_n(\cdot)$ has the same asymptotic behaviour as $\mathbf{D}_n(z) \equiv n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \le z) - F(z,\theta_0) + \varphi(X_i,Y_i,z)] - F_\theta(z,\theta_0)'n^{1/2}(\widehat{\theta} - \theta_0)$, 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 (2003), though now an additional term turns up due to the estimation of parameter θ_0. We can rewrite $\varphi(\cdot,\cdot,\cdot)$ as follows:

$$\varphi(x,y,z) = -\frac{f(z,\theta_0)}{\sigma(x)}(1 - \frac{zm(x)}{\sigma(x)})\{\int_y^\infty(1 - F(v|x))\mathrm{d}v - \int_{-\infty}^{y} F(v|x)\mathrm{d}v\}$$

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

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

$$\varphi(x,y,z) = \varphi(x, m(x)+\sigma(x)\varepsilon, z) = f(z,\theta_0)(\varepsilon + \frac{z}{2}(\varepsilon^2-1)). \tag{5}$$

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 8 and replacing (5) 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

$$\delta_z(\varepsilon) = I(\varepsilon \leq z) + f(z,\theta_0)(\varepsilon + \frac{z}{2}(\varepsilon^2 - 1)) - F_\theta(z,\theta_0)'\psi(X,\varepsilon,\theta_0). \tag{6}$$

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-$\check{C}$ervonenkis 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

[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.

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 ε,

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

For $z \in \mathbb{R}$, we may rewrite (6) 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 8, for fixed $X \in \mathbb{R}$ and $\theta_0 \in \Theta$, 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 8 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, $F(z, \widehat{\theta}) = F(z, \widetilde{\theta}) + F_\theta(z, \theta^{**})(\widehat{\theta} - \widetilde{\theta})$ for some $\widetilde{\theta} \in \Theta$ and θ^{**} a mean value between $\widehat{\theta}$ and $\widetilde{\theta}$. Clearly, under H_0, $\widetilde{\theta} = \theta_0$, and the last term is $O_p(n^{-1/2})$ from assumption 8. Analogously, under H_1, $\widetilde{\theta} = \theta_*$, and the last term is $O_p(n^{-1/2})$ from assumption 9. Thus, irrespective of whether H_0 holds true or not, $\sup_{z \in \mathbb{R}} |F(z, \widehat{\theta}) - F(z, \widetilde{\theta})| = o_p(1)$. Therefore $\sup_{z \in \mathbb{R}} |\widehat{F}_n(z) - F(z, \widehat{\theta})| \xrightarrow{p} \sup_{z \in \mathbb{R}} |F_\varepsilon(z) - F(z, \widetilde{\theta})|$. Under H_1, $\sup_{z \in \mathbb{R}} |F_\varepsilon(z) - F(z, \theta_*)| > 0$ 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 $\widehat{F}_n(\cdot)$. ■

Let us define $\widehat{q}(\cdot)$ in the same way as $q(\cdot)$ but replacing θ_0 by $\widehat{\theta}$. The following two propositions are required in the proof of Theorem 2.

Proposition A1: *Suppose that assumptions 1-7 hold. Then:*

a) If H_0 *holds and assumptions 8 and 11 are satisfied then:*

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

b) If H_1 *holds and assumption 9 and 12 are satisfied then:*

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

Proof of Proposition A1: Under assumption 7, $\widehat{q}(\cdot)$ is continuously differentiable with respect to θ. Thus, by a Taylor expansion we obtain

$$\widehat{q}(\cdot) = q(\cdot) + q_\theta(u)_{|\theta=\theta^*}(\widehat{\theta} - \theta_0)/2,$$

where $q_\theta(\cdot, \theta^*)$ denotes the derivative of $q(\cdot)$ with respect to θ, evaluated at θ^*, and θ^* lies between $\widehat{\theta}$ and θ_0. Observe that

$$\begin{aligned}
&\int_{-\infty}^{+\infty} ||\widehat{q}(u) - q(u)||^2 f(u, \theta_0)du \\
\leq\ & \frac{1}{4}||\widehat{\theta} - \theta_0||^2 \int_{-\infty}^{+\infty} ||q_\theta(u)_{|\theta=\theta^*}||^2 f(u, \theta_0)du \\
\leq\ & \frac{1}{4}||\widehat{\theta} - \theta_0||^2 \sup_{v \in N(\theta_0, M_0)} \int_{-\infty}^{+\infty} ||q_\theta(u)_{|\theta=v}||^2 f(u, \theta_0)du \\
=\ & \frac{1}{4} O_p(n^{-1}) O_p(1) = o_p(1),
\end{aligned}$$

where the first inequality follows using $||\widehat{q}(\cdot) - q(\cdot)||^2 \leq ||q_\theta(\cdot)_{|\theta=\theta^*}||^2||\widehat{\theta} - \theta_0||^2/4$, and the last equality follows using assumptions 8 and 11. More precisely, under assumption 8, it is straightforward to show that $(\widehat{\theta} - \theta_0) = O_p(n^{-1/2})$. Then, $||\widehat{\theta} - \theta_0||^2 = O_p(n^{-1})$. This completes the proof of part **a**. The result of part **b** is obtained along the same line of argument using assumptions 9 and 12. ■

Proposition A2: *Suppose that assumptions 1-7 hold. Then:*

a) If H_0 *holds and assumption 8 is satisfied then:*

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

b) If H_1 *holds and assumption 9 is satisfied then:*

$$\begin{aligned} \sup_{z\in\mathbb{R}}||n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i &\geq z)\{\widehat{q}(\varepsilon_i) - q(\varepsilon_i)\} \\ -\int_z^{+\infty}\{\widehat{q}(u) - q(u)\}f(u,\theta_*)du]|| &= o_p(1). \end{aligned}$$

Proof of Proposition A2: As above, under assumption 7, $\widehat{q}(\cdot)$ is continuously differentiable with respect to θ. Thus, by a Taylor expansion we obtain $\widehat{q}(\cdot) = q(\cdot) + q_\theta(\cdot)_{|\theta=\theta^*}(\widehat{\theta} - \theta_0)/2$, where θ^* lies between $\widehat{\theta}$ and θ_0. Thus, under H_0, observe that

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i &\geq z)\{\widehat{q}(\varepsilon_i)-q(\varepsilon_i)\} - \int_{z}^{+\infty}\{\widehat{q}(u)-q(u)\}f(u,\theta_0)du] \\ &= n^{-1/2}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)\{\widehat{q}(\varepsilon_i)-q(\varepsilon_i)\} - E(I(\varepsilon \geq z)\{\widehat{q}(\varepsilon)-q(\varepsilon)\})] \\ &= n^{-1}\sum_{i=1}^{n}[I(\varepsilon_i \geq z)q_\theta(\varepsilon_i)_{|\theta=\theta^*} - E(I(\varepsilon \geq z)q_\theta(\varepsilon)_{|\theta=\theta^*}]n^{1/2}(\widehat{\theta}-\theta_0)/2 \end{aligned}$$

Under assumption 8, it is straightforward to show that $n^{1/2}(\widehat{\theta}-\theta_0) = O_p(1)$. On the other hand, the first term on the right hand side is $o_p(1)$ using some uniform strong law of large numbers. This completes the proof of part **a**. The result of part **b** is obtained along the same line using assumption 9. ■

Proof of Theorem 2: In the following reasoning we assume that the null hypothesis holds. Interchanging the variables, setting $t = F(z,\theta_0)$, 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[2]. Furthermore, define the linear mapping $\Gamma : D[0,1] \rightarrow D[0,1]$ as follows

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

Let

$$\begin{aligned} Q(t) &= (Q_1(t),Q_2(t),Q_3(t),Q_4(t))' \\ &= (t, f(F^{-1}(t,\theta_0)), f(F^{-1}(t,\theta_0))F^{-1}(t,\theta_0), F_\theta(F^{-1}(t,\theta_0))')', \end{aligned}$$

so that $q(F^{-1}(\cdot,\theta_0))$ is the derivative of $Q(\cdot)$. It is easy to check that

$$\Gamma(Q_l(\cdot)) = Q_l(\cdot), \quad \text{for} \quad l = 1,2,3,4. \tag{7}$$

[2] See Section 14 of Billingsley (1968).

From $C(F^{-1}(s,\theta_0))^{-1}C(F^{-1}(s,\theta_0)) = \mathbf{I}_4$ we have $C(F^{-1}(s,\theta_0))^{-1} \times \{\int_s^1 \dot{Q}(r)\mathrm{d}Q_1(r)\} = (1,0,0,0)'$. Thus $\Gamma(Q_1(\cdot))(t) = \int_0^t \dot{Q}(s)'(1,0,0,0)'\mathrm{d}s = Q_1(t)$. 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

$$\widehat{\mathbf{V}}_n(\cdot) = n^{1/2}[\widehat{F}_n(F^{-1}(\cdot,\theta_0)) - Q_1(\cdot)] + n^{1/2}[Q_1(\cdot) - F(F^{-1}(Q_1(\cdot),\theta_0),\widehat{\theta})]. \quad (8)$$

Using the linearity of $\Gamma(\cdot)$, (4) and (5), 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 (4), 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 $\widetilde{\overline{\mathbf{W}}}_n(\cdot) \equiv \widehat{\mathbf{W}}_n(F^{-1}(\cdot,\theta_0))$. Under assumptions 7 and 8, $f(\cdot,\widehat{\theta}) = f(\cdot,\theta_0) + o_p(1)$ (this follows applying a Taylor expansion). Additionally, 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), that completes the proof of **a**.

On the other hand, under H_1, the assertion can be deduced from the probability limit of $n^{-1/2}\widehat{\mathbf{W}}_n(z)$, which is

$$\Xi(z) \equiv F(z) - \int_{-\infty}^{z} \widetilde{q}(u)\widetilde{C}(u)^{-1}\widetilde{d}_n(u) f(u,\theta_*)\mathrm{d}u\},$$

where

$\widetilde{q}(u) \equiv (1, f_u(u,\theta_*)/f(u,\theta_*), 1 + u f_u(u,\theta_*)/f(u,\theta_*), f_\theta(u,\theta_*)'/f(u,\theta_*))'$,
$\widetilde{C}(u) \equiv \int_u^{+\infty} q(\tau)q(\tau)' f(\tau,\theta_*)\mathrm{d}\tau$,
$\widetilde{d}_n(u) \equiv \int_u^{+\infty} q(\tau) f(\tau)\mathrm{d}\tau$,

It can be easily checked that $\Xi(z) \neq 0$ under H_1. The result of 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. The Review of Economics and Statistics 85, 531-549.

Bickel, P.J., 1982. The 1980 Wald Memorial Lecures: On Adaptive Estimation. The Annals of Statistics 10, 647-671.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley, New York.

Campbell, J.Y. and Hentschel, L., 1992. No news is good news. Journal of Financial Economics 31, 281-318.

Davidson, R. and Flachaire, E., 2001. The Wild Bootstrap, Tamed at last. Queen's Institute for Economic Research Working Paper No. 1000.

Dette, H. and Neumeyer, N., 2003. Testing for symmetric error distribution in nonparametric regression models. Mimeo, Ruhr-universitt Bochum.

Durbin, J., 1973. Weak Convergence of the Sample Distribution Function when Parameters are Estimated. The Annals of Statistics 1, 279-290.

Gouriérox, C., Laurent, J.P. and Scaillet, O., 2000. Sensitivity analysis of value at risk. Journal of Empirical Finance 7, 225-245.

Khmaladze, E.V., 1981. Martingale apporach 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.

Khmaladze, E.V., and Koul, H.L., 2004. Martingale transforms of goodness-of-fit tests in regression models. Annals of Statistics 32, 995-1034.

Loynes, R.M., 1980. The Empirical Distribution Function of Residuals from Generalised Regression. The Annals of Statistics 8, 285-298.

Newey, W.K., 1988. Adaptive Estimation of Regression Models via Moment Restrictions. Journal of Econometrics 38, 301-339.

Powel, J.L., 1994. Estimation of semiparametric models. Handbook of Econometrics IV, Elsevier, Amsterdam, chapter 41, 2443-2521.

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.

Van der Vaart, A.W. and Wellner, J.A., 2000. Weak Convergence and Empirical Processes. Springer, New York.

TABLE 1: Proportion of Rejections of H_0

δ	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(4)}$	$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	$h^{(4)}$
	$n = 100$				$n = 500$			
				$\alpha = 0.01$				
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
				$\alpha = 0.05$				
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
				$\alpha = 0.10$				
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