

ASYMPTOTIC DISTRIBUTION-FREE TESTS FOR SEMIPARAMETRIC REGRESSIONS WITH DEPENDENT DATA

BY JUAN CARLOS ESCANCIANO*, JUAN CARLOS PARDO-FERNÁNDEZ†
AND INGRID VAN KEILEGOM‡

Indiana University, Universidade de Vigo, KU Leuven

This article proposes a new general methodology for constructing nonparametric and semiparametric Asymptotically Distribution-Free (ADF) tests for semiparametric hypotheses in regression models for possibly dependent data coming from a strictly stationary process. Classical tests based on the difference between the estimated distributions of the restricted and unrestricted regression errors are not ADF. In this article we introduce a novel transformation of this difference that leads to ADF tests with well-known critical values. The general methodology is illustrated with applications to testing for parametric models against nonparametric or semiparametric alternatives, and semiparametric constrained mean-variance models. Several Monte Carlo studies and an empirical application show that the finite sample performance of the proposed tests is satisfactory in moderate sample sizes.

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) , $t = 0, \pm 1, \pm 2, \dots$, is strictly stationary and ergodic, and that $E(Y_t^2) < \infty$. Let Y_t be related to X_t through the heteroskedastic regression model

$$(1) \quad Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t,$$

where $\mu(x) = E(Y_t | X_t = x)$ and $\sigma^2(x) = Var(Y_t | X_t = x)$ are almost surely (a.s.) the conditional mean and the conditional variance of Y_t given $X_t = x$, respectively, and ε_t is an error term, which is independent of X_t . Let μ belong to a certain class of measurable functions, \mathcal{M} , that will be specified for each application. This class may consist of semiparametric or nonparametric specifications for μ . In this article, we propose a general methodology

*J.C. Escanciano acknowledges research support from the Spanish Plan Nacional de I+D+I, reference number ECO2014-55858-P.

†J.C. Pardo-Fernández acknowledges research support from the Spanish Ministerio de Economía y Competitividad (grant MTM2014-55966-P).

‡I. Van Keilegom acknowledges support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409), and from IAP Research Network P7/06 of the Belgian State.

for testing parametric or semiparametric hypotheses about the regression function μ . That is, we are interested in testing

$$H_0 : \mu \in \mathcal{M}_0 \quad \text{versus} \quad H_1 : \mu \in \mathcal{M} \setminus \mathcal{M}_0,$$

where

$$\mathcal{M}_0 = \left\{ \mu : \mathbb{R}^d \rightarrow \mathbb{R} \text{ such that } \mu(x) = g(\theta, \eta(x), x), \theta \in \Theta \subset \mathbb{R}^r, \eta \in \mathcal{H} \right\}$$

is a subclass of \mathcal{M} of parametric or semiparametric models defined in terms of a known function g , an unknown parameter $\theta \in \Theta \subset \mathbb{R}^r$ and a possibly infinite-dimensional unknown parameter η belonging to a class of functions \mathcal{H} .

This general formulation covers many testing problems in regression with a parametric/semiparametric null hypothesis versus a semiparametric/nonparametric alternative. 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. See, for instance, the review that [18] offer about this topic. Special cases of semiparametric null hypotheses include the partially linear model with $g(\theta, \eta(x), x) = \theta'x_1 + \eta(x_2)$, $x' = (x'_1, x'_2)$ (where A' denotes the transpose of A), additive models with $g(\theta, \eta(x), x) = \eta_1(x_1) + \eta_2(x_2)$, $\eta = (\eta_1, \eta_2)$, varying-coefficient models $g(\theta, \eta(x), x) = x'_2\eta(x_1)$, and single-index models with $g(\theta, \eta(x), x) = \eta(\theta'x)$, among many others. Our methodology can also be applied to semiparametric alternative hypotheses. For example, a researcher suspecting of non-linear effects might be interested in testing the parametric linear model $g(\theta, \eta(x), x) = \theta'_1x_1 + \theta'_2x_2$, $\theta' = (\theta'_1, \theta'_2)$, against the partially linear model $g(\theta_1, \eta(x), x) = \theta'_1x_1 + \eta(x_2)$, or against the single-index model $g(\theta, \eta(x), x) = \eta(\theta'x)$. The proposed methodology can be equally applied to these different settings.

A new class of examples that we analyze below is that of constrained mean-variance models, where $\eta(x) = \sigma(x)$. The statistical properties of this class of semiparametric models have not been investigated in this generality before. Among other interesting specifications, the choice $g(\theta, \sigma(x), x) = \sigma(x)g_1(\theta, x)$ leads to a parametric model given by $g_1(\theta, x)$ for the nonparametric standardized first moment $\mu(x)/\sigma(x)$. 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. Indeed, there is an extensive literature in financial time series on modelling and testing hypotheses on $\mu(x)/\sigma(x)$, –see, e.g., [28, 29]–, but there are no statistical tests available for parametric specifications of $\mu(x)/\sigma(x)$ when $\mu(x)$ and $\sigma(x)$ are nonparametric.

Tests for parametric and some semiparametric hypotheses on the regression function have been investigated before in the literature, with a special focus on independent and identically distributed (i.i.d) observations; see the above mentioned paper [18] for a recent comprehensive survey on the topic. The two main methodologies are based on comparisons of unrestricted (i.e., nonparametric) estimators of μ and restricted estimators of μ (see, e.g., [21]), or their corresponding cumulative processes (see, e.g., [33, 25, 5, 34]). Test statistics in the first methodology have rates of convergence that depend on suitable bandwidth sequences and lead to tests which are unable to detect local alternatives converging at the parametric rate. Tests based on cumulative processes detect such local alternatives, although their asymptotic local power function is essentially flat (see [23]). Moreover, the “directions” where local power is nontrivial lack interpretation (see [11]). Directional tests that specify the direction of departure exist, see [33] (Section 3) and [34], and although they are easy to interpret, they are inconsistent against all but one fixed alternative.

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, [37] and [7] used this idea to develop goodness-of-fit tests for the parametric form of the regression function and the variance function, respectively. [8] developed goodness-of-fit tests for a multiplicative structure between the regression function and the scale function. Tests using this methodology have parametric rates of convergence, unlike tests based on direct comparisons of unrestricted and restricted estimates of μ . However, their null limit distribution depends on the unknown density of the error distribution and therefore are not Asymptotically Distribution-Free (ADF). Smoothed bootstrap methods have been proposed to approximate the corresponding critical values. These bootstrap methods are computationally intensive and require the choice of additional smoothing parameters.

This article provides three main contributions within the context of the third methodology explained above. First, it proposes a novel transformation of the difference of error distributions that leads to ADF tests. Second, it widens the scope of applications of this third methodology to general semiparametric null and semiparametric/nonparametric alternative hypotheses. Compared to other existing ADF tests available in the literature for dependent data, such as [25]’s ADF test for parametric autoregressions, the proposed tests are ADF in a much more general setting that includes semiparametric null and semiparametric/nonparametric alternative hypotheses.

Third, this article provides a global and local power analysis of tests based on the third methodology. We show that from the local point of view our tests are directional, with a nontrivial local power at the parametric rate that can be tailored to specific alternatives of interest, while from the global point of view our tests are consistent against *all* fixed global alternatives. Thus, the power properties of tests in the third methodology are fundamentally different from those of existing tests based on the first and second methodologies. Monte Carlo experiments show a satisfactory finite sample performance for the proposed tests in three different applications: parametric regression models, mean-variance constrained models and parametric linear regression models against semiparametric partially linear models.

The remainder of this article is organized as follows. In Section 2 we introduce the new methodology in a general context. In Section 3 we discuss practical implementation in several examples. In Section 4 we investigate the asymptotic power properties of the tests. In Section 5 we report some simulation studies and a real data example. Finally, we conclude in Section 6. All mathematical proofs are gathered in an Appendix and in the on-line supplement.

2. General methodology.

2.1. *Test statistics.* This section introduces the general methodology. The discussion here is organized around a few “high-level” assumptions. More specific conditions can be given in specific applications; see Section 3 below for illustrations. 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, for the sake of identifiability, we assume that (θ_0, η_0) exists and is unique. Note that H_0 is equivalent to the fact that $\mu = \mu_0$. Then, define the standardized errors

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

with cumulative distribution functions $F_{\varepsilon_0}(y) = P(\varepsilon_{t0} \leq y)$ and $F_\varepsilon(y) = P(\varepsilon_t \leq y)$, respectively. We show in Theorem 1 below three equivalent formulations for H_0 which motivate the testing methodology introduced in this article.

Theorem 1 *Assume that $E(\varepsilon_{t0}^2) < \infty$. 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 used in recent literature to construct tests for specific examples of H_0 (see, for example, [37] or [8]). In this article we will exploit the equivalence between (i) and (iii), which, as we show below, will allow us to obtain ADF 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) , $t = 1, \dots, T$, is available and construct the estimated residuals

$$(2) \quad \hat{\varepsilon}_{t0} = \frac{Y_t - \hat{\mu}_0(X_t)}{\hat{\sigma}(X_t)} \quad \text{and} \quad \hat{\varepsilon}_t = \frac{Y_t - \hat{\mu}(X_t)}{\hat{\sigma}(X_t)},$$

for $t = 1, \dots, T$, where $\hat{\mu}(x)$ is an estimator of $\mu(x)$ under model \mathcal{M} , $\hat{\sigma}(x)$ is a consistent estimator of $\sigma(x)$ under model \mathcal{M} , and $\hat{\mu}_0(x)$ is a suitable consistent estimator of $\mu_0(x)$ under model \mathcal{M}_0 . For example, if \mathcal{M} is fully nonparametric, then examples of estimators $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ include kernel and sieve estimators. There are also general estimation methods available for semiparametric models that can be used to estimate (θ_0, η_0) , and hence, μ_0 . For example, we can use the sieve least squares estimators $(\hat{\theta}, \hat{\eta}) = \arg \min_{\theta \in \Theta, \eta \in \mathcal{H}_T} T^{-1} \sum_{t=1}^T (Y_t - g(\theta, \eta(X_t), X_t))^2$, where \mathcal{H}_T is a sieve approximation of \mathcal{H} (see e.g. [2] 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) semiparametric least squares estimator $\hat{\theta} = \arg \min_{\theta \in \Theta} T^{-1} \sum_{t=1}^T (Y_t - g(\theta, \hat{\sigma}(X_t), X_t))^2$. Rather than focussing on a specific estimator or class of estimators, here we assume that consistent estimators for $(\mu, \sigma, \theta_0, \eta_0)$, say $(\hat{\mu}, \hat{\sigma}, \hat{\theta}, \hat{\eta})$, are available and satisfy certain conditions, and we refer to the detailed examples below for specific choices of estimators and verification of the conditions. 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

$$\hat{F}_{\varepsilon_0}(y) = \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) I(\hat{\varepsilon}_{t0} \leq y) \quad \text{and} \quad \hat{F}_{\varepsilon}(y) = \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) I(\hat{\varepsilon}_t \leq y),$$

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:

$$(3) \quad \widehat{R}(y) = \frac{f_\varepsilon(y)}{E(w(X_t))} T^{-1/2} \sum_{t=1}^T w(X_t) W_t + o_P(1),$$

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 $0 < \sigma_W^2 < \infty$, where

$$(4) \quad \sigma_W^2 = \frac{E[w^2(X_t)W_t^2]}{(E(w(X_t)))^2}.$$

The random variables W_t and the regularity conditions needed for (3) to hold are of course specific to each application. For instance, [37] obtained this expansion to derive goodness-of-fit tests for parametric regression models, that is, for $g(\theta, \eta(x), x) = g(\theta, x)$, and used it to propose Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM) type statistics. Also [8] gave the analogous expansion and constructed tests for constrained mean-variance models when $g(\theta, \sigma(x), x) = \theta\sigma(x)$ and x is univariate. Tests based on the expansion of $\widehat{R}(y)$ have unknown null limiting distributions, as they depend on the density of the errors $f_\varepsilon(y)$ and other unknown quantities. Both [37] and [8] suggested to implement the tests with the assistance of a smoothed bootstrap procedure, which requires simulations and an additional bandwidth choice.

This article introduces a novel transformation of $\widehat{R}(y)$ that leads to ADF tests. Specifically, we consider the integrated process

$$(5) \quad \widehat{C}(y) = \int_{-\infty}^y \widehat{R}(s) ds,$$

$-\infty < y < \infty$, and show that the expansion

$$\widehat{C}(y) = \frac{F_\varepsilon(y)}{E(w(X_t))} T^{-1/2} \sum_{t=1}^T w(X_t) W_t + o_P(1),$$

holds uniformly in $-\infty < y < \infty$ under the null hypothesis (see Theorem 2 below). This expansion allows to easily obtain the asymptotic null distribution of test statistics that are continuous functionals of $\widehat{C}(\cdot)$. In particular,

taking into account that $\sup_{-\infty < y < \infty} |F_\varepsilon(y)| = 1$ and $\int_{-\infty}^{\infty} F_\varepsilon^2(y) dF_\varepsilon(y) = 1/3$, the expansion above allows us to obtain the asymptotic null distribution of the KS and CM-type statistics

$$KS_T = \sup_{-\infty < y < \infty} \frac{1}{\widehat{\sigma}_W} |\widehat{C}(y)|, \quad \text{and} \quad CM_T = \frac{3}{\widehat{\sigma}_W^2} \int_{-\infty}^{\infty} (\widehat{C}(y))^2 d\widehat{F}_{\varepsilon_0}(y),$$

where $\widehat{\sigma}_W^2$ is a consistent estimator of σ_W^2 in (4), by essentially using the central limit theorem over the asymptotic expansion. We show in the next section that, under H_0 ,

$$(6) \quad KS_T \rightarrow_d |Z| \quad \text{and} \quad CM_T \rightarrow_d Z^2,$$

where $Z \sim N(0, 1)$. Therefore, the proposed tests are very easy to implement for two reasons. First, test statistics are straightforward to compute, since (5) has a closed-form expression as

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

where $a_+ = \max\{a, 0\}$, and therefore,

$$(7) \quad \begin{aligned} KS_T &= \max_{1 \leq t \leq T} \frac{1}{\widehat{\sigma}_W} \max \left\{ |\widehat{C}(\widehat{\varepsilon}_t)|, |\widehat{C}(\widehat{\varepsilon}_{t0})| \right\}, \\ CM_T &= \frac{3}{\widehat{\sigma}_W^2} \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) (\widehat{C}(\widehat{\varepsilon}_{t0}))^2. \end{aligned}$$

Second, critical values are readily available from (6). For instance, the CM test rejects H_0 at significance level α if $CM_T > \chi_{1,1-\alpha}^2$ where CM_T is computed as in (7) and $\chi_{1,1-\alpha}^2$ is the $(1-\alpha)$ -quantile of the chi-square distribution with one degree of freedom.

2.2. A general theory. We investigate now the limit distribution of the test statistics proposed in the previous section under a general set of regularity conditions. To that end, we introduce the following 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 G1 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. [9])

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

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

ASSUMPTION G1: The process (X_t, Y_t) , $t = 0, \pm 1, \pm 2, \dots$, 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 G2:

- (i) 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 $\sup_{x \in R_w} w(x) \leq C$.
- (ii) ε_t is continuous with density $f_\varepsilon(y)$ satisfying $\sup_y f_\varepsilon(y) < \infty$ and $\sup_y |y f_\varepsilon(y)| < \infty$.
- (iii) The errors of the regression model satisfy $E(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}) = E(\varepsilon_t | X_t)$ and $\text{Var}(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}) = E(\varepsilon_t^2 | X_t)$ a.s. Furthermore, ε_t is independent of X_t , with mean zero and unit variance.
- (iv) $\inf_{x \in R_w} \sigma(x) > 0$.

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 . The class of smooth functions $C_M^\tau(R_w)$ is defined in the Appendix (see also p. 154 in [36]).

ASSUMPTION G3: The unrestricted estimators $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ satisfy:

- (i) $\sup_{x \in R_w} |\hat{\mu}(x) - \mu(x)| = o_P(T^{-1/4})$ and $\sup_{x \in R_w} |\hat{\sigma}(x) - \sigma(x)| = o_P(T^{-1/4})$.
- (ii) $P(\hat{\mu} - \mu \in C_M^\tau(R_w)) \rightarrow 1$ as T tends to infinity, where $\tau > d/2$.
- (iii) For a measurable function $l(\cdot)$ satisfying $E(l(Y_t, X_t) | X_t, \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(l^2(Y_t, X_t)) < \infty$, we have

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

ASSUMPTION G4: Under H_0 , the restricted estimator $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\eta}(x), x)$ satisfies:

- (i) $\sup_{x \in R_w} |\hat{\mu}_0(x) - \mu_0(x)| = o_P(T^{-1/4})$.
- (ii) $P(\hat{\mu}_0 - \mu_0 \in C_M^\tau(R_w)) \rightarrow 1$ as T tends to infinity, where $\tau > d/2$.
- (iii) For a measurable function $l_0(\cdot)$ satisfying $E(l_0(Y_t, X_t) | X_t, \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(l_0^2(Y_t, X_t)) < \infty$, we have

$$\int \frac{w(x)}{\sigma(x)} (\hat{\mu}_0(x) - \mu_0(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l_0(Y_t, X_t) + o_P(T^{-1/2}).$$

Assumptions G1-G4 are standard in the literature. If $\widehat{\mu}(x)$, $\widehat{\mu}_0(x)$ and $\widehat{\sigma}(x)$ are kernel estimators, there are well-known conditions for these assumptions to hold, see, for example, [19] for G3(i) and G4(i), and Lemma A.1 in [30] for G3(ii) and G4(ii). Note that [30] assume that the data are i.i.d., whereas here we have dependent data. However, the proof of their Lemma A.1 is based on [27], which is also valid for dependent data. G3(iii) and G4(iii) require a linear asymptotic representation for a linear functional of the estimators. For many examples of nonparametric/semiparametric models and estimators this representation has been already established. [31] (p. 1361) can be used to obtain the expression for $l(Y_t, X_t)$ and $l_0(Y_t, X_t)$ in a general setting. We verify these conditions for our leading examples below. In particular, as we will show later (see Lemma 4 in Section 3), if \mathcal{M} is fully nonparametric and $\widehat{\mu}$ is a kernel-based estimator, then G3(iii) holds with $l(Y_t, X_t) = (Y_t - \mu(X_t))/\sigma(X_t) = \varepsilon_t$.

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

Theorem 2 *Assume G1-G4. Then, under the null hypothesis H_0 the following holds:*

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

uniformly in $-\infty < y < \infty$, where $W_t = l_0(Y_t, X_t) - l(Y_t, X_t)$.

Theorem 2 does not follow directly from (3), since the mapping $R \rightarrow \int_{-\infty}^y R(s) ds$ is not continuous in the space of uniformly bounded functions. Also note that the rate of convergence of \widehat{C} is $T^{-1/2}$ and it does not depend on the dimension of X_t , in contrast to alternative nonparametric tests based on direct comparisons of $\widehat{\mu}(x)$ and $\widehat{\mu}_0(x)$.

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

ASSUMPTION G5: $0 < \sigma_W^2 < \infty$, and there exists a weakly consistent estimator for σ_W^2 , say $\widehat{\sigma}_W^2$.

A natural candidate to estimate σ_W^2 is

$$(8) \quad \widehat{\sigma}_W^2 = \frac{1}{\bar{w}^2 T} \sum_{t=1}^T w^2(X_t) \widehat{W}_t^2,$$

where $\widehat{W}_t = \widehat{l}_0(Y_t, X_t) - \widehat{l}(Y_t, X_t)$ 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 [15]) and the consistency of the estimator $\hat{\sigma}_W^2$.

Corollary 3 *Assume G1-G5 hold. Then, under the null hypothesis H_0 ,*

$$KS_T \rightarrow_d |Z| \quad \text{and} \quad CM_T \rightarrow_d Z^2,$$

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

3. Practical implementation and examples. This section discusses issues of practical implementation. In particular, it provides detailed specific formulae for the estimators, influence functions, asymptotic variances σ_W^2 and their estimators for several examples: (1) parametric models against nonparametric alternatives; (2) parametric linear models against partially linear alternatives; (3) parametric linear models against single-index alternatives; and (4) several constrained mean-variance models. This section also provides primitive regularity conditions under which our general high-level conditions hold for each of these examples. For each example we will explain how to estimate μ and μ_0 .

The implementation of our tests entails computing unrestricted and restricted residuals ($\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t0}$, respectively), the weights $w(X_t)$ and the variance estimator $\hat{\sigma}_W^2$. For the latter we use (8), with an estimated W_t . In all the examples below W_t is a continuous function of ε_t . In our implementations we suggest to replace ε_t by $\hat{\varepsilon}_t$ in W_t , but it is also possible to use $\hat{\varepsilon}_{t0}$ instead, since we only need a consistent estimator of σ_W^2 under the null. Unreported Monte Carlo simulations suggest that the choice of $\hat{\varepsilon}_t$ or $\hat{\varepsilon}_{t0}$ for estimating σ_W^2 has little effect on the finite sample performance of the test statistics.

Although they will not be needed in all examples, we will start by defining nonparametric estimators of $\mu(\cdot)$ and $\sigma(\cdot)$ to ease the presentation. 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

$$(9) \quad \min_{\alpha} \sum_{t=1}^T \left\{ Y_t - P_t(\alpha, x, p) \right\}^2 K_h(X_t - x),$$

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

and $K_h(u) = \prod_{j=1}^d k(u_j/h_j)/h_j$. To estimate $\sigma(x)$, define

$$(10) \quad \hat{\sigma}^2(x) = \hat{\gamma}_0(x) - \hat{\alpha}_0^2(x),$$

where $\hat{\gamma}_0$ is defined in the same way as $\hat{\alpha}_0$, but with Y_t replaced by Y_t^2 in (9), $t = 1, \dots, T$. We note that such an estimator of the conditional variance is not guaranteed to be positive, and we refer to [14] or [26] for alternative positive estimators.

We need assumptions that guarantee that Assumption G3 holds for the local polynomial estimators.

ASSUMPTION N1:

- (i) All partial derivatives of F_X up to order $2d + 1$ exist on the interior of R_w , they are uniformly continuous and $\inf_{x \in R_w} f_X(x) > 0$.
- (ii) All partial derivatives of μ and σ up to order $p + 2$ exist on the interior of R_w , they are uniformly continuous.

ASSUMPTION N2:

- (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 G1.
- (ii) There exists some j' such that for all $j \geq j'$,

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

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

ASSUMPTION N3: 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 N4:

- (i) For all $j = 1, \dots, d$: $h_j/h_1 \rightarrow C_j$, with $0 < C_j < \infty$, and the bandwidth h_1 satisfies $(\log T)^{-1} T^\eta h_1^d \rightarrow \infty$ 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 G1 and N2 respectively, $Th_1^{2d+\delta} \rightarrow \infty$ for some small $\delta > 0$, $Th_1^{2p+2} \rightarrow 0$ for odd p and $Th_1^{2p+4} \rightarrow 0$ 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 $j = 0, \dots, d - 1$.

Assumptions N1-N4 are standard in the literature of nonparametric regression estimation. Assumption N2 and Assumption N4-(i) are taken from [19], and they ensure suitable uniform rates of convergence of the kernel estimators of $\mu(\cdot)$ and $\sigma(\cdot)$.

Lemma 4 *Assume that G1-G2 and N1-N4 hold. Then, Assumption G3 holds for the local polynomial estimator $\hat{\mu}(x)$ with $l(Y_t, X_t) = (Y_t - \mu(X_t))/\sigma(X_t) = \varepsilon_t$.*

The proof is straightforward and is therefore omitted. We proceed to describe semiparametric estimators and regularity conditions for each of the examples enumerated at the beginning of the section.

3.1. *Parametric models against nonparametric alternatives.* Consider the specification $g(\theta, \eta(x), x) = g(\theta, x)$. The testing problem is then a goodness-of-fit test for a parametric form of the regression function $\mu(\cdot)$. The unknown parameter θ_0 can be estimated by the nonlinear least squares estimator $\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T (Y_t - g(\theta, X_t))^2$. The following assumption is needed to verify Assumption G4 in this example:

ASSUMPTION E1: θ_0 belongs to the interior of a compact subset Θ of \mathbb{R}^r . 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. Also, $l_0(Y_t, X_t)$ defined in (11) satisfies $E(l_0^2(Y_t, X_t)) < \infty$. Furthermore, for all $\theta \in \Theta_0$, $g(\theta, \cdot) \in C_M^\tau(R_w)$, for some $\tau > \max(d/2, 1)$.

Lemma 5 *Assume that G1-G2 and E1 hold. Then, Assumption G4 holds for the parametric estimator $\hat{\mu}_0(x) = g(\hat{\theta}, x)$ with*

$$(11) \quad \begin{aligned} l_0(Y_t, X_t) &= w^{-1}(X_t)E(w(X_t)\sigma^{-1}(X_t)g'_{\theta t}(\theta_0)) \\ &\times (E(g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)))^{-1}g_{\theta t}(\theta_0)\sigma(X_t)\varepsilon_t. \end{aligned}$$

In the particular case of testing for a linear model $g(\theta, x) = \theta'x$ and $g_{\theta t}(\theta_0) = X_t$ and thus $l_0(Y_t, X_t) = w^{-1}(X_t)E(w(X_t)\sigma^{-1}(X_t)X_t')(E(X_tX_t'))^{-1}X_t\sigma(X_t)\varepsilon_t$. If the alternative is nonparametric, then $l(Y_t, X_t) = \varepsilon_t$. Hence,

$$W_t = \varepsilon_t \{w^{-1}(X_t)E(w(X_t)\sigma^{-1}(X_t)X_t')(E(X_tX_t'))^{-1}X_t\sigma(X_t) - 1\}.$$

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 (8).

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 [37]. That is, in the definition of ε_{t0} in (2), replace $\widehat{\mu}_0(X_t)$ by $\widetilde{\mu}_0(X_t)$, where $\widetilde{\mu}_0(X_t)$ is obtained in the same way as the nonparametric estimator $\widehat{\mu}(X_t)$, but replacing the responses Y_t by $\widehat{\mu}_0(X_t)$. Our asymptotic theory is still valid when this modification is applied to the estimated residuals.

3.2. Parametric linear model against partially linear model. Our methodology can be also used to test for parametric models against semiparametric alternatives, such as partially linear models. Consider testing for the linear model $Y_t = \theta'_1 X_{1t} + \theta_2 X_{2t} + \sigma \varepsilon_t$, $\theta' = (\theta'_1, \theta_2)$, against the partially linear model $Y_t = \beta' X_{1t} + \eta(X_{2t}) + \sigma \varepsilon_t$, with $\dim(X_{1t}) = d - 1$ and $\dim(X_{2t}) = 1$. Identifiability requires that X_{1t} does not contain an intercept. For simplicity and in order to avoid multivariate estimation of the conditional variance function, in this example we assume a constant error conditional variance, σ^2 . Also, throughout this and the next subsection we assume $w \equiv 1$. For this partially linear model, we take the estimators of β and η given in [32], that is, the unrestricted estimator of μ is $\widehat{\mu}(x) = \widehat{\beta}' x_1 + \widehat{\eta}(x_2)$, where

$$\widehat{\beta} = S_{X,X}^{-1} S_{X,Y} \quad \text{and} \quad \widehat{\eta}(x_2) = \frac{\sum_{t=1}^T k_h(X_{2t} - x_2)(Y_t - \widehat{\beta}' X_{1t})}{\sum_{s=1}^T k_h(X_{2s} - x_2)},$$

where $k_h(u) = h^{-1}k(u/h)$ is the rescaled kernel, $h > 0$ is the smoothing parameter, $S_{X,X} = T^{-1} \sum_{t=1}^T (X_{1t} - \widehat{m}_{X_1}(X_{2t}))(X_{1t} - \widehat{m}_{X_1}(X_{2t}))'$ and $S_{X,Y} = T^{-1} \sum_{t=1}^T (X_{1t} - \widehat{m}_{X_1}(X_{2t}))(X_{1t} - \widehat{m}_Y(X_{2t}))'$, with

$$\widehat{m}_{X_1}(x_2) = \frac{\sum_{t=1}^T k_h(X_{2t} - x_2) X_{1t}}{\sum_{s=1}^T k_h(X_{2s} - x_2)} \quad \text{and} \quad \widehat{m}_Y(x_2) = \frac{\sum_{t=1}^T k_h(X_{2t} - x_2) Y_t}{\sum_{s=1}^T k_h(X_{2s} - x_2)}.$$

Note that $\widehat{m}_{X_1}(x_2)$ and $\widehat{m}_Y(x_2)$ are kernel estimators of $m_{X_1}(x_2) = E(X_{1t}|X_{2t} = x_2)$ and $m_Y(x_2) = E(Y_t|X_{2t} = x_2)$, respectively.

ASSUMPTION E2: Assume that the conditions of the Theorem in [32] (p. 939) are satisfied.

Lemma 6 *Assume that G1-G2 and E2 hold. Then, Assumption G3 holds for the partially linear estimator $\widehat{\mu}(x)$ with $l(Y_t, X_t) = \sigma^{-1} \{Y_t - m_Y(X_{2t}) - \beta' [X_{1t} - m_{X_1}(X_{2t})]\} = \varepsilon_t$.*

Taking into account the influence function of the linear model given in the previous section and that we are working under constant error variance, we obtain that $W_t = (E(X'_t)(E(X_t X'_t))^{-1} X_t - 1) \varepsilon_t$. An estimator of σ_W^2 is then obtained by taking (8) with $\widehat{W}_t = (\{T^{-1} \sum_{t=1}^T X'_t\} \{T^{-1} \sum_{t=1}^T X_t X'_t\}^{-1} X_t - 1) \widehat{\varepsilon}_t$, where $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^T (Y_t - \widehat{\mu}(X_t))^2$ and $\widehat{\varepsilon}_t = \widehat{\sigma}^{-1} \{Y_t - \widehat{m}_Y(X_{2t}) - \widehat{\beta}' [X_{1t} - \widehat{m}_{X_1}(X_{2t})]\}$.

3.3. *Parametric linear model against single-index model.* Next, we consider testing for the linear model $Y_t = X_t'\theta + \sigma\varepsilon_t$, against the single-index model $Y_t = \eta(X_t'\beta) + \sigma\varepsilon_t$. A natural estimator of β under the single-index model is the semiparametric least squares estimator

$$\hat{\beta} = \arg \min_{\beta \in B} \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{\eta}(X_t'\beta|\beta))^2 I(X_t \in A),$$

where A is a subset of \mathbb{R}^d , $\hat{\eta}(u|\beta)$ is a kernel estimator of $\eta(u|\beta) = E(Y_t | X_t'\beta = u)$:

$$\hat{\eta}(u|\beta) = \sum_{t=1}^T \frac{k_h(u - X_t'\beta)Y_t}{\sum_{s=1}^T k_h(u - X_s'\beta)},$$

and B is a compact subset of the unit sphere in \mathbb{R}^d (for identifiability). The estimator of $\mu(x) = \eta(x'\beta)$ is then given by $\hat{\mu}(x) = \hat{\eta}(x'\hat{\beta})$.

ASSUMPTION E3:

- (i) F_X is continuous, and f_X is bounded away from 0 on $A^\delta = \{x \in \mathbb{R}^d : \|x - A\| \leq \delta\}$, where A is an open convex subset of \mathbb{R}^d and $\delta > 0$.
- (ii) η has two bounded, continuous derivatives on $\{x'\beta : x \in A^\delta\}$. Assume that $\int \sup_{\beta \in B} \eta(x'\beta|\beta) dF_X(x) < \infty$.
- (iii) k is supported on $(-1, 1)$ and is a symmetric probability density function with a bounded derivative. Moreover, $Th^2 \rightarrow \infty$ and $Th^4 \rightarrow 0$.
- (iv) $E|\varepsilon_t|^m = M_m < \infty$ for all m .

Assumption E3 is standard in the literature, see [20]. The bounded integral assumption in E3(ii) enables us to interchange integrals with derivatives with respect to β in our next result.

Lemma 7 *Assume that G1-G2 and E3 hold. Then, Assumption G3 holds for the single-index estimator $\hat{\mu}(x)$ with $l(Y_t, X_t) = \sigma^{-1}\{Y_t - \eta(X_t'\beta)\} = \varepsilon_t$.*

3.4. *Constrained mean-variance models.* This section illustrates the general methodology with an application to constrained mean-variance models. In these models $\eta(x) = \sigma(x)$. [8] studied the special case $g(\theta, \sigma(x), x) = \theta\sigma(x)$ and developed bootstrap-based tests for the corresponding hypothesis under the assumption that the covariate is one-dimensional. See also [6] for an alternative test with i.i.d. data. Our more general formulation here is motivated from applications in economics and finance.

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 linear model with $g_1(\theta, x) = \theta'x$ in [35], and the exponential model in [4] and [1], with $g_1(\theta, x) = \exp(\theta'x)$. These examples will be treated in detail in Subsections 3.4.1 and 3.4.2. 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.

In the general case, the parameter θ_0 can be estimated by the following weighted least-squares (LS) estimator

$$(12) \quad \hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{t=1}^T w(X_t) (Y_t - g(\theta, \hat{\sigma}(X_t), X_t))^2,$$

where $\hat{\sigma}(\cdot)$ is defined in (10). See also [13] for a related estimator of θ_0 in a more general context. We now define $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\sigma}(x), x)$. The following assumption is needed to verify Assumption G4 in this general example:

ASSUMPTION E4: The function $g(\theta, u, x)$ is continuously differentiable with respect to the components of θ and u , with derivatives $g_{\theta t}(\theta) = \frac{\partial}{\partial \theta} g(\theta, \sigma(X_t), X_t)$ and $g_{ut}(\theta) = \frac{\partial}{\partial u} g(\theta, u, X_t)|_{u=\sigma(X_t)}$ satisfying that $E(g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0))$ is finite and non-singular, and that $E(g_{ut}^2(\theta_0)) < \infty$. Assume $E(\varepsilon_t^4) < \infty$. Furthermore, for all $\theta \in \Theta_0$, $P(g(\theta, \hat{\sigma}(\cdot), \cdot) \in C_M^\tau(R_w)) \rightarrow 1$, for some $\tau > \max(d/2, 1)$.

Define $u_t = \varepsilon_t - 0.5g_{ut}(\theta_0)(\varepsilon_t^2 - 1)$, $S(\theta_0) = E[w(X_t)g_{\theta t}(\theta_0)g'_{\theta t}(\theta_0)] / E(w(X_t))$, $s(X_t, \varepsilon_t) = S^{-1}(\theta_0)\sigma(X_t)g_{\theta t}(\theta_0)u_t$, and

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

Lemma 8 *Assume that G1-G2, N1-N4 and E4 hold. Then, Assumption G4 holds for the estimator $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\sigma}(x), x)$ with*

$$(13) \quad l_0(Y_t, X_t) = \varphi'(\theta_0)s(X_t, \varepsilon_t) + 0.5g_{ut}(\theta_0)(\varepsilon_t^2 - 1).$$

The first term $s(X_t, \varepsilon_t)$ arises in the linear expansion because (12) satisfies

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

as shown in [13]. The second term in (13) accounts for the effect of estimating σ in $g(\theta, \sigma(X_t), X_t)$. Therefore, for this example G4 holds. As given in Lemma 4, G3(iii) holds for the unrestricted nonparametric estimator with $l(Y_t, X_t) = \varepsilon_t$, and we thus obtain

$$W_t = \varphi'(\theta_0)s(X_t, \varepsilon_t) + 0.5g_{ut}(\theta_0)(\varepsilon_t^2 - 1) - \varepsilon_t = A(X_t)u_t,$$

where $A(X_t) = \varphi'(\theta_0)S^{-1}(\theta_0)\sigma(X_t)g_{\theta t}(\theta_0) - 1$. The variable W_t can be consistently estimated by $\widehat{W}_t = \widehat{A}_t\widehat{u}_t$, with $\widehat{A}_t = \widehat{\varphi}'(\widehat{\theta})\widehat{S}^{-1}(\widehat{\theta})\widehat{\sigma}(X_t)\widehat{g}_{\theta t}(\widehat{\theta}) - 1$,

$$\widehat{u}_t = \widehat{\varepsilon}_t - 0.5\widehat{g}_{ut}(\widehat{\theta})(\widehat{\varepsilon}_t^2 - 1), \quad \widehat{\varphi}'(\widehat{\theta}) = \frac{1}{\bar{w}T} \sum_{t=1}^T \frac{w(X_t)}{\widehat{\sigma}(X_t)} \widehat{g}_{\theta t}(\widehat{\theta}),$$

$\widehat{S}(\widehat{\theta})$ is a consistent estimate for $S(\theta_0)$,

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

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

3.4.1. *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

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

A consistent estimator for σ_W^2 is then given by (8) with $\widehat{W}_t = \widehat{A}(X_t)\widehat{u}_t$, where $\widehat{u}_t = \widehat{\varepsilon}_t - 0.5\widehat{\theta}'X_t(\widehat{\varepsilon}_t^2 - 1)$,

$$\begin{aligned} \widehat{S} &= \frac{1}{T} \sum_{t=1}^T w(X_t)\widehat{\sigma}^2(X_t)X_tX_t', \quad \text{and} \\ \widehat{A}(X_t) &= \left[\frac{1}{T} \sum_{t=1}^T w(X_t)X_t' \right] \widehat{S}^{-1}\widehat{\sigma}^2(X_t)X_t - 1. \end{aligned}$$

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

3.4.2. *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., [4]. In this case the LS estimator in (12) leads to a consistent estimator for θ_0 and σ_W^2 can be estimated by (8) with $\widehat{W}_t = \widehat{A}(X_t)\widehat{u}_t$, where $\widehat{u}_t = \widehat{\varepsilon}_t - 0.5 \exp(\widehat{\theta}'X_t)(\widehat{\varepsilon}_t^2 - 1)$, $\widehat{A}(X_t) = \widehat{\varphi}_1\widehat{S}^{-1}\widehat{\sigma}^2(X_t)X_t \exp(\widehat{\theta}'X_t) - 1$,

$$\begin{aligned} \widehat{\varphi}_1 &= \frac{1}{T} \sum_{t=1}^T w(X_t)X_t' \exp(\widehat{\theta}'X_t), \quad \text{and} \\ \widehat{S} &= \frac{1}{T} \sum_{t=1}^T w(X_t)\widehat{\sigma}^2(X_t)X_tX_t' \exp(2\widehat{\theta}'X_t). \end{aligned}$$

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

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

That is, under this alternative hypothesis H_1 , $P(\mu(X_t) \neq \mu_0(X_t) \mid X_t \in R_w) > 0$. We shall show that under certain conditions the proposed test statistics will diverge to infinity under H_1 as $T \rightarrow \infty$, thereby proving the consistency of the tests against these fixed alternatives. To that end, we show below in Proposition 9 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 $R(s) = E(w(X_t) \{I(\varepsilon_{t0} \leq s) - I(\varepsilon_t \leq s)\})/E(w(X_t))$. Therefore, by the continuous mapping theorem,

$$T^{-1}CM_T = \frac{3}{\sigma_W^2} \int_{-\infty}^{\infty} (C(y))^2 dF_{\varepsilon_0}(y) + o_P(1),$$

and

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

Thus, given that both $C(\cdot)$ and $F_{\varepsilon_0}(\cdot)$ are continuous under our conditions, it follows that the test statistics will diverge to infinity as $T \rightarrow \infty$ if $C(y) \neq 0$ for some $y \in \mathbb{R}$. The next result shows that this is the case under H_1 . The following condition simplifies some of the previous assumptions.

ASSUMPTION G6: It holds:

- (i) $E(\varepsilon_{t0}^2) < \infty$.
- (ii) $\sup_{x \in R_w} |\widehat{\mu}(x) - \mu(x)| = o_P(1)$; $\sup_{x \in R_w} |\widehat{\sigma}(x) - \sigma(x)| = o_P(1)$ and $\sup_{x \in R_w} |\widehat{\mu}_0(x) - \mu_0(x)| = o_P(1)$.

Assumption G6 is standard in the literature. If $\widehat{\mu}(x)$, $\widehat{\mu}_0(x)$ and $\widehat{\sigma}(x)$ are kernel estimators, there are well-known conditions for Assumption G6(ii) to hold, see, for example, [19] and Section 3 of the present article.

Proposition 9 *Assume that G1, G2, G5 and G6 hold. Then, our tests are consistent against H_1 , i.e. the power of our tests tends to 1 when T tends to infinity.*

This shows that our tests are omnibus tests, in the sense that for any fixed alternative in H_1 (i.e. any possible direction $\mu(x) - \mu_0(x)$), the power of our tests goes to one as the sample size goes to infinity.

We consider now the limiting distribution under the local alternative

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

where $a \neq 0$ is the (standardized) direction of departure, such that $E(|a(X_t)|) < \infty$ and $0 < E[a^2(X_t)W_t^2] < \infty$. The following standard assumption is needed to control the behavior of the estimator $\hat{\mu}_0(x)$ under $H_{1T}(a)$:

ASSUMPTION G4BIS: Under $H_{1T}(a)$, the restricted estimator $\hat{\mu}_0(x) = g(\hat{\theta}, \hat{\eta}(x), x)$ satisfies:

- (i) $\sup_{x \in R_w} |\hat{\mu}_0(x) - \mu(x)| = o_P(T^{-1/4})$.
- (ii) $P(\hat{\mu}_0 - \mu_0 \in C_M^\tau(R_w)) \rightarrow 1$ as T tends to infinity, where $\tau > d/2$.
- (iii) For a measurable function $l_0(\cdot)$ satisfying $E(l_0(Y_t, X_t) | X_t, \mathcal{F}_{-\infty}^{t-1}) = 0$ a.s. and $E(l_0^2(Y_t, X_t)) < \infty$, we have

$$\begin{aligned} \int \frac{w(x)}{\sigma(x)} (\hat{\mu}_0(x) - \mu_0(x)) dF_X(x) &= T^{-1} \sum_{t=1}^T w(X_t) l_0(Y_t, X_t) \\ &\quad - T^{-1/2} \int d(x) a(x) dF_X(x) + o_P(T^{-1/2}), \end{aligned}$$

where $d(\cdot)$ depends on the estimator of θ_0 used.

Define

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

Then, we have the following asymptotic result under the local alternative $H_{1T}(a)$.

Corollary 10 *Assume G1-G3, G4bis and G5. Then, under the local alternative $H_{1T}(a)$,*

$$KS_T \rightarrow_d |Z + \delta(a)| \quad \text{and} \quad CM_T \rightarrow_d (Z + \delta(a))^2,$$

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

This result shows that our tests are able to detect local alternatives converging at the parametric rate. Our tests are directional from a local perspective, i.e. the tests are not consistent against all local alternatives, as there are directions a for which $\delta(a) = 0$. This is a generic property of directional

tests, including those in [33] (Section 3) and [34]. On the other hand, note that our tests are consistent against all fixed alternatives and all local alternatives that converge at rate r_T with $1/r_T = o(\sqrt{T})$. In particular for all alternatives of order s_T , with s_T being the rate of convergence of $\widehat{\mu}(\cdot) - \widehat{\mu}_0(\cdot)$, the test statistics diverge to infinity. This is an important advantage of our test with respect to tests that are based on the distance between $\widehat{\mu}(\cdot)$ and $\widehat{\mu}_0(\cdot)$, like e.g. in [21], or the directional tests of [33] (Section 3) and [34] that are inconsistent against a large class of fixed alternatives.

Thus, our tests have the advantages of omnibus tests against fixed global alternatives, and the simplicity of interpretation of directional tests at local alternatives converging at the parametric rate, with local power functions that can be explicitly computed and tailored to specific alternatives of interest. Therefore, from the power point of view our tests have fundamentally different properties from those of existing tests. The following example illustrates these points.

Example. Parametric models, cont.: Suppose we aim to test the linear model $\mu(x) = \theta x$ against all fixed alternatives $\mu(x) \neq \theta x$ with positive probability on R_w (where R_w is a large compact set satisfying our conditions above). Consider, for instance, the CM type test. Proposition 9 shows that our CM-type test is consistent against all fixed alternatives under the conditions of Section 3.1, i.e. under $\mu(x) \neq \theta x$ with positive probability on R_w , $P(CM_T > \chi_{1,1-\alpha}^2) \rightarrow 1$ as $T \rightarrow \infty$. Now, consider, for example, a quadratic local alternative $\mu(x) = \theta x + T^{-1/2}x^2$. From the arguments of Section 3.1 it is straightforward to verify that Assumption G4bis-(iii) holds with

$$d(X_t) = -E(w(X_t)\sigma^{-1}(X_t)X_t)(E(X_t^2))^{-1}X_t\sigma(X_t).$$

We show now that there is a choice of w that makes the drift function $\delta(a) \neq 0$ for the quadratic local alternative $a(x) = \sigma^{-1}(x)x^2$. As in [24] (pp. 996), we can assume without loss of generality that $E[d(X_t)a(X_t)] = 0$. Then, choose $w_0(x) = a(x)I(x \in R_w)$, for a compact set R_w in the support of X_t , so that the corresponding drift is

$$\delta(a) = \frac{E[w_0(X_t)a(X_t)]}{(E[w_0^2(X_t)W_t^2])^{1/2}} \neq 0.$$

Thus, the CM test based on w_0 guarantees a nontrivial local power against quadratic local alternatives, with an asymptotic local power function given by $P(\chi_1^2(\delta(a)) > \chi_{1,1-\alpha}^2)$, where $\chi_1^2(\delta(a))$ is chi-square distribution with one degree of freedom and noncentrality parameter $\delta(a)$. More generally, for any alternative a such that $E[d(X_t)a(X_t)] = 0$ and with non-zero positive part,

we can choose $w(x) = a_+(x)I(x \in R_w)$ for a suitable R_w so that the drift term $\delta(a) \neq 0$. It is in the precise sense that the test can be tailored to specific local alternatives of interest, while preserving the omnibus feature against global alternatives.

5. Numerical examples.

5.1. *Monte Carlo simulations.* In this section we will 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 2000 simulated data sets. The estimation of the finite-dimensional parameter involved in the specification given in the null hypothesis is done by using the least squares estimator, as in (12), and the variance σ_W^2 is estimated as explained in (8). 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 ‘c-v’ in the tables). The nominal level is 0.05 in all cases. Some further numerical results and discussion can be found in the on-line supplement.

5.1.1. *Goodness-of-fit tests for parametric models for μ .* In this first 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 [37],

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

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 constructed on the basis of the regular empirical distributions.

Table 1 summarizes the obtained results for the proposed tests in comparison with the bootstrap-based tests in [37], which are denoted by KS_{VK} and

TABLE 1

Observed rejection proportions in 2000 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 [37], 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. The size of the asymptotic test is adjusted so as to equal the size of the bootstrap test.

model	h	T :	Size-adjusted				Van Keilegom et al. (2008)			
			KS_T		CM_T		KS_{VK}		CM_{VK}	
			100	200	100	200	100	200	100	200
(i)	0.15		.0455	.0410	.0410	.0480	.0455	.0410	.0410	.0480
	0.2		.0500	.0445	.0430	.0455	.0500	.0445	.0430	.0455
	0.25		.0520	.0495	.0460	.0460	.0520	.0495	.0460	.0460
	c-v		.0600	.0490	.0450	.0480	.0600	.0490	.0450	.0480
(ii)	0.15		.8055	.9850	.8435	.9935	.7300	.9715	.8375	.9905
	0.2		.7915	.9805	.8330	.9905	.7350	.9695	.8210	.9865
	0.25		.7710	.9730	.8190	.9855	.7245	.9660	.7970	.9805
	c-v		.8460	.9885	.8555	.9945	.7485	.9725	.8380	.9905
(iii)	0.15		.7330	.9640	.7925	.9860	.6855	.9555	.7910	.9780
	0.2		.7270	.9585	.7710	.9765	.6905	.9480	.7695	.9720
	0.25		.7040	.9500	.7585	.9730	.6775	.9350	.7410	.9595
	c-v		.8015	.9760	.8125	.9875	.6950	.9510	.7800	.9815
(iv)	0.15		.6650	.9115	.3815	.7795	.8395	.9965	.8440	.9980
	0.2		.6675	.9095	.3880	.7135	.8135	.9920	.7930	.9945
	0.25		.6910	.9275	.4445	.7535	.7590	.9810	.7245	.9835
	c-v		.7710	.9440	.4920	.8670	.8635	.9965	.8805	.9990

CM_{VK} . In order to make a fair comparison, we have adjusted the critical values of our tests to match the empirical levels of the bootstrap-based tests. The empirical powers are then size-corrected. To achieve that, instead of taking the critical values from the asymptotic distributions given in Corollary 3, we take them from the empirical distributions of the test statistics obtained from the 2000 simulations under the null hypothesis (model i). The critical value is chosen as the quantile that makes the level equal to the level of the bootstrap-based test and then that critical value is employed to obtain the power in models (ii), (iii) and (iv). We observe that for models (ii) and (iii), the new tests yield better power than the bootstrap-based tests. On the other hand, model (iv) shows an advantage in favor of the bootstrap-based tests, especially in the case of 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 view of the results of this simulation, it seems that the proposed asymptotically distribution-free tests are reasonable competitors.

5.1.2. *Parametric linear model against partially linear model.* In this section we will perform a simulation study to test for the validity of a linear model versus a partially linear model. The alternative hypothesis is therefore semiparametric. More precisely, the null hypothesis is

$$H_0 : \mu(x_1, x_2) = \theta_1 x_1 + \theta_2 x_2,$$

and the alternative hypothesis is

$$H_1 : \mu(x_1, x_2) = \theta_1 x_1 + \eta(x_2),$$

where η is a smooth nonlinear function. We consider the following three specifications of μ :

- (i) $\mu(x_1, x_2) = x_1 + x_2$;
- (ii) $\mu(x_1, x_2) = x_1 + 0.2 \sin(2\pi x_2)$;
- (iii) $\mu(x_1, x_2) = x_1 + 0.2 \exp(x_2)$.

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 two independent uniform distributions on $[0, 1]$. We work under homoscedasticity with $\sigma = 0.25$ and the error follows a standard normal distribution. Since the supports of the covariates are bounded, we take $w \equiv 1$. For the estimation of the partially linear model and the required quantities to construct the test statistics we follow the steps described in Section 3.2. Table 2 collects the results obtained from data sets with sample sizes $T = 200$ and 500 .

Under model (i), a good approximation of the level is observed for both sample sizes. In terms of power –models (ii) and (iii)–, the behaviour is as expected: larger sample sizes produce larger power, with a small advantage of KS_T over CM_T . Finally, it is worth noting that the choice of the smoothing parameter does not have relevant impact on the obtained results. The choice based on cross-validation behaves satisfactorily.

5.1.3. *Constrained mean-variance models.* In this third set of simulations 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) $Y_t = 0.2 + (1 + 0.2X_t^2)^{1/2}\varepsilon_t$;

TABLE 2

Observed rejection proportions in 2000 simulated data sets testing the linear null hypothesis $H_0 : \mu(x_1, x_2) = \theta_1 x_1 + \theta_2 x_2$ versus the semiparametric alternative $H_1 : \mu(x_1, x_2) = \theta_1 x_1 + \eta(x_2)$. Model (i) is under the null hypothesis. Models (ii) and (iii) are under the alternative hypothesis. The significance level is 0.05.

model	h	$T :$	KS_T		CM_T	
			200	500	200	500
(i)	0.15		.0480	.0560	.0590	.0550
	0.20		.0480	.0510	.0500	.0515
	0.25		.0430	.0490	.0460	.0485
	c-v		.0405	.0460	.0505	.0530
(ii)	0.15		.9680	1.000	.8625	.9985
	0.20		.9595	1.000	.8615	.9975
	0.25		.9555	1.000	.8605	.9975
	c-v		.9605	1.000	.8670	.9985
(iii)	0.15		.9530	1.000	.9185	.9995
	0.20		.9480	1.000	.9200	.9995
	0.25		.9425	.9995	.9220	.9995
	c-v		.9380	.9995	.9125	.9995

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

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

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 3 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 behaviour in terms of power is very satisfactory. The results for model (iii) show that KS_T and CM_T produce similar rejection probabilities. On the other hand, the right part of the table shows the results for the null hypothesis (b). In this case, the results for model (ii) show that although the level is overestimated for $T = 200$, the approximation is reasonably good for $T = 500$, especially for fixed values of the bandwidth. The cross-validation bandwidths still lead to an overestimation of the level. The power is excellent under both models (i) and (iii). Finally, although model (iv) is very close to the null hypothesis (b), we can see that both tests reach non-trivial power.

TABLE 3

Observed rejection proportions in 2000 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, and the sample size is $T = 200$ and $T = 500$.

model	h	$T :$	(a) $H_0 : \mu(x) = \theta x \sigma(x)$				(b) $H_0 : \mu(x) = \exp(-\theta x) \sigma(x)$			
			KS_T		CM_T		KS_{VK}		CM_{VK}	
			200	500	200	500	200	500	200	500
(i)	0.50		.0575	.0535	.0565	.0535	1.000	1.000	1.000	1.000
	0.75		.0520	.0515	.0510	.0515	1.000	1.000	1.000	1.000
	1.00		.0500	.0500	.0490	.0500	1.000	1.000	1.000	1.000
	c-v		.0565	.0475	.0545	.0480	1.000	1.000	1.000	1.000
(ii)	0.50		1.000	1.000	1.000	1.000	.0825	.0535	.0795	.0540
	0.75		1.000	1.000	1.000	1.000	.0680	.0540	.0715	.0550
	1.00		1.000	1.000	1.000	1.000	.0640	.0590	.0680	.0600
	c-v		1.000	1.000	1.000	1.000	.0725	.0710	.0700	.0715
(iii)	0.50		.7335	.9810	.6735	.9740	1.000	1.000	1.000	1.000
	0.75		.7210	.9810	.6840	.9780	1.000	1.000	1.000	1.000
	1.00		.7125	.9805	.6825	.9790	1.000	1.000	1.000	1.000
	c-v		.7170	.9830	.6755	.9780	1.000	1.000	1.000	1.000
(iv)	0.50		1.000	1.000	1.000	1.000	.2140	.2900	.1875	.2595
	0.75		1.000	1.000	1.000	1.000	.1715	.2510	.1580	.2390
	1.00		1.000	1.000	1.000	1.000	.1485	.2255	.1410	.2175
	c-v		1.000	1.000	1.000	1.000	.1850	.2655	.1600	.2420

5.2. *Real data example.* As an illustration of our methodology, we will test parametric and semiparametric restrictions for the coefficient of variation or Sharpe ratio in a real data set. We consider the monthly continuously compounded return of the Center for Research on Security Prices (CRSP) value-weighted index (including dividends) during the period 1926-2008. The response variable Y_t is the excess return constructed by subtracting to the CRSP index the monthly return on the 30-day Treasury Bill. The covariate is the one-month lagged value of the response, i.e. $X_t = Y_{t-1}$. A detailed explanation of this data set can be found in [13], where estimation of a semiparametric index model was considered. In this application we wish to test for several parametric and semiparametric specifications for the Sharpe Ratio $\mu(x)/\sigma(x)$ in this data set. We consider three specifications: (i) linear $\mu(x)/\sigma(x) = \theta x$; (ii) exponential $\mu(x)/\sigma(x) = \exp(\theta'x)$ and (iii) proportional to standard deviation $\mu(x)/\sigma(x) = \theta\sigma(x)$. The linear specification (i) is very common in applied work, see, e.g., [35]. The exponential model (ii) was considered in [4] and [1]. The proportional model (iii) is the most popular

TABLE 4

Real data application: p -values of the tests based on KS_T and CM_T to test the specifications given in H_0 . The smoothing parameters are of the form ch_{cv} , where h_{cv} is the cross-validation bandwidth.

H_0	test statistic	$c :$	0.5	1.0	1.5	2.0
$\mu(x)/\sigma(x) = \theta x$	KS_T		.0010	.0033	.0063	.0100
	CM_T		.0010	.0029	.0056	.0086
$\mu(x)/\sigma(x) = \exp(-\theta x)$	KS_T		.0000	.0000	.0000	.0000
	CM_T		.0000	.0000	.0000	.0000
$\mu(x)/\sigma(x) = \theta\sigma(x)$	KS_T		.1999	.4063	.5922	.7751
	CM_T		.2811	.4971	.6706	.8162

model, and it was introduced in the seminal work of [28, 29]. Although there exists an extensive literature using and estimating these models, see, e.g., [22, 16, 17], there are no nonparametric tests available to evaluate their goodness-of-fit.

We now apply our nonparametric tests to this data set. The sample size is 995 and the support of the covariate ranges between -0.31 and 0.67 , approximately. The tests are implemented on the $[5\%, 95\%]$ range of the covariate with a local-linear estimator for $\mu(\cdot)$ and a Nadaraya-Watson estimator for $\sigma(\cdot)$, with a common bandwidth obtained by cross-validation. The cross-validation bandwidth was $h_{cv} = 0.048$. In order to check whether the choice of the bandwidth has an effect on the results, we have also performed the tests with bandwidths of the form ch_{cv} , where $c = 0.5, 1, 1.5, 2$. The results are collected in Table 4. The reported p -values suggest that the linear and exponential specifications are rejected at 1% nominal level for all values of the bandwidth. In contrast, the tests fail to reject [28, 29]’s proportional model for all values of the bandwidth parameter. Thus, our empirical results confirm the suitability of the widely used proportional model in applied work.

6. Conclusions. In this article, we have proposed a general and simple-to-implement methodology for testing parametric or semiparametric hypotheses against semiparametric or nonparametric alternatives in regression models with possibly dependent data. The tests are based on the cumulative difference of the standardized error 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 (see, for example, the discussion in [18]). 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. [10] and [12]. In any case, as shown in our simulations the impact of the smoothing parameter on the results seems rather mild, and the cross-validation bandwidth yielded satisfactory results.

The main contribution of this article is the use of the integrated empirical process of nonparametric residuals to obtain ADF tests for the regression function. Nevertheless, this idea goes beyond inference on the regression function, and it can be used more generally for inferences where suitable tests statistics are functionals of the error distribution. For example, using a similar idea to the one used in this article one can develop ADF nonparametric goodness-of-fit tests for the error distribution in a model with nonparametric conditional mean and variance. This application follows straightforwardly from our results and is of independent interest.

Another extension would be to semiparametric hypotheses on the conditional variance, as in [7]. Our transformation does not lead to asymptotic distribution-free tests in that context, but alternative transformations may exist.

Appendix: Proofs. This appendix contains the proofs of Theorems 1 and 2. The proofs of Lemmas 5-8, Proposition 9 and Corollary 10 can be found in the on-line supplement.

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 hence $Var[(\mu(X_t) - \mu_0(X_t))/\sigma(X_t)] = 0$. We can now conclude that $(\mu(X_t) - \mu_0(X_t))/\sigma(X_t) = 0$ a.s., or, equivalently, $\mu(X_t) = \mu_0(X_t)$ a.s. \square

Proof of Theorem 2. First consider

$$\begin{aligned}
 & \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) \int_{-\infty}^y \left\{ I(\hat{\varepsilon}_t \leq s) - I(\varepsilon_t \leq s) \right\} ds \\
 = & \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) \left\{ I(\hat{\varepsilon}_t \leq y) (y - \hat{\varepsilon}_t) - I(\varepsilon_t \leq y) (y - \varepsilon_t) \right\} \\
 = & \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) I(\varepsilon_t \leq y) (\varepsilon_t - \hat{\varepsilon}_t) \\
 & + \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) \left\{ I(\hat{\varepsilon}_t \leq y) - I(\varepsilon_t \leq y) \right\} (y - \hat{\varepsilon}_t) \\
 = & A(y) + B(y) :: (\text{say}).
 \end{aligned}$$

When $\hat{\varepsilon}_t$ is replaced by $\hat{\varepsilon}_{t0}$ in the above sum, we get a similar decomposition which we denote by $A_0(y) + B_0(y)$. First, consider

$$\begin{aligned}
 & A(y) - A_0(y) \\
 = & \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) I(\varepsilon_t \leq y) (\hat{\varepsilon}_{t0} - \hat{\varepsilon}_t) \\
 = & \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) I(\varepsilon_t \leq y) \frac{\hat{\mu}(X_t) - \hat{\mu}_0(X_t)}{\hat{\sigma}(X_t)} \\
 = & \frac{1}{TE(w(X))} \sum_{t=1}^T w(X_t) I(\varepsilon_t \leq y) \frac{\hat{\mu}(X_t) - \hat{\mu}_0(X_t)}{\sigma(X_t)} + o_P(T^{-1/2}).
 \end{aligned}$$

We will now show that

$$\begin{aligned}
 & \frac{1}{TE(w(X))} \sum_{t=1}^T \left[w(X_t) I(\varepsilon_t \leq y) \frac{\hat{\mu}(X_t) - \hat{\mu}_0(X_t)}{\sigma(X_t)} \right. \\
 (14) \quad & \left. - E \left\{ w(X) I(\varepsilon \leq y) \frac{\hat{\mu}(X) - \hat{\mu}_0(X)}{\sigma(X)} \right\} \right] = o_P(T^{-1/2}),
 \end{aligned}$$

uniformly in $-\infty < y < \infty$, where the latter expectation is taken conditional on $\hat{\mu}$ and $\hat{\mu}_0$, and where the pair (X, ε) has the same distribution as (X_t, ε_t)

and is independent of $(X_1, \varepsilon_1), \dots, (X_T, \varepsilon_T)$. Define the class

$$\mathcal{F} = \left\{ (x, e) \rightarrow w(x)\sigma^{-1}(x)I(e \leq y)v(x) : -\infty < y < +\infty, v \in C_M^\tau(R_w) \right\},$$

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

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

where $\underline{\tau}$ is the largest integer strictly smaller than τ , $k = (k_1, \dots, k_d)$, $D^k = \partial^{k_1} / \partial x_1^{k_1} \dots \partial x_d^{k_d}$, and $k = \sum k_i$. Note that $P(\hat{\mu} - \hat{\mu}_0 \in C_M^\tau(R_w)) \rightarrow 1$ as T tends to infinity, by Assumptions G3(ii) and G4-(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 $u \rightarrow \beta_{[u]}$ ($[u]$ being the integer part of u , and β_t being the mixing coefficient) and Q_g is the inverse cadlag of the tail function $u \rightarrow P(\|g\| > u)$ (see Section 4.3 in [3]). 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 [36] it follows that $N_{[]}(\delta, C_M^\tau(R_w), \|\cdot\|_2) \leq \exp(K\delta^{-d/\tau})$. Moreover, defining $y_j = F_\varepsilon^{-1}(j\delta)$ for $j = 1, \dots, O(\delta^{-1})$, it is easily seen that $N_{[]}(\delta, \mathcal{F}, \|\cdot\|_{2,\beta}) = O(\delta^{-2} \exp(K\delta^{-d/\tau}))$ (see e.g. the proof of Lemma 1 in [8]). It now follows that the class \mathcal{F} is Donsker, provided $\tau > d/2$. Next, note that

$$\text{Var} \left\{ w(X)I(\varepsilon \leq y) \frac{\hat{\mu}(X) - \hat{\mu}_0(X)}{\sigma(X)} \right\} \leq K \sup_{x \in R_w} |\hat{\mu}(x) - \hat{\mu}_0(x)|^2 \xrightarrow{P} 0,$$

where the variance is conditional on $\hat{\mu}$ and $\hat{\mu}_0$, and where the pair (X, ε) is as before independent of $\hat{\mu}$ and $\hat{\mu}_0$. Hence, it follows from Corollary 2.3.12 in [36] that (14) holds true. It remains to calculate

$$\begin{aligned} & \frac{1}{E(w(X))} E \left\{ w(X)I(\varepsilon \leq y) \frac{\hat{\mu}(X) - \hat{\mu}_0(X)}{\sigma(X)} \right\} \\ &= - \frac{F_\varepsilon(y)}{E(w(X))} \int w(x) \frac{\hat{\mu}_0(x) - \mu_0(x)}{\sigma(x)} dF_X(x) \\ & \quad + \frac{F_\varepsilon(y)}{E(w(X))} \int w(x) \frac{\hat{\mu}(x) - \mu(x)}{\sigma(x)} dF_X(x), \end{aligned} \tag{15}$$

which follows from the independence between X and ε .

By Assumption G3(iii),

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

whereas by Assumption G4-(iii) under the null hypothesis

$$\int \frac{w(x)}{\sigma(x)} (\hat{\mu}_0(x) - \mu(x)) dF_X(x) = T^{-1} \sum_{t=1}^T w(X_t) l_0(Y_t, X_t) + o_P(T^{-1/2}).$$

Next, consider the term $B(y)$, which can be bounded as follows :

$$\begin{aligned} B(y) &= \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) [I(\hat{\varepsilon}_t \leq y < \varepsilon_t) (y - \hat{\varepsilon}_t) - I(\varepsilon_t \leq y < \hat{\varepsilon}_t) (y - \hat{\varepsilon}_t)] \\ &\leq \frac{1}{T\bar{w}} \sum_{t=1}^T w(X_t) [I(\hat{\varepsilon}_t \leq y < \varepsilon_t) (\varepsilon_t - \hat{\varepsilon}_t) + I(\varepsilon_t \leq y < \hat{\varepsilon}_t) (\hat{\varepsilon}_t - \varepsilon_t)]. \end{aligned}$$

Each of the above terms is $o_P(T^{-1/2})$ uniformly in y . Indeed, the first term on the right hand side is bounded by

$$\frac{1}{T\bar{w}} \max_{t:w(X_t)>0} |\hat{\varepsilon}_t - \varepsilon_t| \sum_{t=1}^T w(X_t) I(\hat{\varepsilon}_t \leq y < \varepsilon_t).$$

Note that (where the probability $P(\hat{\varepsilon} \leq y)$ is conditional on $\hat{\mu}$ and $\hat{\sigma}$, and where the pair (X, ε) is as before independent of $\hat{\mu}$ and $\hat{\sigma}$)

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T w(X_t) I(\hat{\varepsilon}_t \leq y < \varepsilon_t) \\ &= \frac{1}{T} \sum_{t=1}^T \left[w(X_t) (I(\hat{\varepsilon}_t \leq y) - I(\varepsilon_t \leq y)) - E\{w(X) (I(\hat{\varepsilon} \leq y) - I(\varepsilon \leq y))\} \right] \\ &\quad + E\left[w(X) (I(\hat{\varepsilon} \leq y) - I(\varepsilon \leq y)) \right] \\ &\leq \sup_{x \in \bar{R}_w} w(x) \left\{ \sup_y f_\varepsilon(y) \sup_{x \in \bar{R}_w} |\hat{\mu}(x) - \mu(x)| + \sup_y |y f_\varepsilon(y)| \sup_{x \in \bar{R}_w} |\hat{\sigma}(x) - \sigma(x)| \right\} \\ &\quad + o_P(T^{-1/2}) \end{aligned}$$

uniformly in y (where the term $o_P(T^{-1/2})$ in the last line follows from a combination of the above arguments for the term $A(y)$ and the proof of Lemma 1

in [8]). Hence, the first term of $B(y)$ is of the order $O_P(\sup_{x \in R_w} |\hat{\mu}(x) - \mu(x)|^2) + O_P(\sup_{x \in R_w} |\hat{\sigma}(x) - \sigma(x)|^2) + o_P(T^{-1/2}) = o_P(T^{-1/2})$ by Assumption G3-(i). The second term of $B(y)$ can be bounded in a similar way. Finally, it can be shown that $B_0(y)$ is also $o_P(T^{-1/2})$ by using Assumption G4-(i).

Therefore, by the arguments above, uniformly in y :

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

This finishes the proof. \square

REFERENCES

- [1] BEKAERT, G. and HARVEY, C.R. (1995). Time-varying world market integration. *J. Finance* **50** 403–444.
- [2] 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.
- [3] 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.
- [4] DE SANTIS, G. and GERARD, B. (1997). International asset pricing and portfolio diversification with time-varying risk. *J. Finance* **52** 1881–1912.
- [5] DELGADO, M. and GONZÁLEZ-MANTEIGA, W. (2001). Significance testing in non-parametric regression based on the bootstrap. *Ann. Statist.* **29** 1469–1507.
- [6] DETTE, H., MARCHLEWSKI, M. and WAGENER, J. (2012). Testing for a constant coefficient of variation in nonparametric regression by empirical processes. *Ann. Inst. Statist. Math.* **64** 1045–1070.
- [7] DETTE, H., NEUMEYER, N. and VAN KEILEGOM, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *J. Royal Statist. Soc. - Series B* **69** 903–917.
- [8] DETTE, H., PARDO-FERNÁNDEZ, J.C. and VAN KEILEGOM, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scand. J. Statist.* **36** 782–799.
- [9] DOUKHAN, P. (1994). *Mixing. Properties and Examples*, Springer, Lecture Notes in Statistics.
- [10] EINMAHL, J.H.J. and MASON, M.D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403.
- [11] ESCANCIANO, J.C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory* **25** 162–194.
- [12] 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. *J. Econometrics* **178** 426–443.

- [13] ESCANCIANO, J.C., PARDO-FERNÁNDEZ, J.C. and VAN KEILEGOM, I. (2017). Semi-parametric estimation of risk-return relationships. *J. Bus. Econ. Statistics* **35** 40–52.
- [14] FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660.
- [15] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods*. Springer, New York.
- [16] FERSON, W. (1989). Changes in expected security returns, risk and level of interest rates. *J. Finance* **44** 1191–1217.
- [17] FERSON, W., FOERSTER, S.R. and KEIM, D.B. (1993). General tests of latent variable models and mean-variance spanning. *J. Finance* **48** 131–156.
- [18] GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R.M. (2013). An updated review of goodness-of-fit tests for regression models. *Test* **22** 361–411.
- [19] HANSEN, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748.
- [20] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single index models. *Ann. Statist.* **21** 157–178.
- [21] HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.
- [22] HARVEY, C.R. (1989). Time-varying conditional covariances in tests of asset pricing models. *J. Fin. Econ.* **24** 289–317.
- [23] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253.
- [24] KHMALADZE, E.V. and KOUL, H.L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034.
- [25] KOUL, H.L. and STUTE, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27** 204–236.
- [26] LEE, T.H., TU, Y. and ULLAH, A. (2015). Forecasting equity premium: global historical average versus local historical average and constraints. *J. Bus. Econ. Statist.* **33** 393–402.
- [27] MASRY, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Series Anal.* **17** 571–599.
- [28] MERTON, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica* **41** 867–887.
- [29] MERTON, R.C. (1980). On estimating the expected return on the market. An explanatory investigation. *J. Fin. Econ.* **8** 323–361.
- [30] NEUMEYER, N. and VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multiv. Anal.* **101** 1067–1078.
- [31] NEWEY, W.K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382.
- [32] ROBINSON, P.M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954.
- [33] STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641.
- [34] STUTE, W., XU, W.L. and ZHU, L.X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika* **95** 451–467.

- [35] TANG, Y. and WHITELOW, R.F. (2011). Time-varying Sharpe ratios and market timing. *Quarterly J. Finance* **1** 465–493.
- [36] VAN DER VAART, A.W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer.
- [37] 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.

DEPARTMENT OF ECONOMICS
INDIANA UNIVERSITY
WYLIE HALL, 100 S WOODLAWN AVE
BLOOMINGTON, IN 47405, USA
E-MAIL: jescanci@indiana.edu

DEPARTAMENTO DE ESTATÍSTICA E IO
UNIVERSIDADE DE VIGO
CAMPUS UNIVERSITARIO AS LAGOAS-MARCOSENDE
36310 VIGO, SPAIN
E-MAIL: juanpc@uvigo.es

ORSTAT
KU LEUVEN
NAAMSESTRAAT 69
3000 LEUVEN, BELGIUM
E-MAIL: ingrid.vankeilegom@kuleuven.be