Econometric Theory, **26**, 2010, 744–773.
doi:10.1017/S0266466609990090

ASYMPTOTIC DISTRIBUTION-FREE DIAGNOSTIC TESTS FOR HETEROSKEDASTIC TIME SERIES MODELS

J. CARLOS ESCANCIANO
Indiana University

This article investigates model checks for a class of possibly nonlinear heteroskedastic time series models, including but not restricted to ARMA-GARCH models. We propose omnibus tests based on functionals of certain weighted standardized residual empirical processes. The new tests are asymptotically distribution-free, suitable when the conditioning set is infinite-dimensional, and consistent against a class of Pitman's local alternatives converging at the parametric rate $n^{-1/2}$, with n the sample size. A Monte Carlo study shows that the simulated level of the proposed tests is close to the asymptotic level already for moderate sample sizes and that tests have a satisfactory power performance. Finally, we illustrate our methodology with an application to the well-known S&P 500 daily stock index. The paper also contains an asymptotic uniform expansion for weighted residual empirical processes when initial conditions are considered, a result of independent interest.

1. INTRODUCTION

The correct specification of conditional mean and variance models is of major interest in empirical finance and economics. Model-based economic theories in dynamic contexts, such as asset pricing, portfolio choice, and market risk management, as well as many inference procedures proposed in the econometrics literature, such as the consistency of the quasi-maximum likelihood estimator (QMLE), depend crucially on the correct joint specification of the conditional mean and variance. Despite the significant empirical and theoretical importance of this testing problem, rather few joint specification tests are available in the literature. Furthermore, the existing tests have model-dependent asymptotic distributions and may require bootstrap methods that are difficult to implement in standard econometric packages. The aim of this paper is then to construct simple

Research was funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908, and by the Spanish Ministerio de Educación y Ciencia, reference number SEJ2005-07657/ECON. I would like to thank Miguel A. Delgado, Oliver Linton, Carlos Velasco, and two anonymous referees for helpful comments. I also thank Wenceslao González-Manteiga for pointing out an important reference. All errors are mine. Address correspondence to J. Carlos Escanciano, Department of Economics, Indiana University; 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA; e-mail: jescanci@indiana.edu.

 © Cambridge University Press 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

asymptotic distribution-free (ADF) joint specification tests, especially convenient for financial and economic applications. The new tests are based on certain weighted standardized residual empirical processes with weights equal to proper transformations of the conditioning variables. The weights are chosen orthogonal to the model's scores, so ADF omnibus tests are obtained, with the added property of being flexible to tailor the power against desired directions.

To be more precise, let $\{Y_t\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$. A large body of the time series modeling literature, both theoretical and empirical, has considered the model

$$Y_t = f(I_{t-1}, \theta_0) + h(I_{t-1}, \theta_0)u_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where $I_{t-1} = (Y_{t-1}, u_{t-1}, Y_{t-2}, u_{t-2}, \ldots)'$ is the conditioning set at time $t-1$, and $f(I_{t-1}, \theta_0)$ and $h^2(I_{t-1}, \theta_0)$ are parametric specifications for $f(y) := \mathrm{E}[Y_t \mid I_{t-1} = y]$ almost surely (a.s.) and $h^2(y) := \mathrm{Var}[Y_t \mid I_{t-1} = y]$ a.s. In (1), θ_0 is an unknown parameter in a Euclidean space, $\theta_0 \in \Theta \subset \mathbb{R}^p$, and $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) disturbances, with unknown cumulative distribution function (cdf) F_u. Throughout the paper we assume that u_t is independent of I_{t-1}, for all $t \in \mathbb{Z}$. The specification (1) covers the well-known linear ARMA-GARCH models, as well as nonlinear conditional mean and variance models, see, e.g., Fan and Yao (2003). Our theory is valid for the general heteroskedastic time series model described in (1).

We aim to test under this setup for $H_0 : f(I_{t-1}) = f(I_{t-1}, \theta_0)$ and $h(I_{t-1}) = h(I_{t-1}, \theta_0)$ a.s. for some $\theta_0 \in \Theta$, against nonparametric or parametric alternatives given by the negation of H_0. This testing problem has important implications in econometrics practice and, in particular, in financial econometrics modeling, where parametric models such as (1) are commonly used. An example of the application is to the value at risk (VaR) methodology in market risk management; see Tsay (2005). The VaR at level α in model (1) is given by $VaR_t(\alpha) = f(I_{t-1}, \theta_0) + h(I_{t-1}, \theta_0)F_u^{-1}(\alpha)$, where $F_u^{-1}(\alpha)$ is the standardized error's quantile function evaluated at α. Clearly, the functional forms for f and h play a crucial role in determining the $VaR_t(\alpha)$, whereas θ_0 and $F_u^{-1}(\alpha)$ are nuisance parameters that can be estimated by many $\sqrt{n}-$consistent estimators, e.g., QMLE for θ_0 and quantile estimators for $F_u^{-1}(\alpha)$; see, e.g., Koenker and Zhao (1996).

More generally, model specification within the class (1) is important mainly because a lack of fit in the postulated conditional mean and/or variance can lead to misleading conclusions in statistical inferences, and to suboptimal point forecasts. Therefore, in order to prevent wrong conclusions, every statistical inference that is based on the model $\mathcal{M} = \{f(\cdot, \theta), h(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ should be accompanied by a proper joint diagnostic test, i.e., a test for H_0.

There is a vast amount of literature on testing the specification of parametric dynamic conditional mean models; see Escanciano (2006) and references therein. However, the literature on joint model checks for the conditional mean and

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

variance functions is rather scarce. Classical diagnostic tools for testing H_0 are the Portmanteau tests initially proposed by Box and Pierce (1970) and Ljung and Box (1978), and subsequently extended to some conditional variance models by Li and Mak (1994); see also Lundbergh and Teräsvirta (2002) and Hidalgo and Zaffaroni (2007). Tests based on usual correlation or autocorrelation measures of errors (centered squared errors) are not consistent in any misspecified model that leads to uncorrelated errors (centered squared errors), increasing the Type II error probability. Ngatchou-Wandji (2005) proposed using χ^2-discrepancy measures for testing H_0, whereas Gao and King (2004) have extended the initial smoothed-based approach of Härdle and Mammen (1993) to tests for H_0.

An important limitation of most of the aforementioned articles is that they consider a finite-dimensional information set I_{t-1}, and hence, they are not suitable for testing H_0 here. Moreover, even for the case in which the information set is of finite dimension, d say, most of the proposed tests lead to a poor power performance when d is large or moderate, due to the so-called "curse of dimensionality" problem. Escanciano (2008) overcomes the latter problem using a generalized spectral approach, but resorts to the assistance of a bootstrap approximation that is difficult to implement in standard econometric packages. Alternatively, Hong and Lee (2003) consider a smoothed approach based on the generalized spectral density, which requires a bandwidth choice, without much guidance for its choice for testing purposes.

This article deviates from previous works by proposing a large class of simple ADF tests for H_0 that are robust to the dimension in the information set I_{t-1}. The rationale for our tests follows from the asymptotic properties of the weighted empirical processes of standardized residuals

$$R^1_{n,\widehat{g}}(x) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{g}_c(\widehat{I}_{t-1}, \theta_n) 1(\widehat{u}_t \leq x), \tag{2}$$

where $\widehat{g}_c(\widehat{I}_{t-1}, \theta_n)$ is an estimator for $g_c(I_{t-1}, \theta_0)$, a suitable transformation of the conditioning set that will be specified later on; $\widehat{u}_t$ are standardized residuals obtained from (1) and computed as

$$\widehat{u}_t \equiv \widehat{u}_t(\theta_n) = \frac{Y_t - m(\widehat{I}_{t-1}, \theta_n)}{h(\widehat{I}_{t-1}, \theta_n)}, \qquad 1 \leq t \leq n,$$

where θ_n is a $\sqrt{n}$-consistent estimators for θ_0, e.g., the QMLE; and $\widehat{I}_{t-1}$ is the information set observed at time $t-1$ that contains $(Y_{t-1}, Y_{t-2}, \ldots, Y_0)$ and that may contain some other initial values; see Section 3.

As we shall show, the asymptotic behavior of the process $R^1_{n,\widehat{g}}$ will depend on whether H_0 holds or not. This fact will allow us to base omnibus tests on suitable functionals of $R^1_{n,\widehat{g}}$. We choose the weights $\widehat{g}_c(\widehat{I}_{t-1}, \theta_n)$ in (2) in such a way that simple ADF tests are obtained, avoiding much of the burden of bootstrap approximations or complicated martingale transforms.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Our tests can be seen as generalizations of the tests proposed by Wooldridge (1990). This author considers tests based on

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{g}_c(I_{t-1},\theta_n)\widehat{u}_i \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{g}_c(I_{t-1},\theta_n)\widehat{u}_i^2. \tag{3}$$

So the sample means in (3) contain less "information" than the process $R^1_{n,\widehat{g}}$ itself, resulting in generally less powerful testing procedures; see Section 2.2 for discussion. Note that both quantities in (3) are empirical integrals of $R^1_{n,\hat{g}}$, e.g.,

$$\int_{\mathbb{R}} x R^1_{n,\widehat{g}}(dx) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{g}_c(X_i,\theta_n)\widehat{u}_i.$$

For further comparison of our tests with Wooldridge's (1990) tests, see Example 2 below.

Weighted residual empirical processes have been used in the literature for different testing problems than H_0. Koul and Ling (2006) proposed goodness of fit tests for the error distribution in (1), assuming the conditional mean and variance are correctly specified. Thus, our tests here can be used as a pre-test in Koul and Ling's (2006) goodness of fit test. Also, we stress that in practice the information set I_{t-1} is not observable and must be estimated by $\widehat{I}_{t-1}$. Koul and Ling (2006) did not take into account this estimation of initial values, and unfortunately this makes the application of their results difficult; see the discussion in the next section. Therefore, in order to carry out our program, we need to extend a weak convergence result in Koul and Ling (2006, Lem. 4.1; see also Bai and Ng, 2001, Lem. A.6) to allow for an estimated conditioning set. This extension is not trivial and is of independent interest.

We also note that after the first version of this paper was written, a recent paper by Stute, Xu, and Zhu (2008) came to our attention[1] using weighted residual processes for testing the correct specification of a homoskedastic regression in an i.i.d. setup. These authors developed a principal components decomposition that can be also used in the present framework. Details are omitted to save space. Nevertheless, the testing framework in Stute et al. (2008) is different from ours and uses different tools.

The remainder of this paper is organized as follows: In Section 2 we study the asymptotic theory for a class of weighted residual empirical processes and introduce the tests. In Section 3 we illustrate the applicability of our methods and show that our assumptions are naturally satisfied by the class of ARMA-GARCH models. Section 4 provides evidence of the good finite-sample performance of the proposed tests and apply the new methods to the S&P 500 daily stock index. Section 5 concludes and discusses future extensions. In the Appendix we establish an asymptotic uniform expansion for a weighted residual empirical processes when initial conditions are considered, a result of independent interest. Mathematical proofs are gathered in the Appendix.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

2. JOINT MODEL CHECKS FOR MEAN-VARIANCE MODELS

2.1. Asymptotic Null Distribution

This section introduces omnibus ADF tests for H_0. For the sake of exposition, we shall first consider the limit behavior of the process

$$R^1_{n,g}(x) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_c(\widehat{I}_{t-1},\theta_n)1(\widehat{u}_t \le x)$$

for a centered weight $g_c(I_{t-1},\theta) = g(I_{t-1},\theta) - \overline{g}(\theta)$, $\overline{g}(\theta) = n^{-1}\sum_{t=1}^{n} g(I_{t-1},\theta)$, where $g(I_{t-1},\theta)$ is a generic real-valued measurable transformation of I_{t-1} that is smooth in θ (cf. Assump. A5). Henceforth, the subindexes c and e stand for sample and population centering, respectively, e.g., $g_e(I_{t-1},\theta) := g(I_{t-1},\theta) - \mathrm{E}[g(I_{t-1},\theta)]$.

We introduce some regularity conditions. For any differentiable function g on Θ, let $\dot{g}$ denote its differential. Throughout, A' denotes transposition of the matrix A and Θ_0 denotes a small convex neighborhood of θ_0. For positive finite constants $b > 0$ and $c > 0$, define the set $\mathcal{Z}_{b,c} := \{(z_1,z_2)' \in \mathbb{R}^2 : |z_1| + |z_2| \le b, z_1 \ge -1 + c\}$.

Assumption A1.

A1(a): $\{Y_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process.

A1(b): F_u has an absolutely continuous density f_u such that, for all $0 < b < \infty$ and $0 < c < \infty$,

$$\sup_{x\in\mathbb{R},(z_1,z_2)'\in\mathcal{Z}_{b,c}} |x| f_u(x + xz_1 + z_2) < \infty,$$

$$\sup_{x\in\mathbb{R},\gamma\ge -1+c} \left| x^j f_u(x(1+\gamma)) - x^j f_u(x)\right| \le C\gamma \text{ for } j = 0, 1.$$

Assumption A2. The functions $f(I_{t-1},\cdot)$ and $h(I_{t-1},\cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with derivatives $\dot{f}_t(\theta)$ and $\dot{h}_t(\theta)$, respectively, such that

$$\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{f}_t(\theta)}{h(I_{t-1},\theta_0)}\right|^2\right] < \infty, \qquad \mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}_t(\theta)}{h(I_{t-1},\theta_0)}\right|^2\right] < \infty,$$

$$n^{-1/2}\max_{1\le t\le n}\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{f}(\widehat{I}_{t-1},\theta)}{h(I_{t-1},\theta_0)}\right|\right] = O(1),$$

$$n^{-1/2}\max_{1\le t\le n}\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}(\widehat{I}_{t-1},\theta)}{h(I_{t-1},\theta_0)}\right|\right] = O(1),$$

and

$$n^{-1/2}\max_{1\le t\le n}\sup_{\theta\in\Theta_0}\left|\frac{\dot{f}_t(\theta)}{h(I_{t-1},\theta_0)}\right| = o_P(1), \qquad n^{-1/2}\max_{1\le t\le n}\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}_t(\theta)}{h(I_{t-1},\theta_0)}\right| = o_P(1).$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

In addition, $\min_{t\in\mathbb{Z}} \inf_{\theta\in\Theta_0} h(I_{t-1},\theta) \geq c > 0$ and

$$\max_{1\leq t\leq n} \sup_{\theta\in\Theta_0} \left| \frac{h(I_{t-1},\theta_0)}{h(\widehat{I}_{t-1},\theta)} \right| = O_P(1).$$

Assumption A3. The parametric space Θ is compact in $\mathbb{R}^p$. The parameter θ_0 belongs to the interior of Θ. Under the null hypothesis, there exists a unique $\theta_0 \in \Theta$ such that $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$.

Assumption A4. The observed information set available at period $t-1$, $\widehat{I}_{t-1}$, may contain some assumed initial values and satisfies

$$\lim_{n\to\infty} \sum_{t=1}^{n} \mathrm{E}^{1/2} \left\{ \sup_{\theta\in\Theta_0} \left| \frac{f(\widehat{I}_{t-1},\theta) - f(I_{t-1},\theta)}{h(I_{t-1})} \right|^2 \right\} < \infty,$$

$$\lim_{n\to\infty} \sum_{t=1}^{n} \mathrm{E}^{1/2} \left\{ \sup_{\theta\in\Theta_0} \left| \frac{h(\widehat{I}_{t-1},\theta) - h(I_{t-1},\theta)}{h(I_{t-1})} \right|^2 \right\} < \infty,$$

and

$$n^{-1/2} \sum_{t=1}^{n} \mathrm{E} \sup_{\theta\in\Theta_0} \left| g(\widehat{I}_{t-1},\theta) - g(I_{t-1},\theta) \right| = o_P(1).$$

Assumption A5. The function $g(I_{t-1},\cdot)$ satisfies $\mathrm{E}\left[\sup_{\theta\in\Theta_0} |g(I_{t-1},\theta)|^4\right] < \infty$ and is continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with derivative $\dot{g}_t(\theta)$ such that $\mathrm{E}\left[\sup_{\theta\in\Theta_0} |\dot{g}_t(\theta)|^2\right] < \infty$.

Assumption A1 is a condition on the DGP. Assumption A2 is on the model and is standard in the literature; see, e.g., Koul and Ling (2006). Assumption A3 is satisfied under mild conditions for the nonlinear least squares estimator (or its robust modifications, under further regularity assumptions) or for the QMLE, see Koul (2002, Chs. 5 and 8), Hall and Heyde (1980, Ch. 6), Francq and Zakoïan (2004), Straumann (2005), and Robinson and Zaffaroni (2006), to mention but a few. Assumption A4 is a start-up value condition. It ensures that the impacts of initial values are asymptotically negligible. This condition easily holds for many time series models; see Francq and Zakoïan (2004) for ARMA-GARCH models. We remark that Lemma 4.1 in Koul and Ling (2006) requires

$$\max_{1\leq t\leq n} \left| \frac{h(\widehat{I}_{t-1},\theta_0) - h(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)} \right| = o_P(1),$$

$$\max_{1\leq t\leq n} \left| \frac{m(\widehat{I}_{t-1},\theta_0) - m(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)} \right| = o_P(1),$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

which is difficult to verify and may not hold even for simple models such as ARMA-GARCH. For instance, in a homoskedastic (i.e., $h(I_{t-1},\theta_0)=\sigma$) MA(1) model with estimated information set $\widehat{I}_{t-1}=\{Y_{t-1},\ldots,Y_1,\widehat{u}_0\}$, where $\widehat{u}_0$ is some assumed value for u_0, the conditions in A4 here hold, whereas, if $|\theta_0|<1$,

$$\max_{1\le t\le n}\left|\frac{f(\widehat{I}_{t-1},\theta_0)-f(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)}\right|=\frac{1}{\sigma}\left|\widehat{u}_0-u_0\right||\theta_0|^t,$$

which does not converge to zero in probability. For this reason, we extend their Lemma 4.1 in the Appendix below. This extension is not trivial and has applications beyond the present problem. Finally, Assumption A5 is on the weight function g and can be easily checked.

Define

$$b(x,g,\theta_0):=\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{f}_t(\theta_0)}{h(I_{t-1})}\right]f_u(x)+\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{h}_t(\theta_0)}{h(I_{t-1})}\right]xf_u(x).$$

It is important for our purposes to note that if the moment conditions

$$\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{f}_t(\theta_0)}{h(I_{t-1})}\right]=0 \tag{4}$$

and

$$\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{h}_t(\theta_0)}{h(I_{t-1})}\right]=0 \tag{5}$$

hold, then $b(x,g,\theta_0)\equiv 0$. We are now in position to establish the limit null distribution of $R^1_{n,g}$.

THEOREM 1.

(i) Under Assumptions A1–A5 and H_0, uniformly in $x\in\mathbb{R}$,

$$R^1_{n,g}(x)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_e(I_{t-1},\theta_0)[1(u_t\le x)-F_u(x)]$$
$$+\sqrt{n}(\theta_n-\theta_0)'b(x,g,\theta_0)+o_P(1).$$

(ii) If, in addition, (4) and (5) hold, then, uniformly in $x\in\mathbb{R}$,

$$R^1_{n,g}(x)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_e(I_{t-1},\theta_0)[1(u_t\le x)-F_u(x)]+o_P(1).$$

A corollary of Theorem 1(ii) is that, provided the moment conditions (4) and (5) hold, the null limit distribution of tests based on $R^1_{n,g}$ will be ADF (cf. Corollary 1). A natural question is then how we construct weights satisfying (4) and (5).

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

We address this problem in the next lines. To simplify notation, write $a_{1,t}(\theta) = \dot{f}_t(\theta)/h(I_{t-1},\theta)$, $a_{2,t}(\theta) = \dot{h}_t(\theta)/h(I_{t-1},\theta)$, and

$$X_t(\theta) = (1, a'_{1,t}(\theta), a'_{2,t}(\theta))' \qquad \theta \in \Theta,\ t = 1,\dots,n.$$

Suppose we consider an initial weight $g_0(I_{t-1})$. Then, we proceed by taking as $g_c(\widehat{I}_{t-1},\theta_n)$ the residuals from the least squares regression (provided no exact collinearity exists, otherwise remove the necessary regressors) in

$$g_0(\widehat{I}_{t-1}) = \gamma'\widehat{X}_t(\theta_n) + \xi_t \qquad t = 1,\dots,n, \tag{6}$$

where $\widehat{X}_t(\theta_n)$ uses $\widehat{I}_{t-1}$ instead of I_{t-1}. The initial weight function g_0 is up to the econometrician and gives flexibility to direct the power against desired directions; see the local power analysis in Section 2.1. The least squares estimator in (6) is

$$\widehat{\gamma}_n(\theta_n) = \left(\sum_{t=1}^n \widehat{X}_t(\theta_n)\widehat{X}'_t(\theta_n)\right)^{-1}\left(\sum_{t=1}^n \widehat{X}_t(\theta_n) g_0(\widehat{I}_{t-1})\right).$$

The estimator $\widehat{\gamma}_n(\theta_n)$ estimates $\gamma \equiv \gamma(\theta_0) = (\mathrm{E}[X_t(\theta_0)X'_t(\theta_0)])^{-1}\mathrm{E}[X_t(\theta_0)g_0(I_{t-1})]$, and the weight

$$g(I_{t-1},\theta_0) = g_0(I_{t-1}) - \gamma'(\theta_0)X_t(\theta_0) \tag{7}$$

satisfies (4) and (5) and has zero mean, by construction. The function

$$\widehat{g}(\widehat{I}_{t-1},\theta_n) = g_0(\widehat{I}_{t-1}) - \widehat{\gamma}'_n(\theta_n)\widehat{X}_t(\theta_n) \tag{8}$$

estimates $g(I_{t-1},\theta_0)$ in (7). Finally, our tests are functionals of the weighted residual empirical process

$$R^1_{n,\widehat{g}}(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \widehat{g}(\widehat{I}_{t-1},\theta_n)1(\widehat{u}_t \le x).$$

We introduce regularity conditions that are sufficient for Assumption A2 for the present choice of $\widehat{g}(\widehat{I}_{t-1},\theta_n)$.

Assumption A6. The functions $f(I_{t-1},\cdot)$ and $h(I_{t-1},\cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with derivatives $\dot{f}_t(\theta)$ and $\dot{h}_t(\theta))$ such that

A6(a): $\mathrm{E}\left[\sup_{\theta\in\Theta_0}|X_t(\theta)|^4\right] < \infty$ and $I(\theta) = \mathrm{E}\left[X_t(\theta)X'_t(\theta)\right]$ is positive definite on Θ_0.

A6(b): The function $X_t(\theta)$ is continuously differentiable with respect to $\theta \in \Theta_0$ a.s., with derivative $\dot{X}_t(\theta)$ such that $\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\dot{X}_t(\theta)\right|^2\right] < \infty$.

A6(c): $\mathrm{E}\left[g_0^4(I_{t-1})\right] < \infty$ and $n^{-1/2}\sum_{t=1}^n \left|g_0(\widehat{I}_{t-1}) - g_0(I_{t-1})\right| = o_P(1)$.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Now from Theorem 1 we can construct ADF tests for H_0 as follows: Define $\sigma^2 := \mathrm{E}\left[g_e^2(I_{t-1}, \theta_0)\right]$. The symbol $\Longrightarrow$ denotes weak convergence in the metric space $D([-\infty, \infty])$ of *cadlag* (right-continuous with left limits) functions on $[-\infty, \infty]$, endowed with the Skorohod metric; see Billingsley (1999). Notice that $R^1_{n,\widehat{g}}$ belongs to $D([-\infty, \infty])$, after defining $R^1_{n,\widehat{g}}(-\infty) = 0 = R^1_{n,\widehat{g}}(+\infty)$.

COROLLARY 1. *Under Assumptions A1, A3, A4, and A6 and* H_0,

$$R^1_{n,\widehat{g}}(\cdot) \Longrightarrow \sigma B(F_u(\cdot)),$$

where $B(\cdot)$ *is a standard Brownian bridge on [0,1].*

An application of the continuous mapping theorem (CMT) (see, e.g., Thm. 1.3.6 in van der Vaart and Wellner, 1996) and Lemma 3.1 in Chang (1990) yields

$$CvM_n := \int_{\mathbb{R}} \widehat{\sigma}^{-2} \left| R^1_{n,\widehat{g}}(x) \right|^2 F_{n,u}(dx) \xrightarrow{d} \int_{[0,1]} |B(x)|^2 dx,$$

and

$$KS_n = \sup_{x \in \mathbb{R}} \left| \widehat{\sigma}^{-1} R^1_{n,\widehat{g}}(x) \right| \xrightarrow{d} \sup_{u \in [0,1]} |B(u)|,$$

where $\widehat{\sigma}^2 = n^{-1} \sum_{t=1}^n \widehat{g}^2(\widehat{I}_{t-1}, \theta_n)$ estimates σ^2 and $F_{n,u}(\cdot)$ is the empirical cdf of the standardized residuals $\{\widehat{u}_t\}_{t=1}^n$.

The test statistics CvM_n and KS_n are very easy to compute, as $R^1_{n,\widehat{g}}$ takes at most n different values, which can be computed recursively by ordering the residuals. More concretely, denote by $\widehat{u}_{(1)} \leq \widehat{u}_{(2)} \leq \cdots \leq \widehat{u}_{(n)}$ the ordered residuals and by $\widehat{g}_{c(1)}, \ldots, \widehat{g}_{c(n)}$ the corresponding concomitants, i.e., $\widehat{g}_{c(j)} = \widehat{g}(\widehat{I}_{(j)-1}, \theta_n)$, $1 \leq j \leq n$. Then it is straightforward to prove that

$$CvM_n = \frac{1}{\widehat{\sigma}^2 n^2} \sum_{t=1}^n \left(\sum_{j=1}^t \widehat{g}_{c(j)} \right)^2$$

and

$$KS_n = \max_{1 \leq t \leq n} \left| \frac{1}{\widehat{\sigma}\sqrt{n}} \sum_{j=1}^t \widehat{g}_{c(j)} \right|.$$

Our new tests have well-known asymptotic distributions with critical values that are readily available; see, e.g., Shorack and Wellner (1986). In particular, our Cramér-von Mises (CvM) and Kolmogorov (K) tests reject the null hypothesis at 5% if $CvM_n > 0.46136$ and $KS_n > 1.3581$, respectively.

2.2. Power Properties: Choice of g_0

In this section we discuss global and local power properties of tests based on $R^1_{n,\widehat{g}}$. Proofs of the main results in this section are omitted, as they follow from similar

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

arguments to those of Theorem 1. They are available from the author upon request. Assume that under the alternative hypothesis there exists a pseudo-true value $\theta_* \in \Theta$ such that $|\theta_n - \theta_*| = o_P(1)$ and define $u_t(\theta_*) = (Y_t - f(I_{t-1}, \theta_*))/h(I_{t-1}, \theta_*)$

$$d_{1t*} := \frac{h(I_{t-1}, \theta_*) - h(I_{t-1})}{h(I_{t-1})} \quad \text{and} \quad d_{2t*} := \frac{f(I_{t-1}, \theta_*) - f(I_{t-1})}{h(I_{t-1})}.$$

Note that under the alternative, $d_{1t*} \neq 0$ and $d_{2t*} \neq 0$ with positive probability. Similarly to the proof of Theorem 1, it can be shown that under the alternative, uniformly in $x \in \mathbb{R}$,

$$n^{-1/2} R^1_{n,g}(x) \xrightarrow{P} \Delta_*(x) := \mathrm{E}[g_e(I_{t-1}, \theta_*) 1(u_t(\theta_*) \leq x)].$$

Thus, whether or not tests based on $R^1_{n,\widehat{g}}$ are consistent against fixed alternatives depends in turn on whether or not the function $\Delta_*(x)$ is different from the zero function. Note that $\Delta_*(x) = \mathrm{E}[g_e(I_{t-1}, \theta_*) F_u(x + x d_{1t*} + d_{2t*})]$. It turns out that the function $\Delta_*(x)$ depends on the weight $g_e(I_{t-1}, \theta_*)$, on the probabilistic limit θ_*, the specified and true models, through d_{1t*} and d_{2t*}, and on the unknown cdf F_u. Hence, the function $\Delta_*(x)$ is in general unknown, which makes it difficult to study the global power properties of tests.

The following example[2] illustrates that the global consistency will depend in general on the choice of $g_e(I_{t-1}, \theta_*)$.

Example 1

Consider a Gaussian AR(1)-ARCH(1) model with mean and variance functions

$$\mathrm{E}[Y_t \mid I_{t-1}] = 0.5 Y_{t-1}, \qquad \mathrm{Var}[Y_t \mid I_{t-1}] = 1 + 0.5 Y^2_{t-1},$$

respectively, and Gaussian innovations $u_t \sim N(0, 1)$. The econometrician specifies model (1) with $f(I_{t-1}, \theta_0) = \theta_{01} Y_{t-1}$ and $h(I_{t-1}, \theta_0) = \theta_{02}$, $\theta_0 = (\theta_{01}, \theta_{02})' \in \Theta \subset (-1, 1) \times (0, \infty)$, and proceeds to estimate θ_0 by the ordinary least squares (OLS) estimator $\theta_n = (\theta_{n1}, \theta_{n2})'$. Under mild moment conditions, $|\theta_n - \theta_*| = o_P(1)$ with $\theta_* = (0.5, \sqrt{2})$. For this example,

$$\Delta_*(x) = \mathrm{E}\left[g_e(I_{t-1}, \theta_*) \Phi\left(x \frac{\sqrt{2}}{\sqrt{1 + 0.5 Y^2_{t-1}}} \right) \right],$$

where, henceforth, Φ denotes the standard Gaussian cdf. Note that $X_t(\theta_0) = (1, Y_{t-1}/\theta_{02})'$.Therefore, if we choose $g_0(I_{t-1}) = Y^3_{t-1}$, then $g_e(I_{t-1}, \theta_*) = Y^3_{t-1} - \gamma_1 Y_{t-1}$,for a constant γ_1 that is not relevant for the following conclusions. Note that by symmetry,

$$\mathrm{E}[g_e(I_{t-1}, \theta_*) \mid Y^2_{t-1}] = 0,$$

which implies that $\Delta_*(x)$ will be identically zero, and the test will have no power. Other choices of g_0 for this example lead to consistent tests, such as $g_0(I_{t-1}) = Y^2_{t-1}$; see our Monte Carlo experiments below.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Thus, this example shows that an unfortunate choice for g_0 may lead to inconsistency of the new tests. This problem is not specific to our tests and applies to other specification tests available in the literature. In particular, this drawback also appears in the tests by Koul and Ling (2006) and Stute et al. (2008). This possible inconsistency is the price we pay for the ADF property and simplicity of our tests, and it can be mitigated by making the tests more complicated and losing the asymptotic pivotal property of the tests; see Section 5 for further discussion. Nevertheless, our Monte Carlo experiments below and other unreported simulations show that for the alternatives considered and several choices of g_0, our tests have a satisfactory empirical power; see Section 4.

We now provide a detailed comparison with Wooldridge's (1990) test. It can be shown that Wooldridge's test based on (3) has global power provided

$$\int y\Delta_*(dy) \neq 0 \quad \text{or} \quad \int y^2\Delta_*(dy) \neq 0.$$

Hence, we conclude that Wooldridge's (1990) test has no global power whenever our test has no global power (i.e., if $\Delta_*(y) \equiv 0$). The reciprocal is not true, as there are examples where our test has nontrivial power but Wooldridge's (1990) test has not. The following example illustrates this point; see also our application to the S&P 500.

Example 2

Consider a Gaussian AR(1) model with

$$\mathrm{E}[Y_t \mid I_{t-1}] = \frac{1}{\sqrt{2}}Y_{t-1}, \qquad \mathrm{Var}[Y_t \mid I_{t-1}] = 1,$$

respectively, and Gaussian innovations $u_t \sim N(0,1)$. The econometrician specifies model (1) with $f(I_{t-1},\theta_0) = \theta_0$ and $h(I_{t-1},\theta_0) = 1$, and proceeds to estimate θ_0 by the sample mean θ_n. Under mild moment conditions, $|\theta_n - \theta_*| = o_P(1)$ with $\theta_* = 0$. For this example,

$$\Delta_*(x) = \mathrm{E}\left[g_e(I_{t-1},\theta_*)\Phi\left(x - 2^{-1/2}Y_{t-1}\right)\right].$$

Note that $X_t(\theta_0) = 1$. Therefore, if we choose $g_0(I_{t-1}) = Y_{t-1}^2$, then $g_e(I_{t-1},\theta_*) = Y_{t-1}^2 - 2$. It is straightforward to show that, although $\mathrm{E}[g_e(I_{t-1},\theta_*)Y_t] = 0$, the function $\Delta_*(x)$ is different from the zero function. Wooldridge's (1990) test will have no power against this alternative, whereas our tests are consistent.

Of course, we do not claim in Example 2 that our tests are always better than Wooldridge's (1990) tests; they are complementary rather than competing tests. Moreover, strictly speaking they cannot be compared, since they are valid under different sets of assumptions. Wooldridge's (1990) tests are robust to deviations of the i.i.d. assumption in the innovations, whereas our tests are not. However, being based on a *finite* number of functions of residuals (see (3)), Wooldridge's

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

(1990) tests are not consistent against a large class of alternatives to H_0 that can be picked up by our tests, as they consider an infinite number of transformations of residuals indexed by $x \in \mathbb{R}$.

We now turn to the study of the local power of our tests. This study will shed some light on the problem of the choice of g_0. We shall show that the choice of g_0 allows us to construct ADF tests with good local power against desired alternatives, e.g., threshold alternatives. Consider the following local alternatives within model (1):

$$H_{1n} : f(y) = f(y, \theta_0) + n^{-1/2} s_m(y) \quad \text{and} \quad h(y) = h(y, \theta_0), \tag{9}$$

and

$$H_{2n} : f(y) = f(y, \theta_0) \quad \text{and} \quad h(y) = h(y, \theta_0) + n^{-1/2} s_v(y).$$

Following the arguments in the proof of Theorem 1, one can show that under the local alternatives H_{in}, $i = 1, 2$, the process $R^1_{n,\widehat{g}}$ converges to $\sigma B(F_u(\cdot)) + D_i(x)$, respectively, where for $i = 1, 2$, $D_1(x) = \mathrm{E}[g_e(I_{t-1}, \theta_0) s_m(I_{t-1}) h^{-1}(I_{t-1}, \theta_0)] f_u(x)$ and $D_2(x) = \mathrm{E}[g_e(I_{t-1}, \theta_0) s_v(I_{t-1}) h^{-1}(I_{t-1}, \theta_0)] x f_u(x)$. It is then clear that the optimal choice for $g_0(I_{t-1})$ under H_{1n} is the orthogonal projection of $s_m(I_{t-1}) h^{-1}(I_{t-1}, \theta_0)$ on the orthocomplement of the span generated by $X_t(\theta_0)$. A similar conclusion holds for H_{2n} (details are omitted). Thus, we have shown that, provided $D_i(x) \neq 0$ for $i = 1, 2$, the tests have nontrivial local power.

3. ARMA-GARCH MODELS

This section verifies the applicability of our previous results for the well-known class of ARMA(P, Q)-GARCH(p, q) models. An ARMA(P, Q)-GARCH(p, q) process, $P, Q, p, q \in \mathbb{N}$, satisfies the equations

$$\begin{cases} Y_t = f(I_{t-1}, \vartheta_0) + \varepsilon_t \equiv c_0 + \sum_{i=1}^{Q} a_{0i} Y_{t-i} - \sum_{j=1}^{P} b_{0j} \varepsilon_{t-j} + \varepsilon_t, & \varepsilon_t = h_t u_t, \\ h_t^2 \equiv h^2(I_{t-1}, \theta_0) = w_0 + \sum_{i=1}^{q} \alpha_{0i} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}^2, & t \in \mathbb{Z}, \end{cases} \tag{10}$$

where u_t is a sequence of i.i.d. random variables independent of $I_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots)'$, and $w_0 > 0$, $\alpha_{0i} \geq 0$, $i = 1, \ldots, q$, $\beta_{0j} \geq 0$, $j = 1, \ldots, p$. Denote a generic vector of parameters by $\vartheta = (c, a_1, \ldots, a_Q, b_1, \ldots, b_P)'$, $\varphi = (w, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$, and $\theta = (\vartheta', \varphi')'$, and the parameter space $\Theta \subset \mathbb{R}^{P+Q+1} \times (0, +\infty) \times [0, +\infty)^{p+q}$. The true parameter value is unknown, and it is denoted by $\theta_0 = (\vartheta_0', \varphi_0')'$, with $\vartheta_0 = (c_0, a_{01}, \ldots, a_{0Q}, b_{01}, \ldots, b_{0P})'$ and $\varphi_0 = (w_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$. We assume that θ_0 belongs to the interior of Θ and that Θ is compact.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

For the stationarity of the model, we define the matrix

$$A_t = \begin{bmatrix} \alpha_1 u_t^2 & \cdots & \alpha_{q-1} u_t^2 & \alpha_q u_t^2 & \beta_1 u_t^2 & \cdots & \beta_{p-1} u_t^2 & \beta_p u_t^2 \\ 1 & \cdots & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1 & \cdots & \alpha_{q-1} & \alpha_q & \beta_1 & \cdots & \beta_{p-1} & \beta_p \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

The spectral radius of a square matrix A is denoted by $\rho(A)$. Denote by $\otimes$ the Kronecker product. Define the polynomials $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$, $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^i$, $A_\vartheta(z) = \sum_{i=1}^Q a_i z^i$, $B_\vartheta(z) = 1 - \sum_{j=1}^P b_j z^i$, $z \in \mathbb{C}$. By convention, $\mathcal{A}_\theta(z) \equiv 0$ if $Q = 0$ and $\mathcal{B}_\theta(z) \equiv 1$ if $P = 0$, and similarly for $A_\vartheta(z)$ and $B_\vartheta(z)$. The next assumption is sufficient for strictly stationarity of the model and for the asymptotic normality of the QMLE estimator (see Francq and Zakoïan, 2004, and references therein).

Assumption B1. Assume u_t^2 has a nondegenerate distribution. If $p > 1$ ($P > 1$), $A_{\vartheta_0}(z)$ ($\mathcal{A}_{\theta_0}(z)$) and $B_{\vartheta_0}(z)$ ($\mathcal{B}_{\theta_0}(z)$) have no common roots, $a_{0P} \neq 0$ or $b_{0Q} \neq 0$, $\mathcal{A}_{\theta_0}(1) \neq 0$ and $\alpha_{0q} + \beta_{0p} \neq 0$. Moreover, $\forall \theta \in \Theta$, $A_\vartheta(z) B_\vartheta(z) = 0$ implies $|z| > 1$. $\rho(\mathrm{E}[(A_t \otimes A_t)]) < 1$ and $\forall \theta \in \Theta$, $\sum_{j=1}^p \beta_j < 1$.

In particular, Chen and An (1998) show that $\rho(\mathrm{E}[(A_t \otimes A_t)]) < 1$ is necessary and sufficient for the existence of a strictly stationarity and ergodicity solution to (10) with finite fourth moment. Assume we are given a sample of size n of such stationary solution, $\{Y_t\}_{t=1}^n$ say. ARMA(P, Q)-GARCH(p, q) models are usually estimated by the QMLE; see Li, Ling, and McAleer (2002) and references therein. We describe the estimator as follows. If $q \geq Q$, the initial values $\widehat{I}_0 = (Y_0, \ldots, Y_{1-(q-Q)-P}, \widetilde{\varepsilon}_{-q+Q}, \ldots, \widetilde{\varepsilon}_{1-q}, \widetilde{\sigma}_0^2, \ldots, \widetilde{\sigma}_{1-p}^2)'$ allow us to compute $\widetilde{\varepsilon}_t(\vartheta)$, for $t = -q + Q + 1, \ldots, n$, and $\widetilde{\sigma}_t^2(\theta)$, for $t = 1, \ldots, n$, according to

$$\begin{cases} \widetilde{\varepsilon}_t \equiv \widetilde{\varepsilon}_t(\vartheta) = Y_t - c_0 - \sum_{i=1}^Q a_{0i} Y_{t-i} + \sum_{j=1}^P b_{0j} \widetilde{\varepsilon}_{t-j}(\vartheta), \\ \widetilde{\sigma}_t^2 \equiv \widetilde{\sigma}_t^2(\theta) = w + \sum_{i=1}^q \alpha_i \widetilde{\varepsilon}_{t-i}^2(\vartheta) + \sum_{j=1}^p \beta_j \widetilde{\sigma}_{t-j}^2(\theta). \end{cases}$$

When $q < Q$, the required initial values are $\widehat{I}_0 = (Y_0, \ldots, Y_{1-(q-Q)-P}, \widetilde{\varepsilon}_0, \ldots, \widetilde{\varepsilon}_{1-Q}, \widetilde{\sigma}_0^2, \ldots, \widetilde{\sigma}_{1-p}^2)'$. For simplicity, these initial values will be taken as fixed (neither random nor a function of parameters). Denote $\widehat{I}_t = (Y_t, \ldots, Y_1, \widehat{I}_0)$, $t = 1, \ldots n$.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

A QMLE estimator is defined as any measurable solution $\widehat{\theta}_n$ of

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \widetilde{L}_n(\theta), \tag{11}$$

where

$$\widetilde{L}_n(\theta) = n^{-1}\sum_{t=1}^{n} \widetilde{\ell}_t(\theta), \qquad \widetilde{\ell}_t = \widetilde{\ell}_t(\theta) = \frac{\widetilde{\varepsilon}_t^2(\vartheta)}{\widetilde{\sigma}_t^2(\theta)} + \log\widetilde{\sigma}_t^2(\theta).$$

The asymptotic theory of $\widehat{\theta}_n$ has been studied by numerous authors. For instance, in the present context, Francq and Zakoïan (2004) proved that under Assumption B1,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \longrightarrow_d N(0, J^{-1} I J^{-1}),$$

where $J = \mathrm{E}_{\theta_0}\left[\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right]$, $I = \mathrm{E}_{\theta_0}\left[\frac{\partial \ell_t(\theta_0)}{\partial\theta}\frac{\partial \ell_t(\theta_0)}{\partial\theta'}\right]$ (explicit expressions for $\ell_t(\theta)$, $\partial\ell_t(\theta_0)/\partial\theta$, and $\partial^2\ell_t(\theta_0)/\partial\theta\partial\theta'$ can be found in Francq and Zakoïan, 2004). Therefore, Assumption B1 is sufficient for Assumptions A1(a) and A3 here.

It can be shown (see displays 4.58–4.60 in Francq and Zakoïan, 2004) that under our Assumption B1,

$$\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{f}_t(\theta)}{h(I_{t-1},\theta_0)}\right|^4\right] < \infty, \qquad \mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}_t(\theta)}{h(I_{t-1},\theta_0)}\right|^4\right] < \infty$$

holds, so Assumption A2 (and A6(a)) is also satisfied. As for the initial conditions, $\rho(\mathrm{E}[(A_t \otimes A_t)]) < 1$ implies for a positive constant $\rho < 1$,

$$\mathrm{E}\sup_{\theta\in\Theta_0}\left|\frac{h(\widehat{I}_{t-1},\theta) - h(I_{t-1},\theta)}{h(I_{t-1})}\right|^2 < C\rho^t,$$

$$\mathrm{E}\sup_{\theta\in\Theta_0}\left|\frac{f(\widehat{I}_{t-1},\theta) - f(I_{t-1},\theta)}{h(I_{t-1})}\right|^2 < C\rho^t.$$

Hence, Assumption A4 also holds. Therefore, we observe that our assumptions are naturally satisfied for ARMA-GARCH models.

Some remarks on how to compute our tests for ARMA-GARCH models are in order. First, we compute residuals according to $\widehat{u}_t = \widetilde{\varepsilon}_t(\widehat{\vartheta}_n)/\widetilde{\sigma}_t(\widehat{\theta}_n)$, $t = 1,\ldots,n$, using the QMLE $\widehat{\theta}_n = (\widehat{\vartheta}_n', \widehat{\varphi}_n')'$ and the previous initial conditions. Expressions for the derivatives $\dot{f}_t(\theta_0)$ and $\dot{h}_t(\theta_0)$ can be found in Koul and Ling (2006), among others. We remark that for *zero mean homoskedastic* ARMA models there is no need to run any preliminary regression. In this special case, any sample centered weight $g_0(I_{t-1})$ will work, since it will trivially satisfy (4) and (5).

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

4. FINITE SAMPLE PERFORMANCE AND DATA ANALYSIS

This section examines the performance through some Monte Carlo experiments of the test statistics CvM_n and KS_n for testing H_0. We compare the new tests with Wooldridge's (1990) test based on the quadratic form

$$W_n = S_n' \Sigma_n^{-1} S_n,$$

where $S_n = n^{-1/2} \sum_{t=1}^n \phi(I_t, \widehat{\theta}_n)$ and $\Sigma_n = n^{-1} \sum_{t=1}^n \phi(I_t, \widehat{\theta}_n)\phi'(I_t, \widehat{\theta}_n)$, with $\phi(I_t, \widehat{\theta}_n) = \widehat{g}_c(I_{t-1}, \widehat{\theta}_n) \times (\widehat{u}_t, \widehat{u}_t^2)'$. Under the null hypothesis H_0, W_n converges to a chi-squared distribution with two degrees of freedom. All of the experiments are carried out using the GARCH Toolbox from Matlab 7.1. First, we consider the data generating process and fitted model of Example 1 to illustrate how the choice of g_0 affects the power properties. Table 1 reports the rejection probabilities for several choices of the initial weight g_0 and sample sizes $n = 100$, 500, and 1,000. As expected, our tests have low power against this alternative when asymmetric weights are chosen, even for large sample sizes such as $n = 1,000$. However, for other choices of g_0, our tests present a satisfactory power performance, with moderate power for $g_0(I_{t-1}) = Y_{t-1}Y_{t-2}$ and $\varepsilon_{t-1}^2 1(\varepsilon_{t-1} \leq 0)$, and high power for $g_0(I_{t-1}) = Y_{t-1}^2$, $\cos(Y_{t-1})$, and ε_{t-1}^2. Our tests outperform Wooldridge's (1990) test for almost all sample sizes and choices of g_0. This simulation suggests that it may be important to try several choices of g_0 for power purposes.

In the second experiment, the following AR(1)-GARCH(1,1) is used:

$$Y_t = \mu + aY_{t-1} + \varepsilon_t, \qquad \varepsilon_t = h(I_{t-1}, \theta_0)u_t,$$
$$h^2(I_{t-1}, \theta_0) = \alpha_{00} + \alpha_{01}\varepsilon_{t-1}^2 + \beta_{01}h^2(I_{t-1}, \theta_0), \tag{12}$$

where $\{u_t\}$ are i.i.d. disturbances, the true parameters are $\theta_0 = (\mu, a, \alpha_{00}, \alpha_{01}, \beta_{01}) \in \Theta$, with

$$\Theta = \{(\mu, a, \alpha_0, \alpha_1, \beta_1) \in \mathbb{R}^5 : |a| \leq \overline{a} < 1,$$
$$\alpha_0 \geq c > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \text{ and } \alpha_2 + \alpha_3 \leq \overline{\rho} < 1\}.$$

TABLE 1. Rejection frequencies (over 100) for a Gaussian AR(1)-ARCH(1), 5% level

	$n = 100$			$n = 500$			$n = 1000$		
$g_0(I_{t-1})$	CvM_n	KS_n	W_n	CvM_n	KS_n	W_n	CvM_n	KS_n	W_n
Y_{t-1}^2	16.1	12.4	13.0	87.9	76.0	59.7	97.6	94.1	68.2
Y_{t-1}^3	15.5	4.0	3.4	25.5	9.5	3.5	30.1	13.0	4.0
$Y_{t-1}Y_{t-2}$	6.6	2.7	4.3	46.3	33.7	32.4	79.5	68.9	50.6
$\sin(Y_{t-1})$	12.5	6.1	4.2	14.3	9.7	5.3	16.0	10.2	8.2
$\cos(Y_{t-1})$	24.0	19.7	34.6	98.6	94.3	88.7	100.0	100.0	93.2
$\varepsilon_{t-1}^2 1(\varepsilon_{t-1} \leq 0)$	12.9	7.6	10.7	51.2	41.3	50.7	80.4	71.8	66.0
ε_{t-1}^2	14.7	10.8	8.1	71.3	59.9	57.0	93.3	88.5	66.9

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

The null model in H_0 corresponds to $a = 0$, that is, a pure GARCH(1,1) model. We report in Figure 1 the rejection probabilities for model (12) with a varying from -0.9 to 0.9 in increments of 0.1, and $(\mu, \alpha_0, \alpha_1, \beta_1) = (0, 0.025, 0.25, 0.5)$. The empirical power functions are plotted in Figure 1 for $n = 100$ and the innovations' distributions $N(0, 1)$ and t-student with 3 degrees of freedom. We consider 1,000 replications in all experiments. For the rest of the simulations, the initial weight function g is taken to be $g_0(I_{t-1}) = Y_{t-1}$, $t \in \mathbb{Z}$. Other choices of g_0 were considered but not reported for the sake of space. They are available from the author upon request.

From Figure 1 we see that the size performance is satisfactory. The rejection probabilities for the Gaussian distribution for CvM_n, KS_n, and W_n are, respectively, 0.057, 0.037, and 0.050. For the t distribution with 3 degrees of freedom, the rejection probabilities are, respectively, 0.065, 0.041, and 0.035. These simulations show the robustness of our tests to fat-tail distributions. The empirical power properties of our tests are excellent against these alternatives already for samples as small as $n = 100$. For Gaussian innovations, CvM_n outperforms W_n, which in turn has higher power than KS_n. For t-student innovations our tests have higher power than Wooldridge's (1990) test uniformly in all values of c, with CvM_n attaining the highest rejection probabilities.

In the third experiment, we consider some nonlinear alternatives to the AR(1)-GARCH(1,1) model (12) with mean parameters $(\mu, a) = (0.02, 0.02)$ and variance parameters $(\alpha_0, \alpha_1, \beta_1) = (0.08, 0.1, 0.85)$. These parameter values are motivated

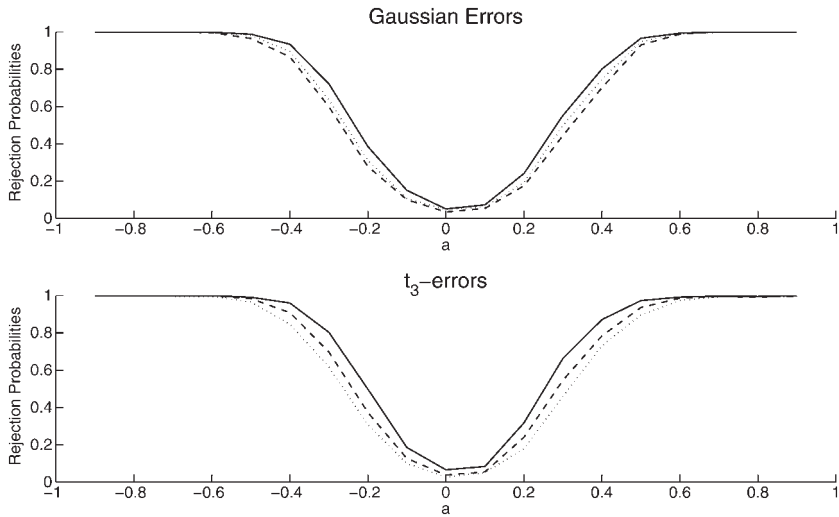


FIGURE 1. Rejection probabilities for *CvM* (solid line), *KS* (dash-dotted line), and *W* (dotted line) for the AR(1)-GARCH(1,1) model. Sample size $n = 100$. Nominal level 5%. Top panel: standard Gaussian innovations. Botton panel: Student-t innovations with three degrees of freedom.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

from our application. We denote the null model by A0. The alternatives considered are:

ALT1: ARMA(1,1)-GARCH(1,1) model: $Y_t = 0.02 + 0.02Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$.

ALT2: TAR model: $Y_t = 0.6Y_{t-1} + \varepsilon_t$ if $Y_{t-1} \leq 1$ and $Y_t = -0.5Y_{t-1} + \varepsilon_t$ if $Y_{t-1} > 1$.

ALT3: EGARCH(1,1) model: $Y_t = h_t u_t$, $\ln h_t^2 = 0.025 + 0.5 \ln h_{t-1}^2 + 0.25 (|u_{t-1}| - (2/\pi)^{1/2}) - 0.8u_{t-1}$.

ALT4: Bilinear model (BIL): $Y_t = 0.6Y_{t-1} + 0.7u_{t-1}Y_{t-2} + u_t$.

ALT5: Nonlinear moving average model (NLMA): $Y_t = 0.8u_{t-1}^2 + u_t$.

In models ALT1–ALT5, ε_t is as in (12) with $(\alpha_0, \alpha_1, \beta_1) = (0.08, 0.1, 0.85)$ and $\{u_t\}$ i.i.d. $N(0, 1)$ variates. Table 2 shows the rejection probabilities for these alternatives for $n = 300$, $n = 600$, and $n = 1,000$ when $g_0(I_{t-1}) = Y_{t-1}$.

Again, Table 2 reveals that the size performance of our tests is satisfactory. Here, KS_n presents some slight underrejection. It is shown that our tests are able to detect all the linear and nonlinear alternatives considered. The highest power is attained at ALT5. Again, CvM_n presents the highest empirical power. KS_n has higher power than W_n against ALT3, ALT4, and ALT5, whereas in general the reciprocal is true for ALT1 and ALT2. Unreported simulations confirm that the excellent empirical power and size performance observed extend to other parameter values $(\alpha_0, \alpha_1, \beta_1)$ in the stationary region. These simulations indicate that the proposed tests have satisfactory size and power performance in moderate finite samples, thereby providing useful model checks for ARMA-GARCH models.

We next apply our testing methodology to the well-known and extensively studied S&P 500 daily stock index. The debate on whether the dynamics of economic and financial time series are determined by the conditional mean or the conditional variance has important implications on many applications including portfolio selection and asset pricing. The S&P 500 daily stock index is a representative of the data for which the GARCH model has been extensively used; see, e.g., Bollerslev, Chou, and Kroner (1992) and references therein. We consider

TABLE 2. Rejection frequencies (over 100), 5% level

n	Test	A0	ALT1	ALT2	ALT3	ALT4	ALT5
300	CvM_n	4.60	48.80	40.50	23.70	18.70	83.90
	KS_n	3.80	38.30	35.30	21.20	16.20	84.00
	W_n	4.70	44.40	31.30	19.90	16.90	63.10
600	CvM_n	4.90	71.20	60.90	50.00	41.30	99.40
	KS_n	3.90	63.10	55.80	44.90	39.80	99.70
	W_n	4.30	65.20	56.60	32.60	35.80	95.60
1000	CvM_n	4.40	86.30	82.30	77.80	67.40	100.0
	KS_n	3.80	77.80	77.80	71.70	65.30	100.0
	W_n	4.40	83.80	81.20	54.40	59.30	99.80

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

several periods of this stock index. The first period spans from 2 January 1990 until 31 December 1993, the second period from 3 January 1994 until 31 December 1997, and the third from 2 January 1998 until 28 August 2002. The martingale properties of these samples were studied in Escanciano and Velasco (2006). The number of observations in each period is 1013, 1011, and 1170, respectively. A common feature of financial data is that the performance of any model is not consistent across different subperiods of the sample, suggesting that the correct model is period-specific; see, e.g., Loudon, Watt, and Yadav (2000).

In this application, we fit an AR(1)-GARCH(1,1) model to the log differences of the S&P 500 (Y_t) as in (12). We estimate the parameters with the QMLE, and the results are reported in Table 3; as usual, the standard errors are in parentheses.

From Table 3 we observe that the mean equation parameters are statistically insignificant in the first and third periods. The latter fact is consistent with the results obtained by Escanciano and Velasco (2006), who found evidence that in the periods 1990–1993 and 1998–2002, the S&P 500 return series is a martingale difference sequence (mds). In the same study, these authors provided evidence that for the period 1994–1997, the S&P 500 is not a mds, which is again consistent with the results of Table 3. Comparing the different periods, we observe a substantial decrease in the persistence of the series.

We proceed by fitting an AR(1)-GARCH(1,1) model as in (12) for the period 1994–1997 and a pure GARCH(1,1) model for the periods 1990–1993 and 1998–2002. We apply our tests and Wooldridge's (1990) test to these data sets with several initial weights $g_0(I_{t-1})$. We report the values of the test statistics $Cv M_n$, KS_n, and W_n in Table 4.

We summarize our findings as follows: We reject the null hypothesis of GARCH(1,1) specification with both statistics $Cv M_n$ and KS_n at the 5% level for essentially all cases within the first period. Wooldridge's (1990) test fails to detect this alternative. Thus, it is shown that a GARCH(1,1) specification for the conditional variance is not adequate for the first period. We conclude that weights involving the first lagged value of the dependent variable or the first lagged conditional mean error are sufficient to account for the lack-of-fit of the GARCH model. Further misspecifications at higher order lags can be investigated similarly. The AR(1)-GARCH(1,1) and GARCH(1,1) specifications for the periods 1994–1997 and 1998–2002 seem satisfactory and no significative test statistic is found at the 5% level with the weights considered. Only the CvM test at 10% uncovers

TABLE 3. Estimates of AR(1)-GARCH(1,1) model for the S&P500 daily stock index

Parameters	1990–1993	1994–1997	1998–2002
a_0	0.0257 (0.0219)	0.0764 (0.0220)	0.0281 (0.0370)
a_1	0.0455 (0.0334)	0.0648 (0.0355)	0.0200 (0.0331)
α_0	4.7e-005 (0.0006)	0.0086 (0.0033)	0.0808 (0.0254)
α_1	0.0176 (0.0033)	0.0772 (0.0108)	0.1000 (0.0169)
β_1	0.9814 (0.0032)	0.9134 (0.0146)	0.8567 (0.0271)

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

TABLE 4. Model checks for the S&P500 daily stock index

	1990–1993			1994–1997			1998–2002		
	GARCH			AR-GARCH			GARCH		
$g_0(I_{t-1})$	CvM_n	KS_n	W_n	CvM_n	KS_n	W_n	CvM_n	KS_n	W_n
$\lvert Y_{t-1}\rvert$	0.58**	1.63**	1.61	0.13	0.74	2.22	0.20	0.95	1.66
Y_{t-1}^2	0.54**	1.43**	1.82	0.15	0.83	1.47	0.12	0.76	1.12
$\varepsilon_{t-1}^2 1(\varepsilon_{t-1} \leq 0)$	0.11	0.76	1.01	0.29	1.03	2.82	0.40*	1.07	0.61
$\lvert\varepsilon_{t-1}\rvert$	0.58**	1.63**	1.48	0.12	0.75	2.24	0.20	0.96	1.61
ε_{t-1}^2	0.54**	1.44**	1.62	0.15	0.82	1.34	0.12	0.76	1.06

Notes: *Significant at the 10% level. **Significant at the 5% level. Asymptotic tests.

some discrepancy with the assumed model in the last period for the asymmetric weight. Unreported simulations confirm our findings with other weights. In particular we have also considered standardized weights $g_0(I_{t-1})h^{-1}(I_{t-1}, \theta_0)$, where $g_0(I_{t-1})$ is as in Table 4. The results with these standardized weights are similar to those with the unstandardized ones, and hence they are omitted.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed ADF diagnostic tests for the adequacy of a class of heteroskedastic time series models, including but not restricted to ARMA-GARCH models. Unlike most tests available in the literature that use empirical processes of covariates weighted by residuals, our tests are based on empirical processes of residuals weighted by appropriate transformations of the possible infinite-dimensional conditioning set. This fact allows us to construct tests that are not affected by the so-called curse of dimensionality in regression model checks. This property is essential if the interest is in non-Markovian models such as ARMA-GARCH models, where the conditioning set can be infinite-dimensional.

The weights in the weighted residual process are orthogonal to the score functions of the model. We show that this orthogonality renders to tests that are ADF. In addition to the ADF property and robustness to high-dimensional conditioning set, another appealing property of our tests is their computational simplicity (to compute the test statistics we just need to run an OLS estimation and order the resulting residuals). The "price" we pay for these appealing properties is that an unfortunate choice of g_0, as well as the particular features of the unknown data generating process and the specified model, may lead to inconsistency of our tests; see Example 1. In fact, as shown recently by Escanciano (2009), this price is, in practice (i.e., in finite samples), paid by *all* omnibus specification tests for regression models, including the consistent ones. It is important to stress that this weakness can be also seen as its strength, since unlike with consistent tests, with our tests the local power properties are "controlled" by the researcher through the choice of g_0; whereas, unlike fully directional tests (e.g. Lagrange multiplier tests), our tests may have global power against a large class of alternatives; see

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Example 2. Thus, our tests provide an interesting compromise between consistent and directional tests.

Two extensions of our methods can be shown to be useful in improving the power properties of our tests, although at the expense of losing the ADF property. First, a simple and trivial extension is to multivariate weights g_0. The principal components decomposition in Stute et al. (2008) can be considered as a method to combine (asymptotic independent) information from different weight functions into a single statistic. The second and more complicated extension is to weights that are function-valued. For instance, if we take $g_0(I_{t-1}) = 1(Y_{t-1} \le y)$ as a function indexed by $y \in \mathbb{R}$, then our theory would lead to tests with possibly better power properties, based on the process

$$R^1_{n,\widehat{g}}(x, y) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(Y_{t-1} \le y) - \widehat{\gamma}'_n(\theta_n, y)\widehat{X}_t(\theta_n)\} 1(\widehat{u}_t \le x),$$

with $\widehat{\gamma}'_n(\theta_n, y)$ denoting the OLS estimator in the regression of $1(Y_{t-1} \le y)$ against $\widehat{X}_t(\theta_n)$. Similarly to Theorem 1, it can be proved that under the conditions of Corollary 1,

$$R^1_{n,\widehat{g}}(x, y) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_e(I_{t-1}, \theta_0, y)[1(u_t \le x) - F_u(x)] + o_P(1),$$

where $g_e(I_{t-1}, \theta_0, y) = 1(Y_{t-1} \le y) - \gamma_0(\theta_0, y)X_t(\theta_0)$. Unlike $R^1_{n,\widehat{g}}(x)$, the null limit of $R^1_{n,\widehat{g}}(x, y)$ is a Gaussian process with zero mean and complicated covariance function. The asymptotic null distribution of functionals of $R^1_{n,\widehat{g}}(x, y)$ can be approximated by bootstrap methods such as the multiplier-type bootstrap described in van der Vaart and Wellner (1996). The power properties of the resulting tests will be investigated elsewhere, in comparison with the simpler and still powerful tests proposed in the present paper.

NOTES

1. We thank Wenceslao González-Manteiga for pointing out this reference.
2. We thank an anonymous referee for providing this example.

REFERENCES

Bai, J. & S. Ng (2001) A test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225–258.

Bera, A.K. & M.L. Higgins (1997) Arch and bilinearity as competing models for nonlinear dependence. *Journal of Business and Economic Statistics* 15, 43–50.

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd ed. Wiley.

Bollerslev, T., R.Y. Chou, & K.F. Kroner (1992) ARCH modelling in finance. *Journal of Econometrics* 52, 5–59.

Box, G. & D. Pierce (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1527.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Chang, N.M. (1990) Weak convergence of a self-consistent estimator of a survival function with doubly censored data. *Annals of Statistics* 18, 391–404.
Chen, M. & H.Z. An (1998) A note on the stationarity and the existence of moments of the GARCH model. *Statistica Sinica* 8, 505–510.
Chen, X., O. Linton, & I. van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.
Delgado, M.A. & J.C. Escanciano (2007) Nonparametric tests for conditional symmetry in dynamic models. *Journal of Econometrics* 141, 652–682.
Escanciano, J.C. (2006). Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* 101, 531–541.
Escanciano, J.C. (2008) Joint and marginal diagnostic tests for conditional mean and variance specifications. *Journal of Econometrics* 143, 74–87.
Escanciano, J.C. (2009) On the lack of power of omnibus specification tests. *Econometric Theory* 25, 1–33.
Escanciano, J.C. & C. Velasco (2006) Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* 134, 151–185.
Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag.
Francq, C. & J.M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH. *Bernoulli* 10, 605–637.
Gao, J. & M. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–882.
Hall, P. & C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.
Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1974.
Hidalgo, J. & P. Zaffaroni (2007) A goodness of fit test for ARCH(∞). *Journal of Econometrics* 141, 835–875.
Hong, Y. & T.H. Lee (2003) Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory* 19, 1065–1121.
Jennrich, R.I. (1969) Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.
Koenker, R. & Q. Zhao (1996) Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12, 793–814.
Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd ed. Lecture Notes in Statistics, vol. 166. Springer.
Koul, H.L. & S. Ling (2006). Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994–1012.
Li, W.K., S. Ling, & M. McAleer (2002) A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Survey* 16, 245–269.
Li, W.K. & T.K. Mak (1994) On the squared residual autocorrelation in nonlinear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.
Ljung, G.M. & G.E.P. Box (1978) A measure of lack of fit in time series models. *Biometrika* 65, 297–303.
Loudon, G.F., W.H. Watt, & P.K. Yadav (2000) An empirical analysis of alternative parametric ARCH models. *Journal of Applied Econometrics,* 15, 117–136.
Lundbergh, S.T. & T. Teräsvirta (2002). Evaluating GARCH models. *Journal of Econometrics* 110, 417–435.
Ngatchou-Wandji, J., (2005) Checking nonlinear heteroscedastic time series models. *Journal of Statistical Planning and Inference* 133, 33–68.
Robinson, P.M. & P. Zaffaroni (2006) Pseudo-maximum likelihood estimation of ARCH(∞) models. *Annals of Statistics* 34, 1049–1074.
Shorack, G. & J. Wellner (1986) *Empirical Processes with Applications to Statistics*. Wiley.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics 181. Berlin-Heidelberg.

Stute, W., W.L. Xu, & L.X. Zhu (2008) Model diagnosis for parametric regression in high dimensional spaces. *Biometrika* 95, 451–467.

Tsay, R.S. (2005). *Analysis of Financial Time Series*, 2nd ed. Wiley-Intersci.

Van der Vaart, A.W. & J.A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

Wooldridge (1990) A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.

APPENDIX: Proofs

In order to handle initial conditions, we need to extend some general results in Koul and Ling (2006). More concretely, Assumption (4.10) in Lemma 4.1 in Koul and Ling seems to be a strong assumption when initial conditions are considered; see the discussion before Theorem 1. To that end, we shall introduce a weak convergence theorem that generalizes Theorem 2.2.5 in Koul (2002) in several aspects. Most notably, unlike the aforementioned results, our weak convergence theorem allows us to consider initial conditions and to study the behavior of our tests under the alternative hypothesis.

We need some notation. For each $n \geq 1$, $\{u_{n,t}\}_{t=1}^{n}$ are i.i.d. variables with cdf F_u. Let $\mathcal{F}_{n,t-1}$ be the σ-field generated by the random variables $\{Y_{n,s}\}_{s=1}^{t-1}$ and $\{u_{n,s}\}_{s=1}^{t-1}$, i.e., $\mathcal{F}_{n,t-1} := \sigma(Y_{n,t-1}, u_{n,t-1}, \ldots, Y_{n,1}, u_{n,1})$. Denote by $(\Omega_n, \mathcal{A}_n, P_n)$, $n \geq 1$, the probability space in which all the random variables $\{u_{n,t}, Y_{n,t}\}_{t=1}^{n}$ are defined. Let $||\cdot||_{2,n}$ indicate the $L_2(P_n)$ norm. Assume that $u_{n,t}$ is independent of $\mathcal{F}_{n,t-1}$ for all $1 \leq t \leq n$ and all $n \in \mathbb{N}$. The reader is referred to Definitions 2.1.5 and 2.1.6 in van der Vaart and Wellner (1996) for definitions of the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_2)$ and the covering number $N(\varepsilon, \mathcal{F}, ||\cdot||)$ with respect to a pseudo-norm $||\cdot||$.

Define the sets $\overline{\mathbb{R}} = [-\infty, \infty]$ and $\Delta_K = \{v \in \mathbb{R}^p : |v| \leq K\}$ for an arbitrary but fixed constant $K > 0$. Let $\{d_{1tn}(v), d_{2tn}(v), g_{n,t}(v)\}_{t=1}^{n}$ be a sequence of random variables such that $d_{1tn}(v)$, $d_{2tn}(v)$, and $g_{n,t}(v)$ are $\mathcal{F}_{n,t-1}$-measurable for each $v \in \mathbb{R}^p$. For a given $K > 0$ and with $z = (z_1, z_2) \in \mathbb{R}^2$, define

$$\rho_K(x_1, x_2) := \sup_{|z| \leq K} |F_u(x_1 + x_1 z_1 + z_2) - F_u(x_2 + x_2 z_1 + z_2)|, \qquad x_1, x_2 \in \overline{\mathbb{R}}.$$

Let $d_K(u_1, u_2) := \rho_K(x_1, x_2) + |v_1 - v_2|$, where $u_1 = (x_1, v_1')'$ and $u_2 = (x_2, v_2')'$ belong to $\Pi_K := \overline{\mathbb{R}} \times \Delta_K$.

For $(x, v) \in \Pi_K$ define $i_{nt}(x, v) := x + x d_{1tn}(v) + d_{2tn}(v)$ and the process

$$V_n(x, v) := \frac{1}{n} \sum_{t=1}^{n} g_{n,t}(v)\{1(u_{n,t} \leq i_{nt}(x, v)) - F_u(i_{nt}(x, v))\},$$

indexed by $\gamma = (x, v) \in \Pi_K$. Consider the following assumptions:

Assumption W1. For any $0 < \delta < 1$,

$$\max_{1 \leq t \leq n, n \geq 1} \mathrm{E}\left[\sup_{|v_1 - v_2| \leq \delta} \left|g_{n,t}(v_1) - g_{n,t}(v_2)\right|\right] < C\delta \quad \text{and}$$

$$\sup_{\varepsilon \in (0,1) \cap \mathbb{Q}} \max_{1 \leq t \leq n} n^{-1} \sum_{t=1}^{n} \varepsilon^{-2} \sup_{|v_1 - v_2| \leq \varepsilon^2} \left|g_{n,t}(v_1) - g_{n,t}(v_2)\right|^2 = O_P(1).$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Moreover, for each $K > 0$, $\max_{1\le t\le n, n\ge 1} \mathrm{E}[\overline{g}^2_{n,t,K}] < \infty$ and $n^{-1}\sum_{t=1}^n \overline{g}^2_{n,t,K} = O_{P_n}(1)$, where $\overline{g}_{n,t,K} = \sup_{v\in\Delta_K} |g_{n,t}(v)|$.

Assumption W2. Assume $\{d_{1tn}(\cdot), d_{2tn}(\cdot) : 1 \le t \le n\}$ are such that, for each $K > 0$, and all $\delta > 0$

$$\sup_{v\in\Delta_K} \max_{1\le t\le n} |d_{1tn}(v)| + |d_{2tn}(v)| = O_{P_n}(1), \tag{A.1}$$

$$\Pr\left(\inf_{v\in\Delta_K} \min_{1\le t\le n} d_{1tn}(v) > c - 1\right) \ge 1 - \delta, \qquad \text{for some } c > 0, \tag{A.2}$$

$$\max_{1\le t\le n, n\ge 1} \mathrm{E}\left[\sup_{|v_1 - v_2|\le\delta} |d_{itn}(v_1) - d_{itn}(v_2)|\right] < C\delta, \qquad i = 1, 2, \quad \text{and} \tag{A.3}$$

$$\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}} n^{-1}\sum_{t=1}^n \varepsilon^{-1}\overline{g}^2_{n,t,K}\left[\sup_{|v_1 - v_2|\le\varepsilon} |d_{itn}(v_1) - d_{itn}(v_2)|\right] = O_{P_n}(1), \qquad i = 1, 2. \tag{A.4}$$

Assumption W3. F_u has an absolutely continuous density f_u such that, for all $0 < b < \infty$ and $0 < c < \infty$,

$$\sup_{x\in\mathbb{R}, (z_1,z_2)'\in\mathcal{Z}_{b,c}} |x|\, f_u(x + xz_1 + z_2) < \infty.$$

THEOREM A1. *Under Assumptions W1–W3, the process $V_n(\gamma)$ is asymptotically tight with respect to $\gamma = (x, v) \in \Pi_K$.*

Proof of Theorem A1: We shall apply Theorem A1 in Delgado and Escanciano (2007) with $w_{nt}(\gamma) = g_{n,t}(v)\{1(u_{n,t} \le i_{nt}(x, v)) - F_u(i_{nt}(x, v))\}$, $\gamma = (x, v) \in \Pi_K$. Define the class of functions $\mathcal{K}_n = \{w_{n\cdot}(\gamma) : \gamma = (x, v) \in \Pi_K\}$ and fix $\gamma_1 = (x_1, v_1) \in \Pi_K$. By Assumption W2, for any $\delta > 0$, there exists $K > 0$ and $c > 0$ such that

$$P(A_{n,K}) \ge 1 - \delta, \qquad P(B_{n,c}) \ge 1 - \delta,$$

where $A_{n,K} \equiv \{\sup_{v\in\Delta_K} \max_{1\le t\le n} |d_{1tn}(v)| + |d_{2tn}(v)| < K\}$ and $B_{n,c} \equiv \{\inf_{v\in\Delta_K} \min_{1\le t\le n} d_{1tn}(v) > c - 1\}$. Now, by Cauchy-Schwartz inequality and triangle inequality, for any $0 < \delta < 1$,

$$\begin{aligned} &\sup_{n\ge 1} \mathrm{E}\left[\sup_{\substack{\gamma_2\in\Pi_K \\ d_K(u_1,u_2)\le\delta}} |w_{nt}(\gamma_1) - w_{nt}(\gamma_2)|\right] \\ &\quad \le C\left(\sup_{n\ge 1} \mathrm{E}\left|g_{n,t}(v_1)\right|^2\right)^{1/2} \\ &\qquad \times \left(\sup_{n\ge 1} \mathrm{E}\left[\sup_{u_2: d_K(u_1,u_2)\le\delta} |F_u(i_{nt}(x_1, v_1)) - F_u(i_{nt}(x_2, v_2))|\right]\right)^{1/2} \end{aligned} \tag{A.5}$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

$$
\begin{aligned}
&+ C \sup_{n\geq 1} \mathrm{E}\left[\sup_{|v_1-v_2|\leq\delta} \left|g_{n,t}(v_1) - g_{n,t}(v_2)\right|\right] \\
&\leq C\left(\sup_{n\geq 1} \mathrm{E}\left[\sup_{x_2:\rho_K(x_1,x_2)\leq\delta} |F_u(i_{nt}(x_1,v_1)) - F_u(i_{nt}(x_2,v_1))|\,1(A_{n,K})\right]\right)^{1/2} \\
&\quad + \left(\sup_{n\geq 1} \mathrm{E}\left[\sup_{u_2:d_K(u_1,u_2)\leq\delta} |F_u(i_{nt}(x_2,v_1)) - F_u(i_{nt}(x_2,v_2))|\,1(B_{n,c})\right]\right)^{1/2} + C\delta^{1/2} \\
&\leq C\delta^{1/2},
\end{aligned}
$$

where the last inequality is due to Assumptions W2 and W3 and the definition of ρ_K.

Theorem 3 in Chen, Linton, and van Keilegom (2003) yields that

$$
\sup_{n\geq 1} N_{[\cdot]}(\varepsilon, \mathcal{K}_n, \|\cdot\|_{2,n}) \leq N((\varepsilon/2C)^2, \overline{\mathbb{R}}, \rho_K) \times N((\varepsilon/2C)^2, \Delta_K, |\cdot|).
$$

Hence, from equation (4.7) in Koul and Ling (2006) and the last display, it follows that

$$
\sup_{n\geq 1} \int_0^\infty \left(\log(N_{[\cdot]}(\varepsilon, \mathcal{K}_n, \|\cdot\|_{2,n}))\right)^{1/2} d\varepsilon < \infty.
$$

Let $\mathcal{B}_\varepsilon = \{B_k; 1 \leq k \leq N_\varepsilon \equiv \sup_{n\geq 1} N_{[]}(\varepsilon, \mathcal{K}_n, \|\cdot\|_{2,n}\}$, with $B_k = [\underline{w}_{k,nt}(\cdot), \overline{w}_{k,nt}(\cdot)]$, be the partition of $\mathcal{K}_n$ in ε-brackets with respect to $\|\cdot\|_{2,n}$, which is bounded by the corresponding product of $(\varepsilon/2C)^2$-nets of Δ_K and $\overline{\mathbb{R}}$ implicit in the proof of Theorem 3 in Chen et al. (2003). Arguing as in (A.5), the conditional quadratic variation of the empirical process V_n on the previous partition is

$$
\begin{aligned}
V_n(\mathcal{B}_\varepsilon) &= \max_{1\leq k\leq N} n^{-1} \sum_{t=1}^n \mathrm{E}\left[\sup_{\gamma_1,\gamma_2\in B_k} |w_{nt}(\gamma_1) - w_{nt}(\gamma_2)|^2 \mid \mathcal{F}_{n,t-1}\right] \\
&\leq C n^{-1} \sum_{t=1}^n \overline{g}_{n,t,K}^2 \left[\sup_{d_K(u_1,u_2)\leq(\varepsilon/2C)^2} \cdot |F_u(i_{nt}(x_1,v_1)) - F_u(i_{nt}(x_2,v_2))|\right] \\
&\quad + C n^{-1} \sum_{t=1}^n \sup_{|v_1-v_2|\leq(\varepsilon/2C)^2} \left|g_{n,t}(v_1) - g_{n,t}(v_2)\right|^2 \\
&\leq C\varepsilon^2 + C n^{-1} \sum_{t=1}^n \overline{g}_{n,t,K}^2 \left[\sup_{|v_1-v_2|\leq(\varepsilon/2C)^2} |d_{1tn}(v_1) - d_{1tn}(v_2)|\right] \\
&\quad + C n^{-1} \sum_{t=1}^n \overline{g}_{n,t,K}^2 \left[\sup_{|v_1-v_2|\leq(\varepsilon/2C)^2} |d_{2tn}(v_1) - d_{2tn}(v_2)|\right] + C\varepsilon^2 + o_P(1).
\end{aligned}
$$

Therefore, the last display and Assumption W2 imply (14) in Delgado and Escanciano (2007, p. 669) for the partition $\mathcal{B}_\varepsilon$. Therefore, Theorem A1 in Delgado and Escanciano (2007) applies and the asymptotic tightness of $V_n(\gamma)$ follows. ■

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Proof of Theorem 1. Define

$$d_{1tn}(v) := \frac{h(\widehat{I}_{t-1}, \theta_0 + n^{-1/2}v) - h(I_{t-1}, \theta_0)}{h(I_{t-1}, \theta_0)},$$

$$d_{2tn}(v) := \frac{f(\widehat{I}_{t-1}, \theta_0 + n^{-1/2}v) - f(I_{t-1}, \theta_0)}{h(I_{t-1}, \theta_0)}.$$

To simplify notation, we write $i_{n,t}(x, v) = x + x d_{1tn}(v) + d_{2tn}(v)$.

Define the process

$$K_n(x, v) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{g}_{n,t}(v) \left\{ 1(u_t \leq i_{n,t}(x, v)) - F_u(x) \right\},$$

with $\widehat{g}_{n,t}(v) = g(\widehat{I}_{t-1}, \theta_0 + n^{-1/2}v) - \mathrm{E}[g(I_{t-1}, \theta_0)]$. By Assumption A4, uniformly in $\gamma = (x, v) \in \Pi_K$,

$$K_n(x, v) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_{n,t}(v) \left\{ 1(u_t \leq i_{n,t}(x, v)) - F_u(x) \right\} + o_P(1),$$

where $g_{n,t}(v) = g(I_{t-1}, \theta_0 + n^{-1/2}v) - \mathrm{E}[g(I_{t-1}, \theta_0)]$. Now, write

$$\begin{aligned} K_n(x, v) = {} & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_{n,t}(v) \left[1(u_t \leq x) - F_u(x) \right] \\ & + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_{n,t}(v) \left[F_u(i_{n,t}(x, v)) - F_u(x) \right] \\ & + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_{n,t}(v) \big[1(u_t \leq i_{n,t}(x, v)) - 1(u_t \leq x) \\ & \qquad - F_u(i_{n,t}(x, v)) + F_u(x) \big] + o_P(1) \\ := {} & A_{1n}(x, v) + A_{2n}(x, v) + A_{3n}(x, v) + o_P(1). \end{aligned} \tag{A.6}$$

To study the asymptotic behavior of $A_{3n}(x, v)$ we shall apply Theorem A1 to the process:

$$\widetilde{K}_n(x, v) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_{n,t}(v) \left[1(u_t \leq i_{n,t}(x, v)) - F_u(i_{n,t}(x, v)) \right].$$

We proceed to verify Assumptions W1–W3 for this particular example. Assumption W3 is directly assumed in Assumption A1(b). Assumption A5 and the ergodic theorem imply W1. It remains to verify Assumption W2.

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

For a sufficiently large n, Assumption A4 implies, uniformly in $\upsilon \in \Delta_K$,

$$\max_{1\le t\le n}|d_{1tn}(\upsilon)| \le \max_{1\le t\le n}\left|\frac{h(\widehat{I}_{t-1},\theta_0+n^{-1/2}\upsilon)-h(I_{t-1},\theta_0+n^{-1/2}\upsilon)}{h(I_{t-1},\theta_0)}\right|$$

$$+\max_{1\le t\le n}\left|\frac{h(I_{t-1},\theta_0+n^{-1/2}\upsilon)-h(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)}\right|$$

$$\le O_P(1)+Kn^{-1/2}\max_{1\le t\le n}\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}_t(\theta)}{h(I_{t-1},\theta_0)}\right| = O_P(1).$$

Similarly, it can be proved that $\sup_{\upsilon\in\Delta_K}\max_{1\le t\le n}|d_{2tn}(\upsilon)| = O_P(1)$. This proves (A.1). Now, Assumption A2 yields (A.3), since

$$\sup_{n\ge1}\mathrm{E}\left[\sup_{|\upsilon_1-\upsilon_2|\le\delta}|d_{1tn}(\upsilon_1)-d_{1tn}(\upsilon_2)|\right]$$

$$\le \sup_{n\ge1}\mathrm{E}\left[\sup_{|\upsilon_1-\upsilon_2|\le\delta}\left|\frac{h(\widehat{I}_{t-1},\theta_0+n^{-1/2}\upsilon_1)-h(\widehat{I}_{t-1},\theta_0+n^{-1/2}\upsilon_2)}{h(I_{t-1},\theta_0)}\right|\right] \le C\delta.$$

Now, we proceed to verify (A.4). By Assumption A4 and the ergodic theorem,

$$\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}} n^{-1}\sum_{t=1}^{n}\varepsilon^{-1}\overline{g}_{n,t,K}^2\left[\sup_{|\upsilon_1-\upsilon_2|\le\varepsilon}|d_{1tn}(\upsilon_1)-d_{1tn}(\upsilon_2)|\right]$$

$$\le C\left\{n^{-1/2}\max_{1\le t\le n}\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left|\frac{\dot{h}(\widehat{I}_{t-1},\theta)}{h(I_{t-1},\theta_0)}\right|\right]\right\}n^{-1}\sum_{t=1}^{n}\overline{g}_{n,t,K}^2 = O_P(1).$$

Similar conclusions hold for d_{2tn}. Hence, Theorem A1 applies to $\widetilde{K}_n(x,\upsilon)$. We use this fact to handle A_{3n} in (A.6).

Note that for a fixed (x,υ), as $n\to\infty$, our assumptions imply by Cauchy-Schwartz inequality, Assumptions A4 and A5, as $n\to\infty$,

$$\mathrm{Var}(A_{3n}(x,\upsilon)) \le \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left[\overline{g}_{n,t,K}^2\left|F_u(i_{n,t}(x,\upsilon))-F_u(x)\right|\right]\to 0. \tag{A.7}$$

Theorem A1 implies that $A_{1n}(x,\upsilon)$ is asymptotically tight in $\ell^\infty(\mathcal{D})$, the Banach space of real-valued bounded functions on $\mathcal{D}$, equipped with the supremum norm $\|z\|_{\mathcal{D}} := \sup_{g\in\mathcal{D}}|z(g)|$. From the latter fact and the asymptotic tightness of $\widetilde{K}_n(x,\upsilon)$ and (A.7), it then follows that

$$\sup_{(x,\upsilon)\in\Pi_K}|A_{3n}(x,\upsilon)| = o_P(1).$$

We proceed to study $A_{2n}(x,\upsilon)$. Define

$$\widehat{d}_{1tn}(\upsilon) := \frac{h(\widehat{I}_{t-1},\theta_0+n^{-1/2}\upsilon)-h(I_{t-1},\theta_0+n^{-1/2}\upsilon)}{h(I_{t-1},\theta_0)}, \quad \text{and}$$

$$\widehat{d}_{2tn}(\upsilon) := \frac{f(\widehat{I}_{t-1},\theta_0+n^{-1/2}\upsilon)-f(I_{t-1},\theta_0+n^{-1/2}\upsilon)}{h(I_{t-1},\theta_0)}.$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

Note that

$$A_{2n}(x,\upsilon) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\left[F_u(i_{n,t}(x,\upsilon)) - F_u(x + xd_{1tn}(\upsilon))\right]$$

$$+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\left[F_u(x + xd_{1tn}(\upsilon)) - F_u(x)\right]$$

$$:= B_{1n}(x,\upsilon) + B_{2n}(x,\upsilon).$$

We write

$$B_{1n}(x,\upsilon) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon) f_u(\widetilde{i}_{n,t}(x,\upsilon))\, d_{2tn}(\upsilon)$$

$$= f_u(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\, d_{2tn}(\upsilon)$$

$$+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\left[f_u(x + xd_{1tn}(\upsilon)) - f_u(x)\right] d_{2tn}(\upsilon)$$

$$+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\left[f_u(\widetilde{i}_{n,t}(x,\upsilon)) - f_u(x + xd_{1tn}(\upsilon))\right] d_{2tn}(\upsilon)$$

$$:= C_{1n}(x,g,\upsilon) + C_{2n}(x,g,\upsilon) + C_{3n}(x,g,\upsilon),$$

for some $\widetilde{i}_{n,t}(x,\upsilon)$ between $i_{n,t}(x,\upsilon)$ and $x + xd_{1tn}(\upsilon)$.

Assumptions A1, A2, and A4 and the ergodic theorem imply that for some $\widetilde{\theta}_n$ with $\left|\widetilde{\theta}_n - \theta_0\right| \le n^{-1/2}\upsilon$, uniformly in υ,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon) d_{2tn}(\upsilon) = \upsilon'\frac{1}{n}\sum_{t=1}^{n} g_{n,t}(\upsilon)\frac{\dot{f}_t(\widetilde{\theta}_n)}{h(I_{t-1},\theta_0)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\widehat{d}_{2tn}(\upsilon)$$

$$= \upsilon'\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{f}_t(\theta_0)}{h(I_{t-1})}\right] + o_P(1).$$

This is so because

$$\sup_{\upsilon\in\Delta_K}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_{n,t}(\upsilon)\widehat{d}_{2tn}(\upsilon)\right| \le n^{-1/2}\max_{1\le t\le n}\overline{g}_{n,t,K}\sup_{\upsilon\in\Delta_K}\left|\sum_{t=1}^{n}\widehat{d}_{2tn}(\upsilon)\right| = o_P(1),$$

and, by the uniform law of large numbers of Jennrich (1969),

$$\sup_{\upsilon\in\Delta_K}\left|\frac{1}{n}\sum_{t=1}^{n} g_{n,t}(\upsilon)\frac{\dot{f}_t(\widetilde{\theta}_n)}{h(I_{t-1},\theta_0)} - \mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{f}_t(\theta_0)}{h(I_{t-1})}\right]\right| = o_P(1).$$

Hence,

$$\sup_{\gamma\in\Pi_K}\left|C_{1n}(x,\upsilon) - \upsilon'\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot{f}_t(\theta_0)}{h(I_{t-1})}\right] f_u(x)\right| = o_P(1).$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

By the same arguments, Assumptions A1 and A4 imply that

$$\begin{aligned}\sup_{\gamma\in\Pi_K}|C_{2n}(x,v)| &\le \sup_{\gamma\in\Pi_K}\left|v'\frac{1}{n}\sum_{t=1}^{n}g_{n,t}(v)\left[f_u(x+xd_{1tn}(v))-f_u(x)\right]\frac{\dot f_t(\widetilde\theta_n)}{h(I_{t-1},\theta_0)}\right| \\ &\quad+\sup_{\gamma\in\Pi_K}\left|\frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)\left[f_u(x+xd_{1tn}(v))-f_u(x)\right]\widehat d_{2tn}(v)\right| \\ &\le CKn^{-1/2}\max_{1\le t\le n}\sup_{\theta\in\Theta_0}\left|\frac{\dot h_t(\theta)}{h(I_{t-1},\theta_0)}\right|\sup_{\gamma\in\Pi_K}\left|\frac{1}{n}\sum_{t=1}^{n}g_{n,t}(v)\frac{\dot f_t(\widetilde\theta_n)}{h(I_{t-1},\theta_0)}\right| \\ &\quad+\sup_{\gamma\in\Pi_K}\left|\frac{1}{n}\sum_{t=1}^{n}g_{n,t}(v)\widehat d_{1tn}(v)\frac{\dot f_t(\widetilde\theta_n)}{h(I_{t-1},\theta_0)}\right| \\ &\quad+2\sup_{x\in\mathbb{R}}|f_u(x)|\sup_{\gamma\in\Pi_K}\left|\frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)\widehat d_{2tn}(v)\right| \\ &= o_P(1).\end{aligned}$$

Similarly, it can be proved that $\sup_{\gamma\in\Pi_K}|C_{3n}(x,g,v)|=o_P(1)$.

As for $B_{2n}(x,v)$, by Assumption A1

$$\begin{aligned}B_{2n}(x,v) &= \frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)f_u(x+x\widetilde{d_{1tn}}(v))xd_{1tn}(v) \\ &= f_u(x)x\frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)d_{1tn}(v) \\ &\quad+\frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)\left[xf_u(x+x\widetilde{d_{1tn}}(v))-xf_u(x)\right]d_{1tn}(v) \\ &:= D_{1n}(x,v)+D_{2n}(x,v),\end{aligned}$$

for some $x+x\widetilde{d_{1tn}}(v)$ between $x+xd_{1tn}(v)$ and x.

Assumptions A1, A2, and A4 and the uniform law of large numbers of Jennrich (1969) imply that

$$\sup_{\gamma\in\Pi_K}\left|D_{1n}(x,g,v)-v'\mathrm{E}\left[g_e(I_{t-1},\theta_0)\frac{\dot h_t(\theta_0)}{h(I_{t-1})}\right]f_u(x)x\right|=o_P(1).$$

Similar arguments to those for C_{jn}, $j=2,3$, imply $\sup_{\gamma\in\Pi_K}|D_{2n}(x,g,v)|=o_P(1)$.

Therefore, summarizing, we have proved for each fixed $K>0$,

$$L_{1,n}:=\sup_{\gamma\in\Pi_K}\left|K_n(x,v)-\frac{1}{\sqrt n}\sum_{t=1}^{n}g_{n,t}(v)[1(u_t\le x)-F_u(x)]-v'b(x,g,\theta_0)\right|=o_P(1). \tag{A.8}$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

To conclude the proof of Theorem 1, note that, denoting $g_e(\cdot,\theta_n)=g(\cdot,\theta_n)-\mathrm{E}[g(I_{t-1},\theta_0)]$,

$$R^1_{n,g}(x)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_e(\widehat{I}_{t-1},\theta_n)\{1(\widehat{u}_t(\theta_n)\le x)-F_u(x)\}$$

$$+\left[\mathrm{E}[g(I_{t-1},\theta_0)]\right]-\overline{g}(\theta_n)]\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\widehat{u}_t(\theta_n)\le x)-F_u(x)\}$$

$$:=A_{1n}(x)+A_{2n}(x).$$

It is straightforward to prove that $A_{2n}(x)=o_P(1)$ uniformly in $x\in\mathbb{R}$. Hence, since $\sqrt{n}(\theta_n-\theta_0)=O_P(1)$, for any $\varepsilon>0$, there exists a $K>0$ such that the event $C_{n,K}=\{\sqrt{n}|\theta_n-\theta_0|>K\}$ satisfies $\Pr(C_{n,K})<\varepsilon/2$ for all n sufficiently large. To simplify notation, define

$$L_{2,n}:=\sup_{x\in\mathbb{R}}\Big|R^1_{n,g}(x)-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g_e(I_{t-1},\theta_n)[1(u_t\le x)-F_u(x)]$$

$$-\sqrt{n}(\theta_n-\theta_0)'b(x,g,\theta_0)\Big|.$$

Thus, because $A_{1n}(x)=K_n(x,\sqrt{n}(\theta_n-\theta_0))$, we conclude using (A.8) that, for any $\delta>0$, we can choose n sufficiently large such that $\Pr(L_{2,n}>\delta)\le\Pr(L_{2,n}>\delta,\sqrt{n}|\theta_n-\theta_0|\le K)+\Pr(C_{n,K})\le\Pr(L_{1,n}>\delta)+\Pr(C_{n,K})\le\varepsilon/2+\varepsilon/2=\varepsilon$.

Since ε and δ were arbitrary, we conclude that $L_{2,n}=o_P(1)$. Finally, Theorem 1 is proved after noticing that, by Assumption A5,

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[g_e(I_{t-1},\theta_n)-g_e(I_{t-1},\theta_0)][1(u_t\le x)-F_u(x)]\right|=o_P(1).$$

■

To emphasize the dependence of $X_t(\theta)$ on I_{t-1}, we write, when it is convenient, $X_t(\theta)\equiv X(I_{t-1},\theta)$.

Proof of Corollary 1. Without loss of generality, assume that $\widehat{X}_t(\widehat{\theta}_n)$ are centered in the sample (otherwise, write the residual with centered regressors). Write

$$K^1_{n,\widehat{g}}(\alpha)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{g_0(\widehat{I}_{t-1})-\widehat{\gamma}'_n(\widehat{\theta}_n)\widehat{X}_t(\widehat{\theta}_n)\}1(\widehat{u}_t\le x)$$

$$=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{g_0(\widehat{I}_{t-1})-\gamma'(\theta_0)\widehat{X}_t(\widehat{\theta}_n)\}1(\widehat{u}_t\le x)$$

$$-(\widehat{\gamma}'_n(\widehat{\theta}_n)-\gamma'(\theta_0))\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{X}_t(\widehat{\theta}_n)1(\widehat{u}_t\le x)$$

$$:=I_{1n}(x)+I_{2n}(x).$$

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press

From the uniform law of large numbers of Jennrich (1969), it follows that $|\widehat{\gamma}'(\widehat{\theta}_n) - \gamma'(\theta_0)| = o_P(1)$, and from Theorem A1,

$$\sup_{x\in\mathbb{R}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{X}_t(\widehat{\theta}_n)1(\widehat{u}_t \le x)\right| = O_P(1).$$

Hence, $\sup_{x\in\mathbb{R}}|I_{2n}(x)| = o_P(1)$.

As for $I_{1n}(x)$, again by Theorem 1 applied to $g(\widehat{I}_{t-1},\widehat{\theta}_n) = \{g_0(\widehat{I}_{t-1}) - \gamma'(\theta_0)\widehat{X}_t(\widehat{\theta}_n)\}$, we obtain

$$\sup_{x\in\mathbb{R}}\left|I_{1n}(x) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(I_{t-1},\theta_0)[1(u_t \le x) - F_u(x)]\right| = o_P(1),$$

with $g(I_{t-1},\theta_0) = g_0(I_{t-1}) - \gamma'(\theta_0)X_t(\theta_0)$. The corollary follows from Theorem A1. ■

https://doi.org/10.1017/S0266466609990090 Published online by Cambridge University Press