<u>INSTITUT DE STATISTIQUE</u> <u>BIOSTATISTIQUE ET</u> <u>SCIENCES ACTUARIELLES</u> <u>(ISBA)</u>

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



DISCUSSION PAPER

2015/01

Asymptotic distribution-free tests for semiparametric regressions

ESCANCIANO, J-C., PARDO-FERNÁNDEZ, J-C. and I. VAN KEILEGOM

ASYMPTOTIC DISTRIBUTION-FREE TESTS FOR SEMIPARAMETRIC REGRESSIONS

Juan Carlos ESCANCIANO Indiana University Juan Carlos PARDO-FERNÁNDEZ Universidade de Vigo

Ingrid VAN KEILEGOM* Université catholique de Louvain

January 9, 2015

Abstract

This article proposes a new general methodology for constructing nonparametric asymptotic distribution-free tests for semiparametric hypotheses in regression models. Tests are based on the difference between the estimated restricted and unrestricted regression errors' distributions. A suitable integral transformation of this difference renders the tests asymptotically distribution-free, with limits that are well-known functionals of a standard normal variable. Hence, the tests are straightforward to implement. The general methodology is illustrated with applications to testing for parametric models, semiparametric constrained mean-variance models and nonparametric significance. Several Monte Carlo studies show that the finite sample performance of the proposed tests is satisfactory in moderate sample sizes.

^{*}Corresponding address: Institute of Statistics, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. Email: ingrid.vankeilegom@uclouvain.be.

1 Introduction

Let Y_t be a response variable and X_t a *d*-dimensional explanatory variable. Assume that the process (X_t, Y_t) , $t = 0, \pm 1, \pm 2, \ldots$, is strictly stationary and ergodic, and that $E[Y_t^2] < \infty$. We assume Y_t is related to X_t through the heteroskedastic regression model

$$Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t,\tag{1}$$

where $\mu(x) = E(Y_t | X_t = x)$ and $\sigma^2(x) = Var(Y_t | X_t = x)$ are the conditional mean and the conditional variance of Y_t given $X_t = x$, respectively, and ε_t is an error term, which is assumed independent of X_t . In this article, we propose a general methodology for testing semiparametric hypotheses about the regression function $\mu(\cdot)$. We are interested in testing

$$H_0: \mu \in \mathcal{M}$$
 versus $H_1: \mu \notin \mathcal{M}$, (2)

where $\mathcal{M} = \left\{ \mu : \mathbb{R}^d \to \mathbb{R} \text{ such that } \mu(x) = g(\theta, \eta(x), x), \ \theta \in \Theta \subset \mathbb{R}^r, \ \eta \in \mathcal{H} \right\}$ is a class of parametric or semiparametric models defined in terms of a known function q, a finitedimensional unknown parameter θ and a possibly infinite-dimensional unknown parameter η . This general formulation covers many testing problems in semiparametric regression. In particular, if $g(\theta, \eta(x), x) \equiv g(\theta, x)$, then the problem reduces to testing for a parametric model for the regression function μ , which is a classical problem in statistics. Special cases of semiparametric null hypotheses include the partially linear model with $g(\theta, \eta(x), x) =$ $\theta' x_1 + \eta(x_2), x = (x_1, x_2), \text{ additive models with } g(\theta, \eta(x), x) = \eta_1(x_1) + \eta_2(x_2), \eta = (\eta_1, \eta_2),$ and single-index models with $g(\theta, \eta(x), x) = \eta(\theta' x)$, among many others. Two examples that we analyze in detail below are constrained mean-variance models and nonparametric significance testing in regression. In the former $\eta(x) = \sigma(x)$, and our general formulation of g allows to test for relationships between the regression function and the variance function; for example, the choice $g(\theta, \sigma(x), x) = \sigma(x)g_1(\theta, x)$ leads to tests for a particular parametric model given by $g_1(\theta, x)$ for the standardized first moment $\mu(x)/\sigma(x)$, which is an interesting problem in the financial literature since the seminal contributions by Merton (1973, 1980). When Y_t is a stock return, $\mu(x)/\sigma(x)$ is called the Sharpe ratio, and its statistical analysis is of practical interest because it measures the stock return per unit of risk. There are no tests available in the literature for general parametric specifications of $\mu(x)/\sigma(x)$, when $\mu(x)$ and $\sigma(x)$ are nonparametric. For other applications of this setting in statistics see McCullagh and Nelder (1989) and Eagleson and Müller (1997), among others. In nonparametric significance testing $g(\theta, \eta(x), x) = \eta(x_1)$, where $\eta(x_1) = E(Y_t \mid X_{1t} = x_1), X_t = (X_{1t}, X_{2t}) \text{ and } x = (x_1, x_2), \text{ that is, in this semiparametric}$

specification the regressor X_{2t} is not significant in the nonparametric regression model for (Y_t, X_t) .

Tests for parametric or semiparametric hypotheses on the regression function have been extensively investigated in the literature, with a special focus on independent and identically distributed (i.i.d) observations; see González-Manteiga and Crujeiras (2013) for a recent comprehensive survey on the topic. The main two methodologies are based on comparisons of unrestricted (i.e. nonparametric) estimators of μ and restricted estimators of μ (see e.g. Härdle and Mammen, 1993), or their corresponding cumulative processes (see, e.g., Stute, 1997, and Delgado and González-Manteiga, 2001). A third methodology is based on the comparison of unrestricted and restricted estimators of the distributions of the standardized errors. This methodology has been used in several regression contexts in the recent literature, and it is widely applicable. For instance, Van Keilegom, González-Manteiga and Sánchez-Sellero (2008) and Dette, Neumeyer and Van Keilegom (2007) used this idea to develop goodness-of-fit tests for the parametric form of the regression function and the variance function, respectively. Dette, Pardo-Fernández and Van Keilegom (2009) developed goodness-of-fit tests for a multiplicative structure between the regression and the scale functions. Their tests are based on Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM) type statistics and they are not asymptotic distributionfree. Bootstrap methods are then used to approximate critical values. In this article, we propose a new methodology based on the cumulative difference of the standardized errors' distributions under H_0 and H_1 . This methodology applies to the general setting described above and it is simple to implement because the test statistics are asymptotic distribution-free, so that resampling methods are not needed. The proposed tests are consistent against fixed alternatives and they are able to detect some, although not all, local alternatives converging to the null at the parametric rate. The tests can be tailored to specific alternatives of interest. Monte Carlo experiments show a satisfactory finite sample performance for the proposed tests in three different applications: parametric models, mean-variance constrained models and nonparametric significance testing.

The remainder of this article is organized as follows. In Section 2 we introduce the new methodology. In Section 3 we provide asymptotic distribution theory for the general framework under the null hypothesis and under fixed and local alternatives. Three applications are studied in detail in Section 4. Section 5 presents simulation studies for the three applications. Finally, we conclude in Section 6 by pointing out further applications and future research. Mathematical proofs are gathered in an Appendix.

2 General methodology

This section introduces the general methodology. The discussion here is organized around a few "high-level" assumptions. More primitive conditions can be given in specific applications of the methodology; see Section 3 below. Our tests are based on the comparison of two estimators of the standardized error's distribution. Define the restricted model for the regression function as $\mu_0(x) = g(\theta_0, \eta_0(x), x)$, where

$$(\theta_0, \eta_0) = \arg\min_{\theta \in \Theta, \eta \in \mathcal{H}} E[(\mu(X_t) - g(\theta, \eta(X_t), X_t))^2].$$

Henceforth, we assume that (θ_0, η_0) exists and is unique (identifiability). Then, define the standardized errors

$$\varepsilon_{t0} = \frac{Y_t - \mu_0(X_t)}{\sigma(X_t)}$$
 and $\varepsilon_t = \frac{Y_t - \mu(X_t)}{\sigma(X_t)}$

with cumulative distribution functions $F_{\varepsilon 0}(y) = P(\varepsilon_{t0} \leq y)$ and $F_{\varepsilon}(y) = P(\varepsilon_t \leq y)$, respectively. Our testing procedure will be based on the integrated difference of the distribution functions of ε_{t0} and ε_t . The following theorem justifies the testing procedure described below. The proof is given in the Appendix.

Theorem 1 The following statements are equivalent:

- (i) H_0 is true.
- (ii) ε_{t0} and ε_t have the same distribution.
- (iii) $D(y) = \int_{-\infty}^{y} (F_{\varepsilon 0}(s) F_{\varepsilon}(s)) ds = 0$ for all $y \in \mathbb{R}$.

The equivalence between statements (i) and (ii) has been extensively used in recent literature to construct tests for H_0 (see, for example, Van Keilegom *et al.*, 2007, or Dette *et al.*, 2009). In this article we will exploit the equivalence between (i) and (iii), which will allow us to obtain asymptotic distribution-free tests.

In practice, the variables ε_t and ε_{t0} are not observable, so they need to be estimated. Assume that a sample (X_t, Y_t) , $t = 1, \ldots, T$ is available and construct the estimated residuals

$$\widehat{\varepsilon}_{t0} = \frac{Y_t - \widehat{\mu}_0(X_t)}{\widehat{\sigma}(X_t)} \quad \text{and} \quad \widehat{\varepsilon}_t = \frac{Y_t - \widehat{\mu}(X_t)}{\widehat{\sigma}(X_t)}, \quad (3)$$

for t = 1, ..., T, where $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ are nonparametric estimators of $\mu(x)$ and $\sigma(x)$, respectively, and $\hat{\mu}_0(x)$ is a suitable consistent estimator of $\mu_0(x)$. More precisely, to estimate nonparametrically $\mu(x)$ we use local polynomial estimators, that is, $\hat{\mu}(x) = \hat{\alpha}_0(x)$, where $\hat{\alpha}_0(x)$ is the first component of the vector $\hat{\alpha}(x)$, which is the solution of the local minimization problem

$$\min_{\alpha} \sum_{t=1}^{T} \left\{ Y_t - P_t(\alpha, x, p) \right\}^2 K_h(X_t - x),$$
(4)

where $P_t(\alpha, x, p)$ is a polynomial of order p built up with all $0 \le i \le p$ products of factors of the form $X_{jt} - x_j$, j = 1, ..., d, and d is the dimension of x. The vector α consists of all coefficients of this polynomial. Here, for $u = (u_1, ..., u_d) \in \mathbb{R}^d$, $K(u) = \prod_{j=1}^d k(u_j)$ is a d-dimensional product kernel, k is a univariate kernel function, $h = (h_1, ..., h_d)$ is a d-dimensional bandwidth vector converging to zero when n tends to infinity, and $K_h(u) = \prod_{j=1}^d k(u_j/h_j)/h_j$. To estimate $\sigma(x)$, define

$$\widehat{\sigma}^2(x) = \widehat{\gamma}_0(x) - \widehat{\alpha}_0^2(x),$$

where $\widehat{\gamma}_0$ is defined in the same way as $\widehat{\alpha}_0$, but with Y_t replaced by Y_t^2 in (4), $t = 1, \ldots, T$.

There are general estimation methods available for semiparametric models that can be used to estimate (θ_0, η_0) . For example, we can use sieve least squares estimators

$$(\widehat{\theta}, \widehat{\eta}) = \arg\min_{\theta \in \Theta, \eta \in \mathcal{H}_T} \frac{1}{T} \sum_{t=1}^T (Y_t - g(\theta, \eta(X_t), X_t))^2,$$

where \mathcal{H}_T is a sieve approximation of \mathcal{H} (see e.g. Chen, 2007, and references therein). In other applications alternative estimators for η_0 , such as kernel estimators, can be used. This is the case for our constrained mean-variance example, where $\hat{\eta} = \hat{\sigma}$, and where θ_0 can be estimated by the (two-step) least squares estimator

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} (Y_t - g(\theta, \widehat{\sigma}(X_t), X_t))^2.$$

Rather than focussing on a specific estimator or class of estimators, here we assume that an estimator for (θ_0, η_0) is available satisfying certain conditions below, and we refer to the detailed examples below for specific choices of estimators. The restricted estimator of $\mu_0(x)$ is then denoted by $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\eta}(x), x)$.

The corresponding distribution functions, $F_{\varepsilon 0}(y)$ and $F_{\varepsilon}(y)$, are estimated by

$$\widehat{F}_{\varepsilon 0}(y) = \frac{1}{T\bar{w}} \sum_{t=1}^{T} w(X_t) I(\widehat{\varepsilon}_{t0} \le y) \quad \text{and} \quad \widehat{F}_{\varepsilon}(y) = \frac{1}{T\bar{w}} \sum_{t=1}^{T} w(X_t) I(\widehat{\varepsilon}_t \le y),$$

respectively, where w is a positive weight function and $\bar{w} = T^{-1} \sum_{t=1}^{T} w(X_t)$. The weights are introduced as a technical device to allow for covariates with non-compact support. Note that when $w \equiv 1$, then the regular empirical distribution functions based on estimated residuals are obtained. Given the standardized difference of empirical distributions

$$\widehat{R}(y) = \sqrt{T} \left(\widehat{F}_{\varepsilon 0}(y) - \widehat{F}_{\varepsilon}(y) \right),$$

 $-\infty < y < +\infty$, under suitable regularity conditions (see below) and the null hypothesis H_0 , one can establish an asymptotic expansion for $\widehat{R}(y)$ as follows:

$$\widehat{R}(y) = \frac{f_{\varepsilon}(y)}{E[w(X_t)]} T^{-1/2} \sum_{t=1}^{T} w(X_t) W_t + o_P(1),$$
(5)

uniformly in $-\infty < y < \infty$, where $f_{\varepsilon}(y)$ is the density function corresponding to $F_{\varepsilon}(y)$, and W_t is a zero-mean random variable which will be defined later. We further assume that

$$0 < \sigma_W^2 = \frac{E[w^2(X_t)W_t^2]}{(E[w(X_t)])^2} < \infty.$$

The random variables W_t and the regularity conditions needed for (5) to hold are of course specific to each application. For instance, Dette *et al.* (2009) established this expansion for constrained mean-variance models when $g(\theta, \sigma(x), x) = \theta \sigma(x)$ and x is univariate, and used this expansion to propose Kolmogorov-Smirnov (KS) and Cramérvon Mises (CM) type statistics. Tests based on the expansion of $\hat{R}(y)$ have null limiting distributions which are unknown, as they depend on the density of the errors and other unknown quantities. These authors suggested to implement the tests with the assistance of a bootstrap procedure. To avoid resampling methods, while keeping the good power properties of the procedure, we consider a transformation of $\hat{R}(y)$ that is well-suited for delivering asymptotic distribution-free inference. Specifically, we consider the integrated process

$$\widehat{C}(y) = \int_{-\infty}^{y} \widehat{R}(s) ds,$$

 $-\infty < y < \infty$, and propose test statistics that are continuous functionals of $\widehat{C}(\cdot)$. In particular, we consider the KS and CM-type statistics

$$KS_T = \sup_{-\infty < y < \infty} \frac{1}{\widehat{\sigma}_W} \left| \widehat{C}(y) \right|,$$

and

$$CM_T = \frac{3}{\widehat{\sigma}_W^2} \int_{-\infty}^{\infty} \left(\widehat{C}(y)\right)^2 d\widehat{F}_{\varepsilon 0}(y),$$

where $\hat{\sigma}_W^2$ is a consistent estimator of σ_W^2 . The proposed test statistics are straightforward to compute, since some simple algebra shows that

$$KS_T = \max_{1 \le t \le T} \frac{1}{\widehat{\sigma}_W} \max\left\{ \left| \widehat{C}(\widehat{\varepsilon}_t) \right|, \left| \widehat{C}(\widehat{\varepsilon}_{t0}) \right| \right\} \quad \text{and} \quad CM_T = \frac{3}{\widehat{\sigma}_W^2} \frac{1}{T\overline{w}} \sum_{t=1}^T w(X_t) \left(\widehat{C}(\widehat{\varepsilon}_{t0}) \right)^2,$$

where

$$\widehat{C}(y) = \frac{1}{\sqrt{T}\overline{w}} \left(\sum_{s=1}^{T} w(X_s) \{ (y - \widehat{\varepsilon}_{s0})_+ - (y - \widehat{\varepsilon}_s)_+ \} \right),$$

and $a_{+} = \max\{a, 0\}$. In the asymptotic results given in Section 3, we will prove that, under H_0 ,

$$KS_T \to_d |Z|$$
 and $CM_T \to_d Z^2$,

where $Z \sim N(0, 1)$. Therefore the tests are very easy to implement. For instance, the CM test rejects H_0 at significance level α level if $CM_T > \chi^2_{1,1-\alpha}$, where $\chi^2_{1,1-\alpha}$ is the $(1 - \alpha)$ -quantile of the chi-square distribution with one degree of freedom.

3 Asymptotic results

This section investigates the limit distribution of the test statistics proposed in the previous section under a set of primitive conditions. To that end, we introduce the following regularity conditions and notations. Let $F_X(x) = P(X_t \leq x)$ and let $F(x, y) = P(X_t \leq x, Y_t \leq y)$ (which under assumption A1 below do not depend on t). Lowercase letters will be used to denote the corresponding density functions. Define the β -mixing coefficients as (see e.g. Doukhan, 1994)

$$\beta_t = \sup_{m \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{t+m}^{\infty}} E \left| P(A | \mathcal{F}_{-\infty}^m) - P(A) \right|,$$

where \mathcal{F}_s^t denotes the σ -algebra generated by the sequence $\{(X_j, Y_j), j = s, \ldots, t\}$ for $s \leq t$. Henceforth, C is a generic constant that may change from expression to expression.

ASSUMPTION A1: The process (X_t, Y_t) , $t = 0, \pm 1, \pm 2, \ldots$, satisfies (1) and is strictly stationary and absolutely regular (β -mixing), with mixing coefficients satisfying $\beta_t = O(t^{-b})$, for some b > 2. Assumption A2:

- (i) θ_0 belongs to the interior of a compact subset Θ of \mathbb{R}^r .
- (ii) The weight function w has a compact support R_w in \mathbb{R}^d and satisfies w(x) > 0 for all $x \in R_w$ and $\sup_{x \in R_w} w(x) \le C$.
- (iii) All partial derivatives of F_X up to order 2d + 1 exist on the interior of R_w , they are uniformly continuous and $\inf_{x \in R_w} f_X(x) > 0$.
- (iv) All partial derivatives of μ and σ up to order p + 2 exist on the interior of R_w , they are uniformly continuous and $\inf_{x \in R_w} \sigma(x) > 0$.

Whenever there is no ambiguity, we use the same notation for any function of X as for its version restricted to the compact set R_w .

Assumption A3:

- (i) $E(|Y_0|^s) < \infty$ and $\sup_{x \in R_X} E(|Y_0|^s \mid X_0 = x) < \infty$ for some s > 2 + 2/(b-2), where b is as in Assumption A1.
- (ii) There exists some j' such that for all $j \ge j'$,

$$\sup_{x_0, x_j \in R_X} E(|Y_0 Y_j|^2 \mid X_0 = x_0, X_j = x_j) f_j(x_0, x_j) < \infty,$$

where $f_j(x_0, x_j)$ denotes the joint density of (X_0, X_j) .

(iii) The errors of the regression model satisfy

$$E(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}) = E(\varepsilon_t | X_t)$$

and

$$\operatorname{Var}(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}) = E(\varepsilon_t^2 | X_t).$$

Furthermore, ε_t is independent of X_t , with mean zero and unit variance.

ASSUMPTION A4: The function F(x, y) is continuous in (x, y), and twice continuously differentiable with respect to x and y. Let L(x, y) denote generically the derivatives $\frac{\partial}{\partial x}F(x, y), \frac{\partial}{\partial y}F(x, y), \frac{\partial^2}{\partial x^2}F(x, y), \frac{\partial^2}{\partial y^2}F(x, y)$ and $\frac{\partial^2}{\partial x \partial y}F(x, y)$. Then, L(x, y) is continuous in (x, y) and satisfies $\sup_{x,y} |y^2 L(x, y)| < \infty$. Assumption A5:

- (i) For all $j = 1, ..., d : h_j/h_1 \to C_j$, with $0 < C_j < \infty$, and the bandwidth h_1 satisfies $(\log T)^{-1}T^{\eta}h_1^d \to \infty$ for $\eta = \frac{b-1-d-(1+b)/(s-1)}{b+3-d-(1+b)/(s-1)}$, where d, b and s are such that $d < \frac{(b-2)(s-2-2/(b-2))+s-2}{s-1}$, and with b and s as defined in Assumption A1 and A3 respectively, $Th_1^{2d+\delta} \to \infty$ for some small $\delta > 0$, $Th_1^{2p+2} \to 0$ for odd p and $Th_1^{2p+4} \to 0$ for even p.
- (ii) The kernel k is a symmetric probability density function on [-1, 1], k is d times continuously differentiable, and $k^{(j)}(\pm 1) = 0$ for $j = 0, \ldots, d 1$.

Assumption A3 and the first condition in Assumption A5-(i) are taken from Hansen (2008), and they ensure suitable rates of convergence of the kernel estimators of $\mu(\cdot)$ and $\sigma(\cdot)$. Our next assumption is on the restricted estimator $\hat{\mu}_0$. The class of smooth functions $C_M^{\alpha}(R_w)$ is defined in the Appendix (see also p. 154 in Van der Vaart and Wellner, 1996).

ASSUMPTION A6: Under H_0 , the restricted estimator $\widehat{\mu}_0(x) = g(\widehat{\theta}, \widehat{\eta}(x), x)$ satisfies:

(i) $\sup_{x \in R_w} |\widehat{\mu}_0(x) - \mu_0(x)| = o_P(T^{-1/4}).$ (ii) $P\left(\widehat{\mu} - \widehat{\mu}_0 \in C^{\alpha}_M(R_w)\right) \to 1$ as T tends to infinity. (iii)

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu_0(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) + o_P(T^{-1/2}),$$

for a measurable function $l(\cdot)$ satisfying $E(l(Y_t, X_t) \mid X_t, \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(\|l(Y_t, X_t)\|^2) < \infty$ (where $\|\cdot\|$ is the Euclidean norm).

Assumption A6 is standard in the literature. If $\hat{\eta}$ is a kernel estimator, the first condition in Assumption A6 can be shown to hold under smoothness conditions on g, using results from Hansen (2008). Likewise, results in Neumeyer and Van Keilegom (2010) can be used to show A6-(ii). Assumption A6-(iii) requires a linear asymptotic representation for the restricted estimator. For many examples of semiparametric models and estimators this representation has been already established. We verify this condition for our leading examples below.

3.1 Asymptotic null distribution

Our next result establishes a uniform expansion of the process $\widehat{C}(y)$ in i.i.d. terms.

Theorem 2 Assume A1-A6. Then, under the null hypothesis H_0 the following holds:

$$\widehat{C}(y) = T^{1/2} \int_{-\infty}^{y} [\widehat{F}_{\varepsilon 0}(s) - \widehat{F}_{\varepsilon}(s)] \, ds = \frac{F_{\varepsilon}(y)}{E[w(X_t)]} T^{-1/2} \sum_{t=1}^{T} w(X_t) W_t + o_P(1) \cdot \frac{1}{2} \sum_{t=1}^{T} w($$

uniformly in $-\infty < y < \infty$, where $W_t = l(Y_t, X_t) - \varepsilon_t$.

The proof of Theorem 2 is given in the Appendix. The result does not follow directly from (5), since the mapping $R \to \int_{-\infty}^{y} R(s) ds$ is not continuous in the space of uniformly bounded functions. The rate of convergence of \widehat{C} is $T^{-1/2}$ and does not depend on the dimension of X_t , in contrast to alternative nonparametric tests based on smoothers.

We now assume that a consistent estimator for $\sigma_W^2 = E(w^2(X_t)W_t^2)/(E[w(X_t)])^2$ exists.

ASSUMPTION A7: $0 < \sigma_W^2 < \infty$, and there exists a weakly consistent estimator for σ_W^2 , say $\hat{\sigma}_W^2$.

Remark 1. Estimation of variance. A natural candidate to estimate σ_W^2 is

$$\widehat{\sigma}_W^2 = \frac{1}{\overline{w}^2 T} \sum_{t=1}^T w^2(X_t) \widehat{W}_t^2,\tag{6}$$

where $\widehat{W}_t = \widehat{l}(Y_t, X_t) - \widehat{\varepsilon}_t$ is a suitable consistent estimator of W_t , see the examples below.

The following Corollary is a consequence of Theorem 2, the continuous mapping theorem, the central limit theorem for mixing sequences (see, for instance, Theorem 2.20 in Fan and Yao, 2003) and the consistency of the estimator $\hat{\sigma}_W^2$.

Corollary 3 Assume A1-A7. Then, under the null hypothesis H_0 ,

 $KS_T \to_d |Z|$ and $CM_T \to_d Z^2$,

where $Z \sim N(0, 1)$.

3.2 Power against fixed and local alternatives

This section investigates the asymptotic power of the proposed tests. We first analyze the power against the fixed alternatives

$$H_1: \mu(\cdot) \neq \mu_0(\cdot)$$
, with positive probability on R_w .

That is, under this alternative hypothesis H_1 , $P(\mu(X_t) \neq \mu_0(X_t) | X_t \in R_w) > 0$. We shall show that under certain conditions the proposed tests will diverge to infinity under H_1 as $T \to \infty$, thereby proving the consistency of the tests against these fixed alternatives. To that end, note that the proof of Theorem 2 shows that under the alternative hypothesis, uniformly in $-\infty < y < \infty$,

$$T^{-1/2}\widehat{C}(y) = \int_{-\infty}^{y} R(s)ds + o_P(1),$$

$$\equiv C(y) + o_P(1),$$

where $R(y) = E(w(X_t) \{ I(\varepsilon_{t0} \le y) - I(\varepsilon_t \le y) \}) / E(w(X_t))$. Therefore, by the continuous mapping theorem,

$$T^{-1}CM_T = \frac{3}{\sigma_W^2} \int_{-\infty}^{\infty} \left(C(y)\right)^2 dF_{\varepsilon}(y) + o_P(1),$$

and

$$T^{-1/2}KS_T = \sup_{-\infty < y < \infty} \frac{1}{\sigma_W} |C(y)| + o_P(1).$$

Thus, the test statistics will diverge to infinity as $T \to \infty$ if $C(y) \neq 0$ for some $y \in \mathbb{R}$. This is evident for the KS test, and using the continuity of C and F_{ε} , the same holds for the CM-type test. The next result shows that under H_1 , $C(y) \neq 0$ for some $y \in \mathbb{R}$. The proof is given in the Appendix.

Proposition 4 Under the alternative hypothesis H_1 , $C(y) \neq 0$ for some $y \in \mathbb{R}$.

We consider now the limiting distribution under the local alternative

$$H_{1T}(a): \frac{\mu(x) - \mu_0(x)}{\sigma(x)} = \frac{a(x)}{\sqrt{T}}$$

so the true $\mu(x)$ equals $\mu(x) = \mu_0(x) + T^{-1/2}\sigma(x)a(x)$, where $a \neq 0$ is the direction of departure, such that $E[w(X_t)a(X_t)] < \infty$ and $0 < E[a^2(X_t)W_t^2] < \infty$. The following assumption is needed to control the behavior of the estimator $\hat{\mu}_0(x)$ under $H_{1T}(a)$:

ASSUMPTION A6BIS: Under $H_{1T}(a)$, the restricted estimator $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\eta}(x), x)$ satisfies: (i) $\sup_{x \in R_w} |\widehat{\mu}_0(x) - \mu(x)| = o_P(T^{-1/4}).$ (ii) $P\left(\widehat{\mu} - \widehat{\mu}_0 \in C^{\alpha}_M(R_w)\right) \to 1$ as T tends to infinity. (iii)

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu_0(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) - T^{-1/2} \int [w(x) + d(x)] a(x) dF_X(x) + o_P(T^{-1/2}),$$

for a measurable function $l(\cdot)$ satisfying $E(l(Y_t, X_t) \mid X_t, \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(\|l(Y_t, X_t)\|^2) < \infty$, and where $d(\cdot)$ depends on the estimator of θ_0 .

Define

$$\delta(a) = \frac{E\left[\left(w(X_t) + d(X_t)\right)a(X_t)\right]}{\left(E\left[w^2(X_t)W_t^2\right]\right)^{1/2}}.$$

Then, we have the following asymptotic result under the local alternative. We refer to the Appendix for the proof.

Corollary 5 Assume A1-A5, A6bis and A7. Then, under the local alternative $H_{1T}(a)$,

 $KS_T \to_d |Z + \delta(a)|$ and $CM_T \to_d (Z + \delta(a))^2$,

where $Z \sim N(0, 1)$.

This result shows that our tests are able to detect local alternatives converging at the parametric rate, provided $\delta(a) \neq 0$. The tests are not consistent against all local alternatives, as there are directions a for which $\delta(a) = 0$. Nevertheless, if one is particularly interested in a direction a^* , one can always choose the weight function w so that $\delta(a^*) \neq 0$, namely by choosing $w(x) = |a^*(x)|$, and a^* such that $E[d(X_t)a^*(X_t)] = 0$. We provide below further details on local power, including the expressions for the drift parameter $\delta(a)$, in the context of some examples. The local power properties of our tests are different from other nonparametric tests based on smoothers (see e.g. Härdle and Mammen, 1993), which are unable to detect local alternatives converging to the null hypothesis at the parametric rate.

4 Examples

This section provides formulae for the estimators, influence functions, drift parameters, asymptotic variances σ_W^2 and their estimators for three examples: (i) parametric models for the regression function μ ; (ii) constrained mean-variance models and (iii) nonparametric significance testing.

4.1 Parametric models

Consider the specification $g(\theta, \eta(x), x) = g(\theta, x)$. The testing problem (2) is then a goodness-of-fit test for a parametric form of the regression function $\mu(x)$. The unknown parameter θ_0 can be estimated by the nonlinear least squares estimator

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \frac{1}{T} \sum_{t=1}^{T} (Y_t - g(\theta, X_t))^2.$$

The following assumption is needed to verify Assumption A6 in this example:

ASSUMPTION E1: The function $g(\cdot, x)$ is continuously differentiable with respect to θ in a neighborhood of θ_0 , say Θ_0 , for all $x \in \mathbb{R}^d$, with derivative $g_{\theta t}(\theta) = \partial g(\theta, X_t)/\partial \theta$ satisfying that $E[g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)]$ is finite and non-singular. Furthermore, for all $\theta \in \Theta_0$, $g(\theta, \cdot) \in C^{\alpha}_M(R_w)$, for an $\alpha > \max(d/2, 1)$.

By Assumption E1 there exists a constant C such that

$$\sup_{x \in R_w} |\widehat{\mu}_0(x) - \mu_0(x)| \le C |\widehat{\theta} - \theta_0| = O_P(T^{-1/2}),$$

which verifies A6-(i). Likewise, E1 and results in Neumeyer and Van Keilegom (2010) show that A6-(ii) holds. To verify A6-(iii) note that by E1 and standard least squares theory,

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu_0(x)) dF_X(x) = E[w(X_t)\sigma^{-1}(X_t)g_{\theta t}(\theta_0)](\widehat{\theta} - \theta_0) + o_P(T^{-1/2})$$
$$= T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) + o_P(T^{-1/2}),$$

with

$$l(Y_t, X_t) = E[w(X_t)\sigma^{-1}(X_t)g_{\theta t}(\theta_0)](E[g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)])^{-1}g_{\theta t}(\theta_0)\sigma(X_t)\varepsilon_t.$$

This verifies Assumption A6-(iii). Similarly, it is straightforward to verify that Assumption A6bis-(iii) holds with

$$d(X_t) = -E[w(X_t)\sigma^{-1}(X_t)g_{\theta t}(\theta_0)](E[g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)])^{-1}g_{\theta t}(\theta_0)\sigma(X_t).$$

This drift term results from the asymptotic non-zero mean of $\sqrt{T}\left(\hat{\theta} - \theta_0\right)$ under local alternatives.

In the particular case of testing for a linear model $g(\theta, x) = \theta' x$, $g_{\theta t}(\theta_0) = X_t$ and then

$$W_t = \varepsilon_t \left\{ E[w(X_t)\sigma^{-1}(X_t)X'_t](E[X_tX'_t])^{-1}X_t\sigma(X_t) - 1 \right\}.$$

A consistent estimator for σ_W^2 is then obtained by replacing $\sigma(X_t)$ and ε_t in the previous expression by $\hat{\sigma}(X_t)$ and $\hat{\varepsilon}_t$, respectively, and then using equation (6).

From a practical point of view, for the goodness-of-fit problem it is also recommendable to apply some smoothing to the restricted estimator of μ as in Van Keilegom *et al.* (2007). That is, in the definition of ε_{t0} in (3), replace $\hat{\mu}_0(X_t)$ by $\tilde{\mu}_0(X_t)$, where $\tilde{\mu}_0(X_t)$ is obtained in the same way as the nonparametric estimator $\hat{\mu}(X_t)$, but replacing the responses Y_t by $\hat{\mu}_0(X_t)$. The asymptotic theory given in section 3 is still valid when this modification is applied to the estimated residuals.

4.2 Constrained mean-variance models

This section illustrates the general methodology with an application to constrained meanvariance models. In these models $\eta(x) = \sigma(x)$ and the null hypothesis becomes

$$H_0: \mu(X_t) = g(\theta_0, \sigma(X_t), X_t), \tag{7}$$

where g is a completely specified function up to the unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^r$. The alternative hypothesis is the negation of the null, i.e.

$$H_1: H_0$$
 is not true.

Dette, Pardo-Fernández and Van Keilegom (2009) studied the special case $g(\theta_0, \sigma(x), x) = \theta_0 \sigma(x)$ in detail and developed bootstrap-based tests for the corresponding hypothesis under the assumption that the covariate is one-dimensional. See also Dette, Marchlewski and Wagener (2012) for an alternative test with i.i.d. data. Our more general formulation here is motivated from applications in economics and finance. We illustrate the general applicability of our null hypothesis (7) with some examples. The first example is motivated by an extensive empirical literature documenting a time-varying standardized first

moment $\mu(X_t)/\sigma(X_t)$; see e.g. Harvey (1989), Ferson (1989), Ferson, Foester and Keim (1993).

Example 1. Models with parametric time-varying coefficient of variation. The general null hypothesis H_0 also incorporates as a special case the hypothesis that $\mu(X_t)/\sigma(X_t)$ follows a specific parametric model. For instance, if the null hypothesis is of a multiplicative form $g(\theta, \sigma(x), x) = \sigma(x)g_1(\theta, x)$, then $\mu(X_t)/\sigma(X_t)$ will have the parametric structure specified in g_1 . Examples are the exponential models in De Santis and Gerard (1997) and Bekaert and Harvey (1995), where $g_1(\theta_0, X_t) = \exp(\theta'_0 X_t)$. Hence, our null hypothesis encompasses tests for the correct specification of a parametric coefficient of variation. To the best of our knowledge, such tests are not available in the literature.

Our next example illustrates that our testing procedure is also useful for testing multiplicative structures in classical time series models using an appropriate definition of Y_t .

Example 2. Testing for multiplicative structure. Several classical time series models for univariate processes present a multiplicative structure of the form $Z_t = \sigma_t \epsilon_t$, where $\sigma_t^2 = m(Z_{t-1}, \ldots, Z_{t-p})$ for some specification of the function m (usually, a linear functional form). The squared values of the process Z_t can be re-expressed as the regression model $Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t$, where $Y_t = Z_t^2$, $X_t = (Z_{t-1}, \ldots, Z_{t-p})$ and $\varepsilon_t = c(\epsilon_t^2 - 1)$, with $c^2 = [E(\epsilon^4) - 1]^{-1}$, and the peculiarity that $\mu(X_t) = m(X_t)$ and $\sigma(X_t) = c^{-1}m(X_t)$. Hence the regression function and the scale function satisfy the relation $\mu(\cdot) = c\sigma(\cdot)$. This is a feature satisfied by many financial models, such as the ARCH(p). Dette, Pardo-Fernández and Van Keilegom (2009) studied this example in detail when the covariate is one-dimensional.

In the general case the parameter θ_0 can be estimated by the following weighted leastsquares (LS) estimator

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \sum_{t=1}^{T} w(X_t) \left(Y_t - g(\theta, \widehat{\sigma}(X_t), X_t) \right)^2.$$
(8)

See also Escanciano, Pardo-Fernández and Van Keilegom (2013) for a related estimator of θ_0 in a more general context. The following assumption is needed to verify Assumption A6 in this general example:

ASSUMPTION E2: The function $g(\theta, u, x)$ is continuously differentiable with respect to the components of θ and u, with derivative $g_{\theta t}(\theta) = \partial g(\theta, X_t) / \partial \theta$ satisfying that $E[g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)]$

is finite and non-singular. Furthermore, for all $\theta \in \Theta_0$, $P(g(\theta, \widehat{\sigma}(\cdot), \cdot) \in C^{\alpha}_M(R_w)) \to 1$, for an $\alpha > \max(d/2, 1)$.

By Assumptions A1-A5 and E2 there exists a constant C such that

$$\sup_{x \in R_w} |\widehat{\mu}_0(x) - \mu_0(x)| \leq C \left\{ |\widehat{\theta} - \theta_0| + \sup_{x \in R_w} |\widehat{\sigma}(x) - \sigma(x)| \right\}$$
$$= O_P(T^{-1/4}),$$

which verifies A6-(i). Likewise, E2 and results in Neumeyer and Van Keilegom (2010) show that A6-(ii) holds.

Consider the notation:

$$g_{\theta t}(\theta) = \frac{\partial}{\partial \theta} g(\theta, \sigma(X_t), X_t),$$

$$g_{ut}(\theta) = \frac{\partial}{\partial u} g(\theta, u, X_t)|_{u=\sigma(X_t)},$$

$$u_t = \varepsilon_t - 0.5g_{ut}(\theta_0)(\varepsilon_t^2 - 1),$$

$$S(\theta_0) = E \left[w(X_t)g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0) \right] / E[w(X_t)],$$

$$s(X_t, \varepsilon_t) = S^{-1}(\theta_0)\sigma(X_t)g_{\theta t}(\theta_0)u_t, \text{ and}$$

$$\varphi(\theta) = \frac{1}{E[w(X_t)]} E \left[\frac{w(X_t)}{\sigma(X_t)}g_{\theta t}(\theta) \right].$$

Then, the following Lemma verifies Assumption A6-(iii).

Lemma 6 Assume A1-A5 and E2. Then, under the null hypothesis H_0 ,

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu_0(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) + o_P(T^{-1/2}),$$

where

$$l(Y_t, X_t) = \varphi'(\theta_0) s(X_t, \varepsilon_t) + 0.5g_{ut}(\theta_0)(\varepsilon_t^2 - 1).$$
(9)

The first term $s(X_t, \varepsilon_t)$ arises in the linear expansion because the least squares estimator $\hat{\theta}$ satisfies

$$\widehat{\theta} - \theta_0 = \frac{1}{E[w(X_t)]T} \sum_{t=1}^T w(X_t) s(X_t, \varepsilon_t) + o_P(T^{-1/2}),$$

as shown in Escanciano, Pardo-Fernández and Van Keilegom (2013). The second term in (9) accounts for the effect of estimating σ in $g(\theta, \sigma(X_t), X_t)$. The arguments of Escanciano, Pardo-Fernández and Van Keilegom (2013) can be also used to show that $\sqrt{T}\left(\hat{\theta} - \theta_0\right)$

has an asymptotic non-zero mean under local alternatives, so that Assumption A6bis-(iii) holds with

$$d(X_t) = -\frac{S^{-1}(\theta_0)w(X_t)g_{\theta t}(\theta_0)\sigma(X_t)}{E[w(X_t)]}.$$

Therefore, for this example A6-A6bis hold and

$$W_t = \varphi'(\theta_0)s(X_t, \varepsilon_t) - u_t$$

which can be consistently estimated by

$$\widehat{W}_t = \widehat{\varphi}'(\widehat{\theta})\widehat{s}(X_t,\widehat{\varepsilon}_t) - \widehat{u}_t,$$

with

$$\widehat{u}_t = \widehat{\varepsilon}_t - 0.5\widehat{g}_{ut}(\widehat{\theta})(\widehat{\varepsilon}_t^2 - 1),$$
$$\widehat{\varphi}(\widehat{\theta}) = \frac{1}{\overline{w}T} \sum_{t=1}^T \frac{w(X_t)}{\widehat{\sigma}(X_t)} \widehat{g}_{\theta t}(\widehat{\theta}),$$

 $\widehat{s}(X_t, \widehat{\varepsilon}_t)$ is a consistent estimate for the influence function $s(X_t, \varepsilon_t)$,

$$\widehat{g}_{ut}(\widehat{\theta}) = \left. \frac{\partial g(\widehat{\theta}, u, X_t)}{\partial u} \right|_{u=\widehat{\sigma}(X_t)} \quad \text{and} \quad \widehat{g}_{\theta t}(\widehat{\theta}) = \frac{\partial g(\widehat{\theta}, \widehat{\sigma}(X_t), X_t)}{\partial \theta}.$$

It is often straightforward to prove the consistency of $\hat{\sigma}_W^2$ in a given problem, as the following examples illustrate.

4.2.1 Example 1: Linear model with time-varying coefficient of variation

If $g(\theta_0, \sigma(x), x) = \sigma(x)\theta'_0 x$, we estimate θ_0 by the LS estimator

$$\widehat{\theta} = \left(\sum_{t=1}^{T} w(X_t) X_t X_t' \widehat{\sigma}^2(X_t)\right)^{-1} \sum_{t=1}^{T} w(X_t) Y_t X_t \widehat{\sigma}(X_t).$$

A consistent estimator for σ_W^2 is then given by (6) with

$$\widehat{W}_t = \widehat{A}(X_t)\widehat{u}_t,\tag{10}$$

where $\widehat{u}_t = \widehat{\varepsilon}_t - 0.5 \widehat{\theta}' X_t (\widehat{\varepsilon}_t^2 - 1)$,

$$\widehat{S} = \frac{1}{T} \sum_{t=1}^{T} w(X_t) \widehat{\sigma}^2(X_t) X_t X_t'$$

and

$$\widehat{A}(X_t) = \left[\frac{1}{T}\sum_{t=1}^T w(X_t)X_t'\right]\widehat{S}^{-1}\widehat{\sigma}^2(X_t)X_t - 1.$$

By the uniform consistency of $\hat{\sigma}^2(\cdot)$ and $\hat{\mu}^2(\cdot)$ on R_w , and the consistency of $\hat{\theta}$ it follows that $\hat{\sigma}_W^2 = \sigma_W^2 + o_P(1)$.

4.2.2 Example 2: Nonlinear model with time-varying coefficient of variation

A nonlinear specification for $\mu(X_t)/\sigma(X_t)$ that has been entertained is $g_1(\theta_0, x) = \exp(\theta'_0 x)$, see, e.g., De Santis and Gerard (1997). In this case the LS estimator in (8) leads to a consistent estimator for θ_0 and σ_W^2 can be estimated by (6) with

$$\widehat{W}_t = \widehat{A}(\widehat{\theta}, X_t)\widehat{u}_t,$$

where $\widehat{u}_t = \widehat{\varepsilon}_t - 0.5 \exp(\widehat{\theta}' X_t) (\widehat{\varepsilon}_t^2 - 1)$,

$$\widehat{A}(\widehat{\theta}, X_t) = \widehat{\varphi}_1 \widehat{S}^{-1} \widehat{\sigma}^2(X_t) X_t \exp(\widehat{\theta}' X_t) - 1,$$
$$\widehat{\varphi}_1 = \frac{1}{T} \sum_{t=1}^T w(X_t) X_t' \exp(\widehat{\theta}' X_t)$$

and

$$\widehat{S} = \frac{1}{T} \sum_{t=1}^{T} w(X_t) \widehat{\sigma}^2(X_t) X_t X_t' \exp(2\widehat{\theta}' X_t).$$

4.3 Nonparametric significance testing

In nonparametric significance testing we are interested in testing

$$H_0: E(Y_t \mid X_t = x) = E(Y_t \mid X_{1t} = x_1),$$

against nonparametric alternatives, where $X_t = (X_{1t}, X_{2t})$ and $x = (x_1, x_2)$. This corresponds in our setting to $g(\theta, \eta_0(x), x) = \eta_0(x_1)$, where $\eta_0(x_1) = E(Y_t \mid X_{1t} = x_1)$. We consider a local polynomial estimator for η_0 , denoted by $\hat{\mu}_0(x)$. The proof of Theorem 2 directly shows that $\hat{\mu}_0(x)$ satisfies A6 with

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu_0(x)) dF_X(x) = \int \iota(x_1) (\widehat{\mu}_0(x_1) - \mu_0(x_1)) dF_{X_1}(x_1)$$
$$= T^{-1} \sum_{t=1}^T \iota(X_{1t}) \varepsilon_t + o_P(T^{-1/2}),$$

with $\iota(x_1) = E(w(X_t)/\sigma(X_t) | X_{1t} = x_1)$. Therefore, in this application $l(Y_t, X_t) = \iota(X_{1t})\varepsilon_t/w(X_t)$ and

$$W_t = \varepsilon_t \left\{ \iota(X_{1t}) / w(X_t) - 1 \right\},\tag{11}$$

which can be consistently estimated by replacing ε_t by $\hat{\varepsilon}_t$ and $\iota(X_{1t})$ by a local polynomial estimator of $E(w(X_t)/\sigma(X_t) \mid X_{1t} = x_1)$. Since the null hypothesis does not involve a

parametric component, it holds that $d(\cdot) \equiv 0$, and the drift under local alternatives for this example is given by

$$\delta(a) = \frac{E[w(X_t)a(X_t)]}{(E[w^2(X_t)W_t^2])^{1/2}},$$

with W_t defined in (11).

5 Simulation study

In this section we will briefly illustrate the finite sample performance of the asymptotic distribution-free tests based on the Kolmogorov-Smirnov (KS_T) and Cramér-von Mises (CM_T) statistics for the applications described in the previous section. In all cases, the rejection probabilities are based on 1000 simulated data sets. The estimation of the finitedimensional parameter involved in the specification given in the null hypothesis is done by using the least squares estimator, as in (8), and the variance σ_W^2 is estimated as explained in (6). Nonparametric estimators of $\mu(\cdot)$ and $\sigma(\cdot)$ are obtained by local-linear estimation and Nadaraya-Watson estimation, respectively, with fixed bandwidths and least-squares cross-validation bandwidths (indicated as 'cv' in the tables). The nominal level is 0.05 in all cases.

5.1 Goodness-of-fit tests for parametric models for μ

In this set of simulations we will deal with the goodness-of-fit problem of parametric models for the regression function $\mu(\cdot)$. The null hypothesis is $H_0: \mu(x) = \theta x$. We generate i.i.d. samples of sizes T = 100 and T = 200 from the model proposed in Van Keilegom *et al.* (2008)

$$Y = \theta X + a(X) + 0.20(1+X)\varepsilon,$$

where the covariate X has a uniform distribution on [0, 1] and the error ε is standard normal. The parameter is fixed at $\theta = 1$. The term a(X) gives different possibilities:

- (i) a(x) = 0;
- (ii) $a(x) = x^2;$
- (iii) $a(x) = 0.5x \exp(x);$
- (iv) $a(x) = 0.3\sin(4\pi x)$.

Model (i) is under the null hypothesis, whereas models (ii), (iii) and (iv) are under the alternative. In this case, since the covariate has compact support, the weight function is $w(x) \equiv 1$ and therefore, the tests will be based on the regular empirical distributions.

Table 1 summarizes the obtained results. For model (i), the critical values obtained from the asymptotic distribution of the test statistics produce a slight overestimation of the nominal level when T = 100, but the approximation of the level is already very good for T = 200 for all values of the smoothing parameter. The results for the other models show that the proposed tests reach good power. For model (ii) and (iii) CM_T yields better power than KS_T , and the contrary happens under model (iv). The influence of the choice of the smoothing parameter only seems to have some relevance for CM_T in model (iv).

Table 1 also shows a comparison with the test statistics proposed in Van Keilegom et al. (2008), which are denoted by KS_{VK} and CM_{VK} . The critical values for these test statistics are approximated by means of a smooth bootstrap of residuals (see details in the paper). The level approximation for the bootstrap tests is good. In terms of power, the results obtained with KS_T and CM_T for models (ii) and (iii) are better than the ones obtained with CM_{VK} and KS_{VK} , respectively. On the other hand, the contrary happens under model (iv), which shows a difference in favor of the bootstrap test CM_{VK} , specially for the CM-type statistics. We must recall that the tests based on a bootstrap approximation require the choice of a second bandwidth and are more computationally demanding. In the view of the results of this simulation, it seems that the proposed asymptotically distribution-free tests are reasonable competitors.

[Table 1 (at the end of the manuscript) to be placed around here]

5.2 Constrained mean-variance models

In this section we will investigate the practical performance of the proposed methodology to test for relationships between $\mu(\cdot)$ and $\sigma(\cdot)$. In this case, we consider data from pairs (X_t, Y_t) with a dependence structure given by $X_t = Y_{t-1}, t \in \mathbb{Z}$. More precisely, sequences of sizes T = 200 and T = 500 are generated from the following data generating processes:

(i)
$$Y_t = 0.1X_t(1+0.2X_t^2)^{1/2} + (1+0.2X_t^2)^{1/2}\varepsilon_t;$$

(ii)
$$Y_t = \exp(-0.1X_t^2)(1+0.2X_t^2)^{1/2} + (1+0.2X_t^2)^{1/2}\varepsilon_t;$$

(iii) $Y_t = 0.2 + (1 + 0.2X_t^2)^{1/2}\varepsilon_t;$

(iv) $Y_t = \exp(-0.1X_t^4)(1+0.2X_t^2)^{1/2} + (1+0.2X_t^2)^{1/2}\varepsilon_t.$

In this case, since the covariate has no compact support, the nonparametric estimation of the conditional mean function and variance function are performed on the [5%, 95%]range of the covariate by conveniently adapting the weight function w.

Two null hypotheses will be tested: (a) $H_0: \mu(x) = \theta x \sigma(x)$, for which model (i) is under the null; and (b) $H_0: \mu(x) = \exp(-\theta x)\sigma(x)$, for which model (ii) is under the null. These specifications provide parametric models for the Sharpe ratio $\mu(X_t)/\sigma(X_t)$.

Table 2 displays the results. In the case of the null hypothesis (a) (left part of the table), the approximation of the level in model (i) is good and the behavior in terms of power is very satisfactory. The results for model (iii) show that KS_T and CM_T yield almost the same power. On the other hand, the right part of the table shows the results for the null hypothesis (b). In this case, the level is well approximated for sample size T = 500. The power is excellent under both models (i) and (iii). Although model (iv) is very close to the null hypothesis, we can see that both tests reach non-trivial power.

[Table 2 (at the end of the manuscript) to be placed around here]

5.3 Nonparametric significance testing

In this section we will perform a small simulation study to test for covariate significance in nonparametric regression. More precisely, the null hypothesis is

$$H_0: E(Y_t \mid X_t = x) = E(Y_t \mid X_{1t} = x_1),$$

where $X_t = (X_{1t}, X_{2t})$ is a bidimensional covariate and $x = (x_1, x_2)$. The null hypothesis states that in the given regression model, only the first component of the covariate is significant. We consider the following specifications of $\mu(x) = E(Y_t \mid X_t = x)$:

- (i) $\mu(x_1, x_2) = x_1;$
- (ii) $\mu(x_1, x_2) = x_1 + x_2;$
- (iii) $\mu(x_1, x_2) = x_1 x_2.$

Model (i) satisfies the null hypothesis, whereas models (ii) and (iii) are under the alternative hypothesis. In the simulations, the bidimensional covariate (X_{1t}, X_{2t}) is drawn from independent uniform distributions on [0, 1], the conditional variance function is $\sigma^2(x_1, x_2) = 0.1 + x_1 + x_2$ and the regression error ε_t is a standard normal. The weight function needed in the construction of the test statistics is $w \equiv 1$.

Table 3 collects the results obtained from data sets with sizes T = 200 and T = 500. To perform the test, bidimensional nonparametric smoothing is required, hence smoothing parameters of the form $h = (h_1, h_2)$ are considered, with $h_1 = h_2 = 0.3, 0.4, 0.5$. In the case of model (i), the asymptotic distributions of the test statistics do not provide a good approximation of the nominal level when the sample size is T = 200. However, there is a clear improvement when the sample size is increased up to T = 500, with reasonable results specially for the test based on KS_T . Models (ii) and (iii) show the consistency of the tests in terms of power, as it increases as the sample size increases. The impact of the choice of the smoothing parameter is more relevant for model (iii) than for model (ii).

[Table 3 (at the end of the manuscript) to be placed around here]

6 Conclusions

In this article, we have proposed a general and simple-to-implement methodology for testing semiparametric hypotheses in regression models. The tests are based on the cumulative difference of the standardized errors' distributions under the null and alternative hypotheses, respectively. The asymptotic null distributions of the tests are known functionals of a standard normal random variable, for which critical values are readily available. The tests are consistent against fixed alternatives and are able to detect local alternatives converging to the null at the parametric rate. Some Monte Carlo experiments have shown a satisfactory finite sample performance for the tests in three different applications.

We now point out several topics for future research. We have not discussed the choice of the bandwidth parameters in our testing problem. Although there exists an extensive literature on bandwidth choice for estimation, there is no general theory available for testing purposes. One possible approach in our context is to choose the bandwidth that maximizes the test statistic subject to convergence constraints on the bandwidth. This procedure is likely to be more stable for our methodology than for alternative nonparametric tests based on smoothers, since the rates of convergence of our tests do not depend on those of the bandwidth under standard rate conditions on bandwidth parameters. To apply these ideas we would need to establish the expansion of Theorem 2 uniformly in the bandwidth parameters in a suitable range that converges to zero. This uniform expansion is feasible given existing results; see e.g. Escanciano, Jacho-Chavez and Lewbel (2014). Another extension would be to semiparametric hypotheses on the conditional variance, as in Dette, Neumeyer and Van Keilegom (2007). Our transformation does not lead to asymptotic distribution-free tests in that context, but alternative transformations may exist. Finally, it would be interesting to investigate whether bootstrap procedures lead to asymptotic refinements when applied to our asymptotic distribution-free tests.

Acknowledgments

The first author acknowledges research support from the Spanish Plan Nacional de I+D+I, reference number ECO2012-33053. The second author acknowledges research support from the Spanish Ministerio de Ciencia e Innovación (project MTM2011-23204, FEDER support included), Xunta de Galicia and Universidade de Vigo. The third author acknowledges research support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650, from IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy), and from the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

Appendix: Proofs

Proof of Theorem 1. The implications $(i) \Rightarrow (ii) \Rightarrow (iii)$ are obvious. To prove $(iii) \Rightarrow (ii)$ it suffices to take the derivative of D(y). Finally, let us prove $(ii) \Rightarrow (i)$. If ε and ε_0 have the same distribution, then it also holds that $E(\varepsilon) = E(\varepsilon_0)$ and $Var(\varepsilon) = Var(\varepsilon_0)$. It is easy to see that $E(\varepsilon_0) = E(\varepsilon) + E[(\mu(X_t) - \mu_0(X_t))/\sigma(X_t)]$, and hence $E[(\mu(X_t) - \mu_0(X_t))/\sigma(X_t)] = 0$. On the other hand, we also have that $Var(\varepsilon_0) = Var(\varepsilon) + Var[(\mu(X_t) - \mu_0(X_t))/\sigma(X_t)]$, and therefore $Var[(\mu(X_t) - \mu_0(X_t))/\sigma(X_t)] = 0$. We can now conclude that $P((\mu(X_t) - \mu_0(X_t))/\sigma(X_t)) = 0$, or $\mu(X_t) = \mu_0(X_t)$ a.s. \Box Proof of Theorem 2. First consider

$$\frac{1}{T\overline{w}}\sum_{t=1}^{T}w(X_t)\int_{-\infty}^{y} \left\{ I(\widehat{\varepsilon}_t \le s) - I(\varepsilon_t \le s) \right\} ds$$

$$= \frac{1}{T\overline{w}}\sum_{t=1}^{T}w(X_t) \left\{ I(\widehat{\varepsilon}_t \le y) \left(y - \widehat{\varepsilon}_t\right) - I(\varepsilon_t \le y) \left(y - \varepsilon_t\right) \right\}$$

$$= \frac{1}{T\overline{w}}\sum_{t=1}^{T}w(X_t)I(\varepsilon_t \le y) \left(\varepsilon_t - \widehat{\varepsilon}_t\right)$$

$$+ \frac{1}{T\overline{w}}\sum_{t=1}^{T}w(X_t) \left\{ I(\widehat{\varepsilon}_t \le y) - I(\varepsilon_t \le y) \right\} (y - \widehat{\varepsilon}_t)$$

$$= A(y) + B(y) :: (say).$$

The term B(y) can be bounded as follows :

$$B(y) = \frac{1}{T\overline{w}} \sum_{t=1}^{T} w(X_t) I(\widehat{\varepsilon}_t \le y < \varepsilon_t) (y - \widehat{\varepsilon}_t)$$
$$-\frac{1}{T\overline{w}} \sum_{t=1}^{T} w(X_t) I(\varepsilon_t \le y < \widehat{\varepsilon}_t) (y - \widehat{\varepsilon}_t)$$
$$\le \frac{1}{T\overline{w}} \sum_{t=1}^{T} w(X_t) I(\widehat{\varepsilon}_t \le y < \varepsilon_t) (\varepsilon_t - \widehat{\varepsilon}_t)$$
$$+\frac{1}{T\overline{w}} \sum_{t=1}^{T} w(X_t) I(\varepsilon_t \le y < \widehat{\varepsilon}_t) (\widehat{\varepsilon}_t - \varepsilon_t).$$

Each of the above terms is $O_P(\max_t(\widehat{\varepsilon}_t - \varepsilon_t)^2) = O_P((Th^d)^{-1}\log T) = o_P(T^{-1/2})$ uniformly in y. Indeed, the first term on the right hand side is bounded by

$$\frac{1}{T\overline{w}}\max_{t}|\widehat{\varepsilon}_{t}-\varepsilon_{t}|\max_{t}w(X_{t})\sum_{t=1}^{T}I(\widehat{\varepsilon}_{t}\leq y\leq \varepsilon_{t}).$$

The sum in the latter expression follows a binomial distribution, conditionally on the estimators $\hat{\mu}(\cdot)$ and $\hat{\sigma}(\cdot)$. The probability of success is $P(\hat{\varepsilon}_t \leq y \leq \varepsilon_t)$, which is bounded by $\sup_y f_{\varepsilon}(y) \max_t |\hat{\mu}(X_t) - \mu(X_t)| + \sup_y |yf_{\varepsilon}(y)| \max_t |\hat{\sigma}(X_t) - \sigma(X_t)|$. Hence, the first term above is of the order $O_P(\max_t |\hat{\mu}(X_t) - \mu(X_t)|^2) + O_P(\max_t |\hat{\sigma}(X_t) - \sigma(X_t)|^2) = O_P((Th^d)^{-1}\log T) = o_P(T^{-1/2})$. The second term can be bounded in a similar way. Next, it can be shown that $B_0(y)$, obtained by replacing $\hat{\varepsilon}_t$ with $\hat{\varepsilon}_{t0}$ in the definition of B(y), is

also $o_P(T^{-1/2})$ by using Assumption A6-(i). Consider now

$$A(y) - A_0(y)$$

$$= \frac{1}{T\overline{w}} \sum_{t=1}^T w(X_t) I(\varepsilon_t \le y) \left(\widehat{\varepsilon}_{t0} - \widehat{\varepsilon}_t\right)$$

$$= \frac{1}{T\overline{w}} \sum_{t=1}^T w(X_t) I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\widehat{\sigma}(X_t)}$$

$$= \frac{1}{TE[w(X_t)]} \sum_{t=1}^T w(X_t) I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)} + o_P(T^{-1/2}).$$

We will now show that

$$\frac{1}{TE[w(X_t)]} \sum_{t=1}^{T} \left[w(X_t) I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)} - E\left\{ w(X_t) I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)} \right\} \right] = o_P(T^{-1/2}), \quad (12)$$

uniformly in $-\infty < y < \infty$. Define the class

$$\mathcal{F} = \Big\{ (x, e) \to w(x)\sigma^{-1}(x)I(e \le y)v(x) : -\infty < y < +\infty, v \in C^{\alpha}_M(R_w) \Big\},\$$

where $C_M^{\alpha}(R_w)$ is the space of continuous functions v defined on the compact set R_w , for which

$$\|v\|_{\alpha} = \max_{k \le \underline{\alpha}} \sup_{x} |D^{k}v(x)| + \max_{k = \underline{\alpha}} \sup_{x,x'} \frac{|D^{k}v(x) - D^{k}v(x')|}{\|x - x'\|^{\alpha - \underline{\alpha}}} \le M < \infty,$$

where $\underline{\alpha}$ is the largest integer strictly smaller than α (which we choose later in the proof), $k = (k_1, \ldots, k_d),$

$$D^k = \frac{\partial^{k.}}{\partial x_1^{k_1} \dots \partial x_d^{k_d}},$$

and $k_{\cdot} = \sum k_i$. Note that

$$P\left(\widehat{\mu} - \widehat{\mu}_0 \in C^{\alpha}_M(R_w)\right) \to 1$$

as T tends to infinity, by Assumption A6-(ii). We will show that this class is Donsker. A sufficient condition for the class \mathcal{F} to be Donsker is that

$$\int_{0}^{2M} \sqrt{\log N_{[]}(\delta, \mathcal{F}, \|\cdot\|_{2,\beta})} \, d\delta < \infty,$$

where for any function g,

$$\|g\|_{2,\beta}^2 = \int_0^1 \beta^{-1}(u) Q_g^2(u) du,$$

and where β^{-1} is the inverse cadlag of the decreasing function $u \to \beta_{\lfloor u \rfloor}$ ($\lfloor u \rfloor$ being the integer part of u, and β_t being the mixing coefficient) and Q_g is the inverse cadlag of the tail function $u \to P(||g|| > u)$ (see Section 4.3 in Dedecker and Louhichi, 2002). Here, $N_{[]}(\delta, \mathcal{F}, || \cdot ||_{2,\beta})$ is the δ -bracketing number of the class \mathcal{F} , i.e. it is the smallest number of δ -brackets needed to cover the space \mathcal{F} , where a δ -bracket is the set of all functions h such that $h_{\ell} \leq h \leq h_u$ and where (h_{ℓ}, h_u) satisfy $||h_u - h_{\ell}||_{2,\beta} \leq \delta$. From Corollary 2.7.2 in Van der Vaart and Wellner (1996) it follows that $N_{[]}(\delta, C_M^{\alpha}(R_w), || \cdot ||_2) \leq \exp(K\delta^{-d/\alpha})$. Moreover, defining $y_j = F_{\varepsilon}^{-1}(j\delta)$ for $j = 1, \ldots, O(\delta^{-1})$, it is easily seen that $N_{[]}(\delta, \mathcal{F}, || \cdot ||_{2,\beta}) = O(\delta^{-1} \exp(K\delta^{-d/\alpha}))$. It now follows that the class \mathcal{F} is Donsker, provided $\alpha > d/2$. Next, since

$$\operatorname{Var}\left\{w(X_t)I(\varepsilon_t \le y)\frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)}\right\} \le K \sup_{x \in R_w} |\widehat{\mu}(x) - \widehat{\mu}_0(x)|^2 \xrightarrow{P} 0,$$

it follows from Corollary 2.3.12 in Van der Vaart and Wellner (1996) that (12) holds true. It remains to calculate

$$\frac{1}{E(w(X_t))} E\left\{w(X_t)I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)}\right\} = -\frac{F_{\varepsilon}(y)}{E(w(X_t))} \int w(x) \frac{g(\widehat{\theta}, \widehat{\sigma}(x), x) - g(\theta_0, \sigma(x), x)}{\sigma(x)} dF_X(x) + \frac{F_{\varepsilon}(y)}{E(w(X_t))} \int w(x) \frac{\widehat{\mu}(x) - \mu(x)}{\sigma(x)} dF_X(x),$$
(13)

which follows from the independence between X_t and ε_t .

It is straightforward to show that

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}(x) - \mu(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) \varepsilon_t + o_P(T^{-1/2}),$$

whereas by Assumption A6-(iii) under the null hypothesis

$$\int \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) + o_P(T^{-1/2}).$$

Therefore, by the arguments above, uniformly in y:

$$\int_{-\infty}^{y} [\widehat{F}_{\varepsilon 0}(s) - \widehat{F}_{\varepsilon}(s)] \, ds = A_0(y) - A(y) = \frac{F_{\varepsilon}(y)}{TE[w(X_t)]} \sum_{t=1}^{T} w(X_t) W_t + o_P(T^{-1/2}).$$

This finishes the proof.

Proof of Proposition 4. First note that by the fundamental theorem of calculus $C \equiv 0$ if and only $R \equiv 0$. Then, arguing as in the proof of Theorem 1, it can be shown that $R \equiv 0$ implies that

$$P(\mu(X_t) \neq \mu_0(X_t) \mid X_t \in R_w) = 0.$$

This finishes the proof.

Proof of Corollary 5. The proof parallels the proof of Theorem 2, except for the calculations starting from equation (13). Write

$$\frac{1}{E(w(X_t))} E\left\{w(X_t)I(\varepsilon_t \le y) \frac{\widehat{\mu}(X_t) - \widehat{\mu}_0(X_t)}{\sigma(X_t)}\right\} \\
= -\frac{F_{\varepsilon}(y)}{E(w(X_t))} \int w(x) \left\{\frac{g(\widehat{\theta}, \widehat{\sigma}(x), x) - g(\theta_0, \sigma(x), x)}{\sigma(x)} - \frac{a(x)}{\sqrt{T}}\right\} dF_X(x) \\
+ \frac{F_{\varepsilon}(y)}{E(w(X_t))} \int w(x) \frac{\widehat{\mu}(x) - \mu(x)}{\sigma(x)} dF_X(x).$$

Using Assumption A6bis-(iii), the first term on the right hand side equals

$$\int \left\{ \frac{w(x)}{\sigma(x)} (\widehat{\mu}_0(x) - \mu(x)) - \frac{w(x)a(x)}{\sqrt{T}} \right\} dF_X(x)$$

= $T^{-1} \sum_{t=1}^T w(X_t) l(Y_t, X_t) - T^{-1/2} \int [w(x) + d(x)] a(x) F_X(x) + o_P(T^{-1/2}).$

It now follows that

$$\int_{-\infty}^{y} [\widehat{F}_{\varepsilon 0}(s) - \widehat{F}_{\varepsilon}(s)] ds$$

= $\frac{F_{\varepsilon}(y)}{TE[w(X_t)]} \sum_{t=1}^{T} w(X_t) W_t + T^{-1/2} \frac{E[(w(X_t) + d(X_t)) a(X_t)]}{E[w(X_t)]} F_{\varepsilon}(y) + o_P(T^{-1/2}).$

From this and from the formula of σ_W^2 , the expression of the non-centrality parameter $\delta(a)$ follows.

Proof of Lemma 6. If we consider $g(\theta, u, x)$ as a function in the two first arguments and apply a Taylor expansion, we obtain

$$g(\widehat{\theta}, \widehat{\sigma}(x), x) - g(\theta_0, \sigma(x), x) = (\widehat{\theta} - \theta_0)' \frac{\partial g(\theta_0, \sigma(x), x)}{\partial \theta} + (\widehat{\sigma}(x) - \sigma(x)) \left. \frac{\partial g(\theta_0, u, x)}{\partial u} \right|_{u = \sigma(x)} + o_P(T^{-1/2}).$$

For simplicity we take d = 1 and p = 0 in what follows, but the results can be easily extended to d > 1 and p > 0. Denote the derivatives of μ and σ by $\dot{\mu}(x) = d\mu(x)/dx$ and $\dot{\sigma}(x) = d\sigma(x)/dx$, and consider

$$\widehat{\sigma}(x) - \sigma(x) = \left[2\sigma(x)f_X(x)\right]^{-1}(Th)^{-1}\sum_{t=1}^T K\left(\frac{x-X_t}{h}\right) \left[(Y_t - \mu(x))^2 - \sigma^2(x)\right] + o_P(T^{-1/2}).$$

Hence, it can be easily shown that using a Taylor expansion around $x = X_t$, we obtain (provided $Th^4 \rightarrow 0$)

$$\int \frac{w(x)}{\sigma(x)} \frac{\partial g(\theta_0, u, x)}{\partial u} \Big|_{u=\sigma(x)} (\widehat{\sigma}(x) - \sigma(x)) dF_X(x)$$

$$= \frac{1}{2T} \sum_{t=1}^T \frac{w(X_t)}{\sigma^2(X_t)} \frac{\partial g(\theta_0, u, X_t)}{\partial u} \Big|_{u=\sigma(X_t)} \left[(Y_t - \mu(X_t))^2 - \sigma^2(X_t) \right] + o_P(T^{-1/2})$$

$$= \frac{1}{2T} \sum_{t=1}^T w(X_t) \frac{\partial g(\theta_0, u, X_t)}{\partial u} \Big|_{u=\sigma(X_t)} (\varepsilon_t^2 - 1) + o_P(T^{-1/2}).$$

If we take into account the representation for $\hat{\theta} - \theta_0$ given after the Lemma, we then obtain the following expansion, uniformly in y:

$$\begin{split} &\int_{-\infty}^{y} [\widehat{F}_{\varepsilon 0}(s) - \widehat{F}_{\varepsilon}(s)] \, ds \\ = & \varphi'(\theta_0) \frac{F_{\varepsilon}(y)}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t) s(X_t, \varepsilon_t) \\ & \quad + \frac{1}{2} \frac{F_{\varepsilon}(y)}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t) \left. \frac{\partial g(\theta_0, u, X_t)}{\partial u} \right|_{u=\sigma(X_t)} (\varepsilon_t^2 - 1) \\ & \quad - \frac{F_{\varepsilon}(y)}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t) \varepsilon_t + o_P(T^{-1/2}) \\ = & \frac{F_{\varepsilon}(y)}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t) \Big\{ \varphi'(\theta_0) s(X_t, \varepsilon_t) + \frac{1}{2} \left. \frac{\partial g(\theta_0, u, X_t)}{\partial u} \right|_{u=\sigma(X_t)} (\varepsilon_t^2 - 1) - \varepsilon_t \Big\} \\ & \quad + o_P(T^{-1/2}). \end{split}$$

This finishes the proof.

References

- Bekaert, G. and Harvey, C. R. (1995). Time-varying world market integration. *Journal* of Finance, 50, 403-444.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In: Handbook of Econometrics (James J. Heckman and Edward E. Leamer, editors), Volume 6, Part 2, pages 5549-5632.
- Dedecker, J. and Louhichi, L. (2002). Maximal inequalities and empirical central limit theorems. In: *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, editors). Birkhäuser, Boston.
- De Santis, G. and Gerard, B. (1997). International asset pricing and portfolio diversification with time-varying risk. *Journal of Finance*, 52, 1881-1912.
- Delgado, M. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics*, 29, 1469-1507.
- Dette, H., Pardo-Fernández, J.C. and Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics*, 36, 782-799.
- Dette, H., Marchlewski, M. and Wagener, J. (2012). Testing for a constant coefficient of variation in nonparametric regression by empirical processes. *Annals of the Institute of Statistical Mathematics*, 64, 1045-1070.
- Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society - Series B*, 69, 903-917.
- Doukhan, P. (1994). Mixing. Properties and Examples, Springer, Lecture Notes in Statistics.
- Eagleson, G. K. and Müller, H. G. (1997). Transformations for smooth regression models with multiplicative errors. *Journal of the Royal Statistical Society - Series B*, 59, 173-189.
- Escanciano, J.C., Jacho-Chávez, D. and Lewbel, A. (2014). Uniform convergence of weighted sums of non- and semi-parametric residuals for estimation and testing. *Jour*nal of Econometrics, 178, 426-443.
- Escanciano, J.C., Pardo-Fernández, J.C. and Van Keilegom, I. (2013). Semiparametric estimation of risk-return relationships. Unpublished manuscript.

- Fan, J. and Yao, Q. (2003). Nonlinear Time Series. Nonparametric and Parametric Methods. Springer, New York.
- Ferson, W. (1989). Changes in expected security returns, risk and level of interest rates, Journal of Finance, 44, 1191-1217.
- Ferson, W., Foerster, S. R. and Keim, D. B. (1993). General tests of latent variable models and mean-variance spanning. *Journal of Finance*, 48, 131-156.
- González-Manteiga, W. and Crujeiras, R.M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, 22, 361-411.
- Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24, 726-748.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. Annals of Statistics, 21, 1926-1947.
- Harvey, C.R. (1989). Time-varying conditional covariances in tests of asset pricing models, Journal of Financial Economics, 24, 289-317.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.
- Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41, 867-887.
- Merton, R.C. (1980). On estimating the expected return on the market. An explanatory investigation. *Journal of Financial Economics*, 8, 323-361.
- Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, 101, 1067-1078.
- Stute, W. (1997). Nonparametric model checks for regression. Annals of Statistics, 25, 613-641.
- Van der Vaart, A.W. and Wellner, J.A. (1996). Weak Convergence and Empirical Processes. With Applications to Statistics. Springer.
- Van Keilegom, I., González-Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, 17, 401-415.

Table 1: Observed rejection proportions in 1000 simulated data sets when the null hypothesis is $H_0: \mu(x) = \theta x$ with the tests based on CM_T and KS_T and those proposed in Van Keilegom *et al.* (2008), indicated as CM_{VK} and KS_{VK} . Model (i) is under the null; models (ii), (iii) and (iv) are under the alternative hypothesis. The significance level is 0.05.

			K	S_T	Cl	M_T	KS	$\tilde{\mathcal{O}}_{VK}$		CN	I_{VK}
model	h	T:	100	200	100	200	100	200	_	100	200
(i)	0.15		0.071	0.061	0.082	0.061	0.038	0.042		0.032	0.047
	0.2		0.064	0.055	0.073	0.056	0.039	0.043		0.036	0.039
	0.25		0.063	0.055	0.070	0.061	0.042	0.048		0.037	0.043
	C-V		0.072	0.058	0.079	0.059	0.054	0.046		0.038	0.048
(ii)	0.15		0.881	0.991	0.927	0.995	0.736	0.975		0.839	0.989
	0.2		0.847	0.982	0.897	0.993	0.740	0.975		0.816	0.986
	0.25		0.795	0.970	0.871	0.986	0.719	0.967		0.793	0.977
	C-V		0.875	0.992	0.927	0.998	0.745	0.974		0.833	0.991
(iii)	0.15		0.825	0.975	0.882	0.989	0.681	0.964		0.786	0.975
	0.2		0.779	0.963	0.848	0.980	0.698	0.953		0.761	0.968
	0.25		0.730	0.949	0.814	0.971	0.676	0.935		0.732	0.957
	C-V		0.833	0.982	0.892	0.994	0.700	0.955		0.781	0.983
(iv)	0.15		0.767	0.945	0.604	0.854	0.846	0.995		0.852	0.997
	0.2		0.717	0.930	0.517	0.777	0.813	0.994		0.798	0.991
	0.25		0.720	0.933	0.524	0.781	0.764	0.982		0.740	0.981
	C-V		0.821	0.963	0.698	0.914	0.865	0.997		0.887	0.998

Table 2: Observed rejection proportions in 1000 simulated data sets when the null hypothesis is (a) H_0 : $\mu(x) = \theta x \sigma(x)$ (left panel, model (i) is under the null), or (b) H_0 : $\mu(x) = \exp(-\theta x)\sigma(x)$ (right panel, model (ii) is under the null). The significance level is 0.05.

		test:	(a) $H_0: \mu(x) = \theta x \sigma(x)$			(b) $H_0: \mu(x) = \exp(-\theta x)\sigma(x)$					
			KS_T		CM_T		K	KS_T		M_T	
models	h	T:	200	500	200	500	200	500	200	500	
(i)	0.50		0.051	0.036	0.054	0.037	1.000	1.000	1.000	1.000	
	0.75		0.047	0.034	0.052	0.034	1.000	1.000	1.000	1.000	
	1.00		0.048	0.033	0.049	0.034	1.000	1.000	1.000	1.000	
	cv		0.047	0.035	0.054	0.035	1.000	1.000	1.000	1.000	
(ii)	0.50		1.000	1.000	1.000	1.000	0.084	0.059	0.079	0.057	
	0.75		1.000	1.000	1.000	1.000	0.068	0.054	0.069	0.056	
	1.00		1.000	1.000	1.000	1.000	0.059	0.058	0.062	0.057	
	cv		1.000	1.000	1.000	1.000	0.075	0.072	0.074	0.072	
(iii)	0.50		0.753	0.980	0.735	0.980	1.000	1.000	1.000	1.000	
	0.75		0.733	0.981	0.743	0.984	1.000	1.000	1.000	1.000	
	1.00		0.726	0.982	0.742	0.984	1.000	1.000	1.000	1.000	
	cv		0.733	0.983	0.740	0.983	1.000	1.000	1.000	1.000	
(iv)	0.50		1.000	1.000	1.000	1.000	0.219	0.269	0.183	0.235	
	0.75		1.000	1.000	1.000	1.000	0.173	0.230	0.155	0.215	
	1.00		1.000	1.000	1.000	1.000	0.145	0.199	0.141	0.189	
	cv		1.000	1.000	1.000	1.000	0.174	0.240	0.159	0.219	

			K_{i}	S_T	CM_T		
model	(h_1, h_2)	T:	200	500	200	500	
(i)	(0.3, 0.3)		0.104	0.067	0.163	0.076	
	(0.4, 0.4)		0.082	0.059	0.109	0.069	
	(0.5, 0.5)		0.103	0.082	0.113	0.082	
(ii)	(0.3, 0.3)		0.965	1.000	0.985	1.000	
	(0.4, 0.4)		0.958	1.000	0.980	1.000	
	(0.5, 0.5)		0.949	0.999	0.971	0.998	
(iii)	(0.3, 0.3)		0.374	0.536	0.521	0.678	
	(0.4, 0.4)		0.319	0.497	0.418	0.622	
	(0.5, 0.5)		0.313	0.470	0.395	0.565	

Table 3: Observed rejection proportions in 1000 simulated data sets for the significance test H_0 : $E(Y_t | X_t = x) = E(Y_t | X_{1t} = x_1)$. Model (i) is under the null hypothesis. Models (ii) and (iii) are under the alternative hypothesis. The significance level is 0.05.