Econometric Theory, **17**, 2001, 188–221. Printed in the United States of America.

A CONSISTENT TEST FOR CONDITIONAL HETEROSKEDASTICITY IN TIME-SERIES REGRESSION MODELS

CHENG HSIAO
University of Southern California
and
Hong Kong University of Science and Technology

QI LI
Texas A&M University and University of Guelph

We show that the standard consistent test for testing the null of conditional homoskedasticity (against conditional heteroskedasticity) can be generalized to a time-series regression model with weakly dependent data and with generated regressors. The test statistic is shown to have an asymptotic normal distribution under the null hypothesis of conditional homoskedastic error. We also discuss extension of our test to the case of testing the null of a parametrically specified conditional variance. We advocate using a bootstrap method to overcome the issue of slow convergence of this test statistic to its limiting distribution.

1. INTRODUCTION

There are plenty of papers on testing for heteroskedasticity. For instance, Bickel (1978), Breusch and Pagan (1979), Eubank and Thomas (1993), Glejser (1969), Godfrey (1978), Goldfeld and Quandt (1965), Harvey (1976), Hildreth and Houck (1968), Koenker and Bassett (1982), Lee (1992), Newey and Powell (1987), Szroeter (1978), Whang and Andrews (1993), and White (1980) have proposed tests either with or without the prior knowledge of heteroskedasticities. See Pagan and Pak (1993) for a recent review on testing for heteroskedasticity. However, as shown by Hong (1993), many of the existing tests are not consistent in the sense that under certain forms of heteroskedasticity the tests have only trivial power asymptotically.[1]

We thank three referees and Bruce Hansen for insightful comments that greatly improved the paper. Cheng Hsiao's research is supported by National Sciences Foundation grant SBR 96-19330. Qi Li's research is supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences Engineering Research Council of Canada, Ontario Premier's Reseach Excellence Awards, and the Bush Program in the Economics of Public Policy. Address correspondence to: Qi Li, Department of Economics, Texas A&M University, College Station, TX 77843, USA.

 © 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Recently, Hong (1993) and Zheng (1994, 1996) have proposed consistent tests for heteroskedasticity under the assumption that both the regressand and regressors are independently and identically distributed (i.i.d.). This assumption may be a reasonable approximation to cross-sectional observations. But it can be overly restrictive for time-series data. For instance, for linear regression models using time series data even though the error terms may be independently distributed over time, the regressors are often serially correlated. Another example is the autoregressive conditional heteroskedasticity (ARCH) type model of Engle (1982). For ARCH type models, the error terms are conditional homoskedastic if the conditional variables are exogenous regressors. However, conditional on the residual values of the errors in the early periods, the error variance is a function of them and hence heteroskedastic. Therefore, an application of the tests of Hong (1993) or Zheng (1994, 1996) tests will have asymptotic power equal to the sizes of the tests if the conditional set contains the exogenous regressors only. On the other hand, if we enlarge the conditional set to include the residuals in early periods, the independence assumption is violated.

When the independence assumption is violated, the conventional central limit theorems of second order degenerate *U*-statistics (Hall, 1984; de Jong, 1987) no longer apply. Neither can we say for sure that the asymptotic distribution of the test statistic using "generated regressors" will remain the same as using the regressors that are observed. As is well known, although the rate of convergence remains at $\sqrt{n}$, the asymptotic distribution of an estimator based on a parametric or a semiparametric regression model using generated regressors is typically different from the one based on observed regressors (e.g., Hsiao and Mountain, 1985; Pagan, 1984; Pagan and Ullah, 1988; Powell, 1987). Moreover, the normalization factor of our test statistic is of an order greater than the conventional $\sqrt{n}$; it is not clear, a priori, what impact it has on the asymptotic distribution of the test.

In this paper, using the central limit theorem for second order degenerate *U*-statistic with weakly dependent data (Fan and Li, 1999), we show that with a proper choice of the conditional set, the test independently proposed by Hong (1993) and Zheng (1994) is applicable to cases involving weakly dependent data or generated regressors. The regularity conditions we use are quite weak in the sense that they are similar to the conditions used for independent data case (e.g., Hong, 1993). We also show, through Monte Carlo studies, that the proposed test suffers some finite sample size distortions due to its slow speed of convergence to its asymptotic distribution. Therefore, we advocate using some bootstrap methods to construct the critical values for conducting the test with appropriate sizes.

In Section 2, a general methodology of constructing consistent model specification tests and the application of the test to conditional heteroskedasticity are presented. In Section 3 we derive the asymptotic distribution of the test statistic when the conditional variables are observable. Section 4 proposes a bootstrap procedure as an alternative method to approximate the null distribu-

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

tion of the test. In Section 5 we derive the asymptotic distribution of the proposed test when the conditional variables also include generated regressors (e.g., Engle, 1982; Hsiao and Mountain, 1985; Pagan, 1984; Pagan and Ullah, 1988). In Section 6, we generalize our test of conditional homoskedasticity to the case of testing the null of a parametrically specified conditional variance. Section 7 uses Monte Carlo simulations to evaluate the finite sample performances of the test statistics proposed in Sections 3–5. We demonstrate that there are severe undersize problems using normal approximation in finite samples. We also show that this issue can be successfully overcome using bootstrap methods.

Throughout the paper, $\sum_t = \sum_{t=1}^n$, $\sum_t \sum_{s\neq t} = \sum_{t=1}^n \sum_{s\neq t, s=1}^n$, etc.

2. A CONSISTENT TEST FOR HETEROSKEDASTICITY

We consider a stationary time-series regression model of the form

$$Y_t = g(W_t, \gamma_0) + U_t, \tag{1}$$

where $g(\cdot,\cdot)$ is a known function with γ_0 being an $r \times 1$ unknown parameter, $W_t \in R^q$, and U_t the error term. We are interested in deriving a consistent test of conditional heteroskedasticity in the context of (1). The null and the alternative hypotheses are, respectively,

$$H_0 : P[E(U^2|X) = \sigma_0^2] = 1, \qquad (\sigma_0^2 = E(U^2) > 0)$$

and

$$H_1 : P[E(U^2|X) = \sigma_0^2] < 1,$$

where $X \in R^p$, which may be different from the regressors W. We note that if we define $\epsilon_t = U_t^2 - \sigma_0^2$, then $E(\epsilon_t|X_t) = 0$ almost everywhere (a.e.) if and only if H_0 is true. Therefore, a consistent test of conditional heteroskedasticity can be constructed analogous to the consistent tests for model specification. In such a framework, the null hypothesis is written as $E(\epsilon|X) = 0$ for some appropriately chosen random variables (vectors) ϵ and X. For example, in the context of testing a parametric regression functional form, say, linearity: $Y = X'\gamma_0 + \epsilon$, the null hypothesis of $E(Y|X) = X'\gamma_0$ a.e. is equivalent to $E(\epsilon|X) = 0$ a.e. Other examples of null hypotheses in the form of $E(\epsilon|X) = 0$ a.e. include (i) the test of omitted variables (e.g., Ait-Sahalia, Bickel, and Stoker, 1994; Fan and Li, 1996; Lavergne and Vuong, 2000); (ii) the test of a semiparametric partially linear model or a single index model (e.g., Horowitz and Härdle, 1994; Fan and Li, 1996); and (iii) the test of conditional symmetry (e.g., Zheng, 1998).

When the null hypothesis is written as H_0: $E(\epsilon|X) = 0$, we can construct consistent tests of H_0 versus H_1 using various distance measures. When the nonparametric kernel estimation method is used, the measure $J = E[\epsilon E(\epsilon|X) f(X)]$ turns out to be a convenient choice to use because $J = E\{[E(\epsilon|X)]^2 f(X)\} \geq 0$ and $J = 0$ a.e. if and only if H_0 is true. Hence, J can serve as a proper candidate for consistent testing of H_0. The advantages of constructing a test statistic based on

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

the measure J include the following: (i) the leading term of the test statistic has zero mean under the null hypothesis (i.e., no nonzero center terms), and its asymptotic variance is the same whether the data are undersmoothed, optimally smoothed, or oversmoothed; (ii) the leading term of the test statistic has a *simple* form so as to make some complex analysis much simpler, such as the higher order bootstrap analysis (see Fan and Linton, 1997; Li and Wang, 1998); (iii) the regularity conditions for deriving desirable properties of this type of test statistics are quite weak (for a discussion on this in the context of testing a parametric regression functional form, see Li and Wang, 1998). In contrast to the tests based on J, the test statistic based on the differences between the parametric sum of squared residuals (SSR) and the nonparametric kernel SSR contains several center terms, and the asymptotic variance of the test depends on whether the data are undersmoothed, optimally smoothed, or oversmoothed (e.g., Fan and Li, 1992).

The sample analog of J is $n^{-1}\sum_{t=1}^{n}\epsilon_t E(\epsilon_t|X_t)f(X_t)$. Replacing $E(\epsilon_t|X_t)\times f(X_t)$ by a kernel estimator $(na^p)^{-1}\sum_{s\neq t}\epsilon_s K_{ts}$, where $K_{ts}=K((X_t-X_s)/a)$ is the kernel function and a is the smoothing parameter, we get

$$J_n=\frac{1}{n^2a^p}\sum_t\sum_{s\neq t}\epsilon_t\epsilon_s K_{ts}, \tag{2}$$

where J_n is a second order degenerate U-statistic with variable kernel. Often in practice ϵ_t is not observable and is replaced by some estimator, say, $\hat{\epsilon}_t$. In our case, $\epsilon_t=U_t^2-\sigma_0^2$ is not observable; we replace ϵ_t by the estimator $\hat{\epsilon}_t=\hat{U}_t^2-\hat{\sigma}_0^2$, where $\hat{U}_t=Y_t-g(W_t,\hat{\gamma})$ and $\hat{\gamma}$ is some root-n consistent estimator of γ_0, say, $\hat{\gamma}$ is the nonlinear least squares estimator of γ_0, and $\hat{\sigma}_0^2=n^{-1}\sum_t\hat{U}_t^2$. Therefore, the sample analog of (2) becomes

$$\hat{J}_n=\frac{1}{n^2a^p}\sum_t\sum_{s\neq t}\hat{\epsilon}_t\hat{\epsilon}_s K_{ts}\equiv\frac{1}{n^2a^p}\sum_t\sum_{s\neq t}(\hat{U}_t^2-\hat{\sigma}_0^2)(\hat{U}_s^2-\hat{\sigma}_0^2)K_{ts}. \tag{3}$$

The preceding $\hat{J}_n$ test is the same as the one considered by Hong (1993) and Zheng (1994) for independent data, although Hong (1993) derived it from a different approach.

3. THE ASYMPTOTIC DISTRIBUTION OF THE TEST STATISTIC

In this section we derive the asymptotic distribution of $\hat{J}_n$. We show in Appendix A that the leading term of $\hat{J}_n$ is $(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2-\sigma_0^2)(U_s^2-\sigma_0^2)K_{ts}\equiv(n^2a^p)^{-1}\sum_t\sum_{s\neq t}\epsilon_t\epsilon_s K_{ts}$, a second order degenerate U-statistic of the form of (2). A general second order variable kernel U-statistic has the form of $\mathcal{U}_n=\sum_t\sum_{s<t}H_n(Z_t,Z_s)$, where $H_n(\cdot,\cdot)$ is a symmetric function and $H_n(\cdot,\cdot)$ depends on n. The term $\mathcal{U}_n$ is said to be a degenerate U-statistic if $\int H_n(x,y)dF(x)=0$ for all y, where $F(\cdot)$ is the marginal distribution function of X. Hall (1984) and de Jong (1987) derived central limit theorems for general second order degen-

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

erate *U*-statistics with independent data. Fan and Li (1999) generalize Hall's (1984) result to weakly dependent data case. Hence, the asymptotic distribution of $\hat{J}_n$ can be derived using a result of Fan and Li (1999), which is restated in the following lemma for ease of reference.

LEMMA 3.1. *Let* $Z_t = (\epsilon_t, X_t')'$, $\epsilon_t \in R$, $X_t \in R^p$, $K(\cdot)$ *be a kernel function, and a be the smoothing parameter. Denote* $\sigma_\epsilon^2(x) = E[\epsilon_t^2 | X_t = x]$ *and define*

$$J_n = \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} \epsilon_t \epsilon_s K\left(\frac{X_t - X_s}{a}\right).$$

Then under Assumptions (A1) *and* (A2) *of Appendix* A,

$$na^{p/2} J_n \to N(0, \Omega) \text{ in distribution,}$$

where $\Omega = 2E[\sigma_\epsilon^4(X_t) f(X_t)][\int K^2(u)du]$ *and* $f(\cdot)$ *is the marginal density function of* X_t.

Proof. This is a special case of Theorem 2.1 of Fan and Li (1999).

Using Lemma 3.1 to derive the asymptotic distributions of $\hat{J}_n$, we need the following regularity conditions.

(C1)

(i) With probability one, $E[U_t | \mathcal{M}_{-\infty}^t(X), \mathcal{M}_{-\infty}^{t-1}(Y)] = 0$, where $M_s^t(Z)$ is the sigma algebra generated by $(Z_s, \ldots, Z_t)$ for $s < t$;
(ii) $E[|U_t|^{8+\eta}] < \infty$ and $E[|U_{t_1}^{2i_1} U_{t_2}^{2i_2}, \ldots, U_{t_l}^{2i_l}|^{1+\xi}] < M < \infty$ for some arbitrarily small $\eta > 0$ and $\xi > 0$, where $2 \leq l \leq 4$ is an integer, $0 \leq i_j \leq 4$, and $\sum_{j=1}^l i_j = 8$;
(iii) let $\sigma^2(x) = E(U_t^2 | X_t = x)$ and $\mu_j(x) = E(U_t^j | X_t = x)$, $j = 4,6,8$. Then $\sigma^2(x)$ and $\mu_j(x)$ all satisfy some Lipschitz conditions: $|m(x+u) - m(x)| \leq D(x)\|u\|$, where $m(\cdot) = \sigma^2(\cdot)$ or $\mu_j(\cdot)$ for $j = 4,6,8$, and $\|\cdot\|$ denotes the Euclidean norm of $\cdot$; also $E[|D(X)|^{2+\eta'}] < \infty$ for some $\eta' > 0$;
(iv) let $f_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ be the joint probability density function of $(W_1, W_{1+\tau_1}, \ldots, W_{1+\tau_l})$ $(1 \leq l \leq 3)$. Then $f_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ exists and satisfies a Lipschitz condition: $|f_{\tau_1,\ldots,\tau_l}(w_1 + u_1, w_2 + u_2, \ldots, w_l + u_l) - f_{\tau_1,\ldots,\tau_l} \times (w_1, w_2, \ldots, w_l)| \leq D_{\tau_1,\ldots,\tau_l}(w_1, w_2, \ldots, w_l)\|u\|$, where $D_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ is integrable and satisfies the condition that $\int D_{\tau_1,\ldots,\tau_l}(w,\ldots,w)\|w\|^{2\xi} dw < M < \infty$, $\int D_{\tau_1,\ldots,\tau_l}(w_1,\ldots,w_l) f_{\tau_1,\ldots,\tau_l}(w_1,\ldots,w_l)\, dw < M < \infty$ for some $\xi > 1$.

(C2)

(i) $\nabla g(w,\cdot)$ and $\nabla^2 g(w,\cdot)$ are continuous in w and dominated by functions $M_1(w)$ and $M_2(w)$, respectively; $M_1(w)$ has finite fourth moment and $M_2(w)$ has finite second moment, where $\nabla g(w,\cdot)$ and $\nabla^2 g(w,\cdot)$ are $r \times 1$ vector of first order partial derivatives and $r \times r$ matrix of second order partial derivatives of g with respect to γ, respectively;
(ii) $E[\nabla g(W,\gamma)\nabla' g(W,\gamma)]$ is nonsingular for γ in a neighborhood of γ_0.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

(C3)

(i) $K(\cdot)$ is bounded and symmetric with $\int K(u)du = 1$ and $\int \|u\|^2 K(u)du < \infty$;
(ii) the smoothing parameter $a = O(n^{-\bar{\alpha}})$ for some $0 < \bar{\alpha} < (\frac{7}{8})p$.

(C4) The process $\{Z_t\} = \{U_t, W_t', X_t'\}'$ is a strictly stationary absolutely regular process with mixing coefficient $\beta_\tau = O(\rho^\tau)$ for some $0 < \rho < 1$.

Remark 3.1. (C1)(i) is a martingale difference assumption on the error U_t. The martingale error assumption is used, mainly for simplifying the proof, in developing the central limit theorem for degenerate U-statistics with weakly dependent data in Fan and Li (1999a) (e.g., Lemma 3.1). Because the asymptotic analysis of this paper relies heavily on the result of Lemma 3.1, we also need this martingale difference error assumption. (C1)(ii) and (iii) impose some moment and smoothness conditions on $\{U_t\}$. For i.i.d. data, (C1)(ii) requires $E[|U_t|^{8+\eta}] < \infty$ for some small $\eta > 0$. It can be weakened to $E[|U_t|^{4+\eta}] < \infty$ with a much more tedious proof.[2] (C1)(iv) consists of a Lipschitz condition on the joint density function of $f_{\tau_1,\ldots,\tau_l}(w_1, w_2, \ldots, w_l)$ and some integrability conditions. (C2)(i) and (ii) are standard assumptions used in nonlinear regression models. (C3)(i) is a standard (and quite weak) condition on the kernel function (i.e., the standard second order kernel function). (C3)(ii) allows a wide range of smoothing parameter choice. It is only slightly stronger than the usual conditions of $a \to 0$ and $na^p \to \infty$. (C4) requires the mixing coefficient β_m to have a geometric decay rate. (C2) and (C4) ensure that $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$.

THEOREM 3.2. *Let $\{Z_t\}$ be a strictly stationary, absolutely regular process and assume that Assumptions* (C1)–(C4) *hold. Then under H_0,*

$$\hat{I}_n \overset{def}{=} na^{p/2}\hat{J}_n \big/ \sqrt{\hat{\Omega}} \to N(0,1) \quad \textit{in distribution},$$

where $\hat{\Omega} = 2(n^2a^p)^{-1}\sum_t \sum_{s\neq t} \hat{\epsilon}_t^2 \hat{\epsilon}_s^2 K_{ts}$ is a consistent estimator for $\Omega = 2\{E[\sigma_\epsilon^4(X)f(X)]\}[\int K^2(u)\,du]$ with $\sigma_\epsilon^2(x) = E[\epsilon_t^2 | X_t = x]$, $\epsilon_t = U_t^2 - \sigma_0^2$, and $\hat{\epsilon}_t = \hat{U}_t^2 - \hat{\sigma}_0^2$.

Proof. The proof of $\hat{\Omega} = \Omega + o_p(1)$ is given in Appendix A. We prove $na^{p/2}\hat{J}_n \to N(0,\Omega)$ in distribution here. By adding and subtracting terms, we have

$$\begin{aligned}
\hat{J}_n &= (n^2a^p)^{-1}\sum_t \sum_{s\neq t} (\hat{U}_t^2 - \hat{\sigma}_0^2)(\hat{U}_s - \hat{\sigma}_0^2)K_{t,s} \\
&= (n^2a^p)^{-1}\sum_t \sum_{s\neq t} \{[(U_t^2 - \sigma_0^2) + (\hat{U}_t^2 - U_t^2) - (\hat{\sigma}_0^2 - \sigma_0^2)] \\
&\qquad \times [(U_s^2 - \sigma_0^2) + (\hat{U}_s^2 - U_s^2) - (\hat{\sigma}_0^2 - \sigma_0^2)]\}K_{ts} \\
&= (n^2a^p)^{-1}\sum_t \sum_{s\neq t} (U_t^2 - \sigma_0^2)(U_s^2 - \sigma_0^2)K_{ts} + R_n \equiv J_n + R_n,
\end{aligned}$$

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

where $J_n = (n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2 - \sigma_0^2)(U_s^2 - \sigma_0^2)K_{ts}$ and $R_n = J_n - \hat{J}_n$. In Appendix A, we show that $R_n = o_p((na^{p/2})^{-1})$. Thus, $na^{p/2}\hat{J}_n = na^{p/2}J_n + o_p(1)$. Obviously J_n is a second order degenerate U-statistic. We can therefore, use Lemma 3.1 to derive the asymptotic distribution of J_n. Let $\epsilon_t = U_t^2 - \sigma_0^2$. It is easy to check that (C1)–(C4) imply the conditions (A1) and (A2) of Appendix A. Hence,

$na^{p/2}\hat{J}_n \to N(0,\Omega)$ in distribution, by Lemma 3.1. ■

The consistency of the test $\hat{J}_n$ follows from the next theorem.

THEOREM 3.3. *Under* (C1)–(C4), *if* H_1 *is true*, *then*

$$\text{Prob}[na^{p/2}\hat{J}_n/\sqrt{\hat{\Omega}} > B_n] \to 1$$

for any nonstochastic sequence $B_n = o(na^{p/2})$.

Proof. Using (C1)–(C4), one can show by using arguments similar to those we used in the proof of Theorem 3.2 that, under H_1, (i) $\hat{J}_n = J + o_p(1)$, where $J = E[(E(U_t^2|X_t) - \sigma_0^2)^2 f(X_t)] > 0$, and (ii) $\hat{\Omega} = O_p(1)$. Theorem 3.3 follows from (i) and (ii). ■

Theorems 3.2 and 3.3 show that the $\hat{I}_n$ test is asymptotically standard normal under H_0 and it diverges to $+\infty$ under H_1. Therefore, $\hat{I}_n$ is a one-sided test. In practice one rejects H_0 if $\hat{I}_n > Z_\alpha$, where Z_α is the αth percentile of a standard normal variate and α is the level of the test.

The last result of this section is to derive the asymptotic distribution of $\hat{I}_n$ under Pitman local alternatives. The Pitman local alternative in our context is given by

$$H_L: E(U^2|X) = \sigma_0^2 + \frac{1}{n^{1/2}a^{p/4}}\,g(x) \quad \text{a.e.} \quad \text{for some non-negative function } g(\cdot).$$

THEOREM 3.4. *Under similar conditions as in Theorem* 3.2, *and also assuming that* $g^2(x)f(x)$ *is a bounded function*, *under* H_L

$na^{p/2}\hat{J}_n \to N(\mu,\Omega)$ *in distribution*,

where $\mu = E[g^2(X)f(X)]$.

The proof of Theorem 3.4 is given in Appendix A.

Theorem 3.4 shows that the $\hat{J}_n$ test can detect Pitman (regular or low frequency) local alternatives that approach the null model at a rate of $O_p((na^{p/2})^{-1/2})$, which has an order larger than $O_p(n^{-1/2})$. Therefore, unlike the usual parametric tests for heteroskedasticity, the $\hat{J}_n$ test cannot detect Pitman local alternatives that approach the null at the rate of $O_p(n^{-1/2})$. However, following the proof of Fan and Li (2000), one can show that the $\hat{J}_n$ test

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

can detect certain "singular" (high frequency) local alternatives that approach the null at a rate arbitrarily close to $O_p(n^{-3/4}) = o_p(n^{-1/2})$, whereas parametric tests cannot detect high frequency local alternatives that approach the null at a rate of $o_p(n^{-1/2})$.

4. THE BOOTSTRAP METHOD

The simulations in Section 7 show that in finite samples the $\hat{J}_n$ test is severely undersized when using asymptotic standard normal critical values. We therefore suggest using some bootstrap methods to approximate the null distribution of the test statistic $\hat{J}_n$ in finite samples. We note that we cannot use block resampling methods because the error terms from the bootstrap method will then have the same property as the original errors U_t, which will not be conditional homoskedastic under H_1.[3] We shall also restrict ourselves to the case in which W_t and X_t contain only strictly exogenous variables and predetermined variables (the previous periods' errors are also considered predetermined when there is no Granger feedback). In the case in which $\{W_t, X_t\}$ are strictly exogenous, we will use the standard bootstrap method. When W_t and/or X_t contain lagged dependent variables, we will use the recursive resampling method.

For ease of exposition we consider a linear regression model ($g(W_t, \gamma_0) = W_t'\gamma_0$). Suppose that

$$Y_t = W_t'\gamma_0 + U_t. \tag{4}$$

The exact bootstrap method is presented in two parts. We first assume that there is a procedure of generating the bootstrap sample $\{W_t^*, X_t^*\}_{t=1}^n$ and describe how to obtain the bootstrap test statistic. We will then discuss how to generate the bootstrap sample $\{W_t^*, X_t^*\}_{t=1}^n$. Let $\mathcal{F}_n$ denote $\{(Y_t, W_t, X_t)\}_{t=1}^n$ and denote $E^*(.) = E(.|\mathcal{F}_n)$. A general bootstrap method is as follows.

Bootstrap steps of generating $\hat{I}_n^*$ conditional on $\{W_t^*, X_t^*\}_{t=1}^n$.

Step (i). Obtain the least squares residual $\hat{U}_t = Y_t - W_t'\hat{\gamma}$, where $\hat{\gamma}$ is the least squares estimator of γ_0 based on model (4).

Step (ii). For $t = 1, \ldots, n$, obtain the bootstrap error U_t^* by random resampling with replacement from $\{\hat{U}_s - \bar{U}\}_{s=1}^n$, where $\bar{U} = n^{-1}\sum_{t=1}^n \hat{U}_t$. Let (W_t^*, X_t^*) be a bootstrap sample obtained by some specific method (which will be discussed subsequently). Then generate $Y_t^* = W_t^*\hat{\gamma} + U_t^*$, $t = 1, \ldots, n$.

Step (iii). Regress Y_t^* on W_t^* to obtain the least squares estimator $\hat{\gamma}^*$ and get the bootstrap residual $\hat{U}_t^* = Y_t^* - W_t^{*\prime}\hat{\gamma}^*$ for $t = 1, \ldots, n$. Compute the bootstrap statistic $\hat{I}_n^* = na^{p/2}\hat{J}_n^*/\sqrt{\hat{\Omega}^*}$, where $\hat{J}_n^*$ and $\hat{\Omega}^*$ are the same as $\hat{J}_n$ and $\hat{\Omega}$ except that $(\hat{U}_t, X_t)$ is replaced by $(\hat{U}_t^*, X_t^*)$.

Step (iv). Repeat step (ii) to (iii) a number of times, say, $(B-1)$ times. These bootstrap test statistics together with the original test statistic $\hat{I}_n = na^{p/2}\hat{J}_n/\sqrt{\hat{\Omega}}$ form the bootstrap distribution of $\{\hat{I}_{n(j)}^*\}_{j=1}^B$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Let C_α be the upper α percentile of the preceding bootstrap distribution. We will reject H_0 at significant level α if $\hat{I}_n > C_\alpha$. Heuristically, steps (ii) and (iii) ensure that conditional on the random sample $\mathcal{F}_n = \{Y_t, W_t, X_t\}_{t=1}^n$, the bootstrap sample $\{Y_t^*, W_t^*, Z_t^*\}_{t=1}^n$ is generated by the null model because conditional on $\mathcal{F}_n$, U_t^* has mean zero and a conditional homoskedastic variance of $n^{-1}\sum_t(\hat{U}_t - \bar{U})^2$ for all $t = 1,\ldots,n$. Therefore, the bootstrap distribution obtained in step (iii), conditional on the random sample $\mathcal{F}_n$, approximates the *null* distribution of the test statistic $\hat{I}_n$ even when the null hypothesis is false.

Now we discuss how to obtain the bootstrap sample $\{W_t^*, X_t^*\}_{t=1}^n$. Note that we cannot obtain $\{W_t^*, X_t^*\}_{t=1}^n$ using the block bootstrap method from $\{W_t, X_t\}_{t=1}^n$. This is because if $W_t = Y_{t-1}$, then we must have $W_t^* = Y_{t-1}^*$ (with high probabilities) for the bootstrap sample to mimic a data generating process (DGP) that is consistent with model (4). However, Y_{t-1}^* generated in step (ii) of the bootstrap steps of generating $\hat{I}_n^*$ is unlikely to be equal to W_t^* obtained from the block bootstrap method.

Bootstrap method of generating $\{W_t^*, X_t^*\}$. For ease of exposition, we assume that $W_t = (Y_{t-1}, W_{2,t})$ and $X_t = (Y_{t-1}, X_{2,t})$, where $W_{2,t}$ and $X_{2,t}$ are strictly exogenous. We write $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2')'$, where $\hat{\gamma}_1$ is the first component of $\hat{\gamma}$ and $\hat{\gamma}_2$ is the remaining components of $\hat{\gamma}$. We use the following recursive parametric bootstrap method.

Because W_{2t} and X_{2t} are strictly exogenous, we set $W_{2,t}^* = W_{2,t}$ and $X_{2,t}^* = X_{2,t}$. Randomly choose the initial Y_1^* from $N(\bar{Y}, \hat{\sigma}_y^2)$, where $\bar{Y}$ and $\hat{\sigma}_y$ are the sample mean and sample standard deviation of $\{Y_t\}_{t=1}^n$. Generate the remaining Y_t^* by $Y_t^* = \hat{\gamma}_1 Y_{t-1}^* + W_{2,t}'\hat{\gamma}_2 + U_t^*$ for $t = 2,\ldots,n$, where U_t^* is obtained as in step (i) of the bootstrap steps of generating $\hat{I}_n^*$. The bootstrap sample of $\{W_t^*, X_t^*\}$ is then obtained by $W_t^* = (Y_{t-1}^*, W_{2,t})$ and $X_t^* = (Y_{t-1}^*, X_{2,t})$, $t = 2,\ldots,n$.

The validity of the preceding bootstrap method is established in the next theorem.

THEOREM 4.1. *In addition to* (C1)–(C4), *we also assume that* (W_t, X_t) *only contain predetermined and strictly exogenous variables. Let* $\hat{I}_n^* = na^{p/2}\hat{J}_n^*/\sqrt{\hat{\Omega}^*}$. *Then we have*

$$\hat{I}_n^* \,|\, \mathcal{F}_n \to N(0,1) \quad \textit{in distribution.}$$

The proof of Theorem 4.1 is given in Appendix B.

Remark 4.1. We have made an additional assumption to prove Theorem 4.1. This condition requires that (W_t, X_t) are predetermined and/or strictly exogenous. When this assumption is violated, the result of Theorem 4.1 may not hold.

Remark 4.2. Theorem 4.1 only shows that the bootstrap method performs at least as well as the normal approximation with the error rate of $o_p(1)$ (like

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

the asymptotic test of Theorem 3.2). However, it is natural to expect that the bootstrap method should perform better than this. Our numerical experience strongly suggests that this is the case. For the independent data case, Fan and Linton (1997) and Li and Wang (1998) show that bootstrap methods indeed provide higher order accuracy for this type of tests. We conjecture that similar higher order accuracy improvement holds true for our test (with weakly dependent data). Unfortunately, we can not provide a not-too-lengthy proof for this conjecture.

5. THE CASE WITH GENERATED REGRESSORS

Up to now we have assumed that the conditional variable X_t is observable in testing the null of $E(U_t^2|X_t) = \sigma_0^2$. In this section we generalize the result of Theorems 3.2 and 3.3 to the case that X_t is not observable and has to be estimated. Many applications in econometrics involve generated regressors (e.g., Engle (1982), Hsiao and Mountain (1985), Pagan (1984), Pagan and Ullah (1988)). For expositional simplicity, we will only consider the linear regression model in this section:

$$Y_t = W_t'\gamma_0 + U_t, \tag{5}$$

where $W_t = (W_{1t},\ldots,W_{qt})' \in R^q$. Suppose that we are interested in testing whether the error is conditional homoskedastic or follows some type of ARCH error process. Then the null hypothesis of H_0 is $E(U_t^2|U_{t-1}) = \sigma_0^2$ a.e., and the alternative hypothesis is H_1: $E(U_t^2|U_{t-1}) = \theta(U_{t-1}) \neq \sigma_0^2$ on a set with positive measure (for some function $\theta(\cdot)$). Alternatively, the null hypothesis can be written as H_0^a: $E(U_t^2|Y_{t-1}, W_{t-1}') = \sigma_0^2$. We note that H_0^a implies H_0 because $U_{t-1} = Y_{t-1} - W_{t-1}'\gamma_0$ is a function of (Y_{t-1}, W_{t-1}'). Therefore, in principle, one can use the observable variables $(Y_{t-1}, W_{t-1}') \in R^{q+1}$ to construct a consistent test for H_0^a. However, it is better to construct a test statistic for H_0 based on a univariate U_{t-1} because the test for H_0^a using (Y_{t-1}, W_{t-1}') as the conditional variable suffers from the "curse of dimensionality" problem, especially when q is large.

To avoid introducing too many notations, in the remaining part of this section we will only consider the case of $X_t = (U_{t-1},\ldots,U_{t-p})'$. Obviously our result applies to other types of generated regressors. We will use $\hat{X}_t = (\hat{U}_{t-1},\ldots,\hat{U}_{t-p})'$ to replace the unobservable X_t in $\hat{J}_n$, where $\hat{U}_s = Y_t - W_t'\hat{\gamma}$ and $\hat{\gamma}$ is an $\sqrt{n}$-consistent estimator of γ based on (5) (say, the least squares estimator of γ). Thus, the test statistic becomes

$$\tilde{J}_n = \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} (\hat{U}_t^2 - \hat{\sigma}_0^2)(\hat{U}_s^2 - \hat{\sigma}_0^2)\hat{K}_{ts}, \tag{6}$$

where $\hat{K}_{ts} = K((\hat{X}_t - \hat{X}_s)/a)$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

To derive the asymptotic distribution of $\tilde{J}_n$, we need the following additional assumption.

(C5) The kernel function $K(\cdot)$ has third order continuous derivatives; $K(\cdot)$ and its derivatives (up to third order) are all bounded and integrable. Moreover, $K(\cdot)$ satisfies a Lipschitz-type condition: $|K(u + v) - K(u) - Q(u,v)| \leq H(u)\|v\|^4$ for all $u, v \in R^p$, where Q is a third-degree homogeneous polynomial in $u - v$ with coefficients the partial derivatives of orders one through three, $H(u)$ is bounded and $\int H(u)du < \infty$.

THEOREM 5.1. *In addition to* (C1)–(C5), *we assume that* W_t *has finite sixth moments and* $na^{4-p/2} \to \infty$ *as* $n \to \infty$. *Then under* H_0

$$\tilde{I}_n = na^{p/2}\tilde{J}_n/\sqrt{\tilde{\Omega}} \to N(0,1) \quad \textit{in distribution,}$$

where $\tilde{\Omega} = 2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(\hat{U}_t^2 - \hat{\sigma}_0^2)^2(\hat{U}_s^2 - \hat{\sigma}_0^2)^2\hat{K}_{ts}^2$ *is a consistent estimator of* Ω. *The term* Ω *is defined in Theorem* 3.2.

The proof of Theorem 5.1 is given in Appendix C. Theorem 5.1 generalizes the result of Theorem 3.2 to the case in which the conditional variables are generated regressor, and it shows that the conclusion of Theorem 3.2 remains unchanged when the conditional variables contain generated regressors.

Remark 5.1. The condition of $na^{4-p/2} \to \infty$ in Theorem 5.1 is imposed because we assume that the kernel function admits a Taylor expansion with the remainder term of fourth order, i.e., the remaining term in the Taylor expansion of $\hat{K}_{ts}$ has a factor $\|\hat{\gamma} - \gamma_0\|^4/a^4 = O_p(n^{-2}a^{-4})$. We need this term times the multiplicative constant $na^{p/2}$ to be $o_p(1)$, i.e., $na^{p/2}/(n^2a^4) = o(1)$, which is equivalent to $na^{4-(p/2)} \to \infty$. For a detailed proof of this, see Appendix B. This condition can be weakened to $(n^{(m-1)}a^{2m-(p/2)}) \to \infty$ for some integer $m \geq 3$ provided that $K(\cdot)$ has a Taylor expansion up to order $(m + 1)$ and further strengthens the moments condition on W_t.

The consistency of the test $\tilde{J}_n$ follows from the next theorem.

THEOREM 5.2. *Under* (C1)–(C5), *if* H_1^b *is true*, *then*

$$\text{Prob}[na^{p/2}\tilde{J}_n/\sqrt{\hat{\Omega}} > B_n] \to 1$$

for any nonstochastic sequence $B_n = o(na^{p/2})$.

The proof of Theorem 5.2 follows similar arguments to those used in the proof of Theorem 3.3 and is omitted here.

Remark 5.2. For the choice of $X_t = U_{t-1}$, our $\tilde{J}_n$ test is obviously consistent against an ARCH(1) alternative process. In fact, the same choice of $X_t = U_{t-1}$ can be used to test the alternative of a generalized autoregressive conditional heteroskedasticity (GARCH) process. To see this, consider a GARCH(1,1) pro-

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

cess of $U_t = V_t\sqrt{h_t}$, where V_t is an i.i.d. process with $E(V_t) = 0$ and $\text{var}(V_t) = 1$, $h_t = E(U_{t-1}^2 | \mathcal{M}_{-\infty}^{t-1}(U)) = \alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 h_{t-1}$ with $\alpha_0 \geq 0$, $\alpha_1 > 0$, and $\alpha_2 > 0$.[4] Then for $X_t = U_{t-1}$ we have

$$\begin{aligned} E(U_t^2 | U_{t-1}) &= E(h_t | U_{t-1}) = \alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 E(h_{t-1}) \\ &= \alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 \sigma_0^2 \\ &\neq \sigma_0^2 \equiv E(U_t^2). \end{aligned}$$

Obviously, a GARCH(1,1) process belongs to the alternative hypothesis (model) H_1, and this would imply that $\tilde{J}_n$ is a consistent test against a GARCH(1,1) process provided one can show that the conditions of Theorem 5.2 hold for the GARCH(1,1) process. However, it is not clear that a GARCH(1,1) process satisfies the conditions of Theorem 5.2. In particular it is currently unknown under what conditions, a GARCH(1,1) process is β-mixing. However, the conditions in Theorem 5.2 are sufficient (not necessary) conditions. It is conceivable that $\tilde{J}_n$ can still be a consistent test with much less restrictive assumptions.

Remark 5.3. We assume that (Y_t, W_t, X_t) is a strictly stationary process in establishing Theorems 5.1 and 5.2. As pointed out by a referee, the assumption that Y_t is strictly stationary can be relaxed to, say, a unit root process: $Y_t = Y_{t-1}\gamma_0 + U_t$ with $\gamma_0 = 1$ and U_t a strictly stationary martingale difference process. However, we also need the conditional variable X_t to be strictly stationary to use Lemma 3.1. Therefore, X_t cannot contain lagged values of Y_t if Y_t follows a unit root process. One can show that Theorems 5.1 and 5.2 hold when Y_t is a unit root process provided (U_t, X_t) satisfy the conditions given in (C1)–(C5) (say, $X_t = U_{t-1}$). This is because our proof of Theorem 5.1 mainly relies on the fact that $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$. When Y_t follows a unit root process, we have $\hat{\gamma} - \gamma_0 = O_p(n^{-1})$ (an even faster rate of convergence), and hence the proof of Theorem 5.1 remains valid for a unit root process.

Finally we note that the bootstrap counterpart of $\tilde{I}_n \equiv na^{p/2}\tilde{J}_n$, say, $\tilde{I}_n^*$, can be obtained similarly to the bootstrap test $\hat{I}_n^*$ as described in Section 4. Specifically, let $\hat{U}_t^*$ and W_t^* be obtained the same way as in Section 4. Then $\hat{X}_t^* = (\hat{U}_{t-1}^*, \ldots, \hat{U}_{t-p}^*)$ and $\tilde{I}_n^*$ is obtained from $\hat{I}_n^*$ by replacing $K_{ts}^* = K((X_t^* - X_s)/a)$ by $\hat{K}_{ts}^* = K((\hat{X}_t^* - \hat{X}_s^*)/a)$. Similarly to the proof of Theorem 4.1, one can show that the bootstrap test $\tilde{I}_n^* | \mathcal{F}_n$ has an asymptotic standard normal distribution.

6. EXTENSION: TESTING A PARAMETRICALLY SPECIFIED CONDITIONAL VARIANCE

In this section we generalize our early results of testing conditional homoskedasticity to test the null hypothesis of a parametrically specified conditional variance.[5] Suppose that the conditional variance of U_t in model (1)

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

is conditional heteroskedastic with a known parametric functional form, say, $E(U_t^2|X_t) = h(X_t, \alpha)$, where the functional form of $h(\cdot, \alpha)$ is known and α is a $d \times 1$ vector of unknown parameter. Then we may formulate our null hypothesis as

$$H_0^b : P[E(U^2|X) = h(X, \alpha)] = 1 \quad \text{for some } \alpha \in \mathcal{D} \in R^d,$$

where $\mathcal{D}$ is a compact subset of R^d, and

$$H_1^b : P[E(U^2|X) = h(X, \alpha)] < 1 \quad \text{for all } \alpha \in \mathcal{D},$$

where as before, $X \in R^p$, which may be different from the regressor W. Define $\eta_t = U_t^2 - h(X_t, \alpha)$;[6] then $E(\eta_t|X_t) = 0$ almost everywhere if and only if H_0^b is true. We first consider the case that X_t is observable.

(I) The Case of Observed X_t. Let $\hat{\alpha}$ be an estimator of α. Our test statistic is obtained from $\hat{J}_n$ by replacing $\hat{\sigma}_0^2$ by $h(X_t, \hat{\alpha})$:

$$\hat{J}_{n,b} = \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} \hat{\eta}_t \hat{\eta}_s K_{ts}$$

$$\equiv \frac{1}{(n^2 a^p)} \sum_t \sum_{s \neq t} (\hat{U}_t^2 - h(X_t, \hat{\alpha}))(\hat{U}_s^2 - h(X_t, \hat{\alpha}))K_{ts}, \tag{7}$$

where $\hat{\eta}_t = \hat{U}_t^2 - h(X_t, \hat{\alpha})$.

An additional assumption is needed to establish the asymptotic distribution of $\hat{J}_{n,b}$.

(C6)

(i) $\hat{\alpha} - \alpha = O_p(n^{-1/2})$ under H_0^b,
(ii) $\nabla h(x, \cdot)$ and $\nabla^2 h(x, \cdot)$ are continuous and dominated by functions $D_1(x)$ and $D_2(x)$, respectively. The terms $D_1(x)$ and $D_2(x)$ have finite fourth moments, where $\nabla h(x, \cdot)$ and $\nabla^2 h(x, \cdot)$ are $d \times 1$ vector of first order partial derivatives and $d \times d$ matrix of second order partial derivatives of h with respect to α.

The following theorem gives the asymptotic null distribution of $\hat{J}_{n,b}$.

THEOREM 6.1. *Suppose that Assumptions* (C1)–(C6) *hold. Then under* H_0^b,

$$na^{p/2} \hat{J}_{n,b} \Big/ \sqrt{\hat{\Omega}_b} \to N(0,1) \quad \textit{in distribution},$$

where $\hat{\Omega}_b = 2(n^2 a^p)^{-1} \sum_t \sum_{s \neq t} [\hat{U}_t^2 - h(X_t, \hat{\alpha})]^2 [\hat{U}_s^2 - h(X_s, \hat{\alpha})]^2 K_{t,s}^2$ *is a consistent estimator for* $\Omega_b = 2E[\sigma_\eta^4(X_t) f(X_t)][\int K^2(u)\, du]$, *where* $\eta_t = U_t^2 - h(X_t, \alpha)$ *and* $\sigma_\eta^2(X_t) = E[\eta_t^2 | X_t]$.

Proof. Taking a Taylor expansion, we have $h(X_t, \hat{\alpha}) = h(X_t, \alpha) + (\hat{\alpha} - \alpha)' \times \nabla h(X_t, \alpha) + (\hat{\alpha} - \alpha)' \nabla^2 h(X_t, \tilde{\alpha})(\hat{\alpha} - \alpha)/2$, where $\tilde{\alpha}$ is in the line segment

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

between $\hat{\alpha}$ and α. Define the shorthand notation: $A_{0t} = h(X_t, \alpha)$, $A_{1t} = \nabla h(X_t, \alpha)$, and $\tilde{A}_{2t} = \nabla^2 h(X_t, \tilde{\alpha})/2$. Then

$$\hat{\eta}_t = \hat{U}_t^2 - h(X_t, \hat{\alpha}) = \hat{U}_t^2 - \{A_{01} + (\hat{\alpha} - \alpha)' A_{1t} + (\hat{\alpha} - \alpha)' \tilde{A}_{2t} (\hat{\alpha} - \alpha)\}. \quad \textbf{(8)}$$

Using (8) and the fact that $(\hat{\alpha} - \alpha) = O_p(n^{-1/2})$ under H_0^b, we have

$$\begin{aligned}
\hat{J}_{n,b} &= \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} \hat{\eta}_t \hat{\eta}_s K_{ts} \\
&\equiv \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} \{(\hat{U}_t^2 - A_{0t})(\hat{U}_s^2 - A_{0s}) \\
&\qquad + 2(\hat{\alpha} - \alpha)'(\hat{U}_t^2 - A_{0t}) A_{1t}\} K_{ts} + O_p(n^{-1}) \\
&\equiv \hat{J}_{nb,1} + 2(\hat{\alpha} - \alpha)' \hat{J}_{nb,2} + O_p(n^{-1}),
\end{aligned}$$

where the definitions of $\hat{J}_{nb,1}$ and $\hat{J}_{nb,2}$ should be apparent and the $O_p(n^{-1})$ term corresponds to terms that contain at least two $(\hat{\alpha} - \alpha)$'s. Here $na^{p/2} \hat{J}_{nb,1} \to N(0, \Omega_b)$ in distribution (under H_0^b) by the same proof as Theorem 3.2.

By adding and subtracting terms in $\tilde{J}_{n,b}$, we have

$$\hat{J}_{nb,2} = (n^2 a^p)^{-1} \sum_t \sum_{s \neq t} [(\hat{U}_t^2 - U_t^2) + (U_t^2 - h_t)] K_{ts} \equiv G_{1n} + G_{2n}.$$

Note that $\hat{U}_t^2 - U_t^2 = (\hat{U}_t - U_t)(\hat{U}_t + U_t)$ and $\hat{U}_t - U_t$ contains a factor of $(\hat{\gamma} - \gamma_0) = O_p(n^{-1/2})$ (see equation (A.1) in Appendix A). One can easily show that $G_{1n} = O_p(n^{-1/2})$; $G_{2n} = o_p((na^p)^{-1/2})$ by Lemma A.1. Therefore, $\hat{J}_{nb,2} = O_p(n^{-1/2}) + o_p((na^p)^{-1/2}) = o_p((na^p)^{-1/2})$ and $(\hat{\alpha} - \alpha)' \hat{J}_{nb,2} = O_p(n^{-1/2}) o_p((na)^{-1/2}) = o_p((na^{p/2})^{-1})$. Hence, $na^{p/2} \hat{J}_{n,b} \to N(0, \Omega_b)$ in distribution.

Finally $\hat{\Omega}_b = \Omega_b + o_p(1)$ follows the same arguments as the proof of $\hat{\Omega} - \Omega = o_p(1)$. This completes the proof of Theorem 6.1. ■

By a similar argument as in the proof of Theorem 3.3, one can easily show that $\hat{J}_{n,b}$ is a consistent test.

(II) The Case of Generated Regressors. Similarly to the procedure in Section 5 we consider the case $X_t = (U_{t-1}, \ldots, U_{t-p})'$, which is not observable. We estimate X_t by $\hat{X}_t = (\hat{U}_{t-1}, \ldots, \hat{U}_{t-p})'$. In this case our test statistic is

$$\tilde{J}_{n,b} = \frac{1}{n^2 a^p} \sum_t \sum_{s \neq t} (\hat{U}_t^2 - h(\hat{X}_t, \hat{\alpha}))(\hat{U}_s^2 - h(\hat{X}_t, \hat{\alpha})) \hat{K}_{ts}, \quad \textbf{(9)}$$

where as in Section 5, $\hat{K}_{ts} = K((\hat{X}_t - \hat{X}_s)/a)$. The next theorem gives the asymptotic null distribution of $\tilde{J}_{n,b}$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

THEOREM 6.2. *Suppose that Assumptions* (C1)–(C6) *hold. Then under* H_0^b,

$na^{p/2}\tilde{J}_{n,b}/\sqrt{\tilde{\Omega}_b} \to N(0,1)$ in distribution,

where $\tilde{\Omega}_b = 2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}[\hat{U}_t^2 - h(\hat{X}_t,\hat{\alpha})]^2[\hat{U}_s^2 - h(\hat{X}_s,\hat{\alpha})]^2\hat{K}_{t,s}^2$ *is a consistent estimator for* Ω_b *and* Ω_b *is the same as defined in theorem* 6.1.

Proof. Taking a Taylor expansion of $K((\hat{X}_t - \hat{X}_s)/a)$ at $K((X_t - X_s)/a)$, and using the same arguments as in the proof of Theorem 5.1, it is straightforward to show, though tedious, that $\tilde{J}_{n,b} = \hat{J}_{nb,1} + o_p((na^{d/2})^{-1})$, where

$$\hat{J}_{nb,1}(n^2a^p)^{-1}\sum_t\sum_{s\neq t}[\hat{U}_t^2 - h(X_t,\hat{\alpha})][\hat{U}_s^2 - h(X_s,\hat{\alpha})]K((X_t - X_s)/a).$$

Theorem 6.2 then follows from Theorem 6.1. ■

Remark 6.1. Similarly to the proofs of Theorems 3.3 and 5.2, it can be shown that both $\hat{J}_{n,b}$ and $\tilde{J}_{n,b}$ are consistent tests, i.e., under H_1^b: $\text{Prob}[na^{p/2}\hat{J}_{n,b}/\sqrt{\hat{\Omega}_b} > B_n] \to 1$ and $\text{Prob}[na^{p/2}\tilde{J}_{n,b}/\sqrt{\tilde{\Omega}_b} > B_n] \to 1$ for any nonstochastic sequence $B_n = o(na^{p/2})$.

Remark 6.2. Our test statistic $\tilde{J}_{n,b}$ can be used to test the null hypothesis of an ARCH(p) error process: $h(X_t,\alpha) = \alpha_0 + \alpha_1 U_{t-1}^2 + \ldots + \alpha_p U_{t-p}^2$. However, it does not cover the case of testing the null of a GARCH error process (e.g., Bollerslev, 1986; Bollerslev, Chou, and Kroner, 1992). Consider a GARCH(1,1) process, say, $U_t|\mathcal{M}_{t-1} \sim (0,h_t)$ with $h_t = \alpha_0 + \alpha_1 U_{t-1}^2 + \alpha_2 h_{t-1}$, where $h_t = E(U_t^2|\mathcal{M}_{-\infty}^{t-1})$. When $\alpha_2 > 0$, there does not exist a *finite* value of p such that $h_t = E(U_t^2|U_{t-1},\ldots,U_{t-p})$. Therefore, consistent testing of the null of a GARCH(1,1) error process is not covered in Theorems 6.1 and 6.2.

Remark 6.3. As suggested by a referee, it might be possible to modify our testing procedure to test the null of a GARCH(1,1) process by testing whether the standardized error $U_t/\sqrt{h_t}$ is conditional homoskedastic. Such an extension requires some additional work for the following reasons.

(i) One needs to establish the convergence rate of $\hat{h}_t$ to h_t, where $\hat{h}_t$ is a consistent estimator of h_t under H_0^b, say, the quasi maximum likelihood estimator as in Lee and Hansen (1994). Lee and Hansen have established the $\sqrt{n}$-normality results for the estimators of $\alpha = (\alpha_0,\alpha_1,\alpha_2)$ for a GARCH(1,1) model ($\hat{\alpha} - \alpha = O_p(n^{-1/2})$) and established uniform convergence of $\hat{h}_t$ to h_t. The results of Lee and Hansen should be useful in deriving the convergence rate of $\hat{h}_t$ to h_t.

(ii) To use Lemma 3.1, one needs to show that under some conditions a GARCH(1,1) process is β-mixing. Hansen (1991) has shown that the GARCH(1,1) process is near epoch dependent under quite general conditions. However, it is currently unknown under what conditions a GARCH(1,1) process is β-mixing.

(iii) Finally one needs to choose a proper conditional variable X_t for the test to be consistent, i.e., to have asymptotic power one against any other form (other than GARCH(1,1)) of conditional heteroskedasticity. One possibility is to choose

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

$X_t = (U_{t-1}, h_{t-1})$ However, this choice of X_t may not lead to a consistent test because under H_1^b the probability limit of $\hat{h}_{t-1}$ is not clear. Another possible choice is $X_t = (U_{t-1}, U_{t-2}, \ldots, U_{t-p})$. Again, it may not lead to a consistent test because the conditional set does not include U_{t-s} for $s > p$. A third choice is to choose $(U_{t-1}, U_{t-2}, \ldots, U_{t-p_n})$ as the conditional variables, where p_n is not fixed, $p_n \to \infty$ as $n \to \infty$ but at a slower rate than n (e.g., $p_n/n \to 0$). Such a choice may lead to a consistent test; however, the asymptotic analysis will be quite complex (e.g., de Jong, 1996).

Therefore, a rigorous treatment of this extension is beyond the scope of the paper and is left for future research.

7. MONTE CARLO RESULTS

We report some Monte Carlo results to study the finite sample performances of our proposed tests in this section. Section 7.1 studies the finite sample performance of the $\hat{J}_n$ test with observable conditional variables, and Section 7.2 examines the finite sample behavior of the $\tilde{J}_n$ test where the conditional variables are generated regressors.

7.1. The $\hat{J}_n$ Test (Observed X)

In this section we report some Monte Carlo simulation results of the test statistic $\hat{I}_n = na^{p/2}\hat{J}_n/\sqrt{\hat{\Omega}}$ using both asymptotic and bootstrap critical values. We use the following DGP's:

$$DGP1: \quad Y_t = 0.5Y_{t-1} + U_t,$$

$$DGP2: \quad Y_t = 1 + X_t + U_t,$$

$$DGP3: \quad Y_t = 0.5Y_{t-1} + U_t\eta_t,$$

$$DGP4: \quad Y_t = 1 + X_t + U_t\eta_t,$$

where $X_t = 0.5X_{t-1} + V_t$ and U_t and V_t are independent processes, both of which are i.i.d. $N(0,1)$. Here $\eta_t = \sqrt{(0.2 + 0.8Z_t^2)/c}$, where $Z_t = Y_{t-1}$ for DGP3 and $Z_t = X_t$ for DGP4, $c = 0.2 + 0.8/(1 - 0.5^2)$ so that the unconditional error variance is one. DGP1 and DGP2 correspond to the null hypothesis of conditional homoskedastic error, i.e., H_0 is true, where DGP3 and DGP4 correspond to H_1 of conditional heteroskedastic errors with the conditional variables being the same as the regressors. Note that DGP1 is tested against DGP3 (DGP1 as H_0 and DGP3 as H_1) and DGP2 is tested against DGP4.

We use a standard normal kernel function with smoothing parameter chosen via $a = cy_{-1,sd}n^{-1/5}$ for DGP1 and DGP3 and $a = cx_{sd}n^{-1/5}$ for DGP2 and DGP4, respectively, where $y_{-1,sd}$ and x_{sd} are sample standard deviations of

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

$\{Y_{t-1}\}_{t=2}^{n}$ and $\{X_t\}_{t=1}^{n}$, respectively. We use $c = 0.5,1,2$. Our test statistic is the standardized test $\hat{I}_n = na^{1/2}\hat{J}_n/\sqrt{\hat{\Omega}}$, where

$$\hat{J}_n = \frac{1}{n^2 a}\sum_t \sum_{s\neq t}(\hat{U}_t^2 - \hat{\sigma}_0^2)(\hat{U}_s^2 - \hat{\sigma}_0^2)K\left(\frac{Z_t - Z_s}{a}\right),$$

and $\hat{\Omega} = 2(n^2 a)^{-1}\sum_t \sum_{s\neq t}(\hat{U}_t^2 - \hat{\sigma}_0^2)^2(\hat{U}_s^2 - \hat{\sigma}_0^2)^2 K^2((Z_t - Z_s)/a)$ and $Z_t = Y_{t-1}$ for DGP1 and DGP3 and $Z_t = X_t$ for DGP2 and DGP4. To save space, we only report the results of $c = 0.5$ and $c = 1$ subsequently, and we will briefly mention the result for $c = 2$.

For the $\hat{J}_n$ test using the asymptotic standard normal critical values, the number of replication is 2,000. For the $\hat{J}_n$ test based on the wild bootstrap critical values (the procedure is described in Section 4), the number of replication is 1,000, and within each replication, 399 bootstrap test statistics are generated to give the 1%, 5%, and 10% bootstrap critical values. To compare the performances of our $\hat{J}_n$ test against some popular parametric tests, we also compute a Lagrange multiplier (LM) test that tests $\alpha_1 = 0$ in the following regression equation:

$$\hat{U}_t^2 = \alpha_0 + \alpha_1 Z_t^2 + \text{error},$$

where $\hat{U}_t$ is the least squares residual, $Z_t = Y_{t-1}$ for DGP1 and DGP3, and $Z_t = X_t$ for DGP2 and DGP4.

The simulation results for DGP1 are given in Table 1. From Table 1 we observe that the $\hat{J}_n$ test based on the asymptotic standard normal critical values underestimates the nominal sizes for most cases considered. The estimated sizes seem to converge to their nominal sizes as n increases, however, at a fairly slow rate. This is understandable because the rate that $\hat{J}_n$ converges to a standard normal variate is at best of the order of $O_p(a^{p/2})$ (see, e.g., Li and Wang, 1998). Overall, the estimated sizes are closer to their nominal sizes for $c = 0.5$ than for $c = 1$. The estimated sizes for $c = 2$ are even more downsized than for $c = 1$ (which is not reported here to save space).

For the $\hat{J}_n$ test based on bootstrap critical values, we observe that, for both $c = 0.5$ and $c = 1$, the estimated sizes are quite close to their nominal sizes. Thus, the bootstrap procedure provides much better finite sample approximations to the null distributions of the $\hat{J}_n$ test. Also the bootstrap test is much less sensitive to the different choices of the smoothing parameter than the asymptotic test. The parametric LM test has good estimated sizes as expected.

Table 2 gives the estimation result for DGP2. The result is very similar to that of Table 1. The asymptotic $\hat{J}_n$ is sensitive to the different choices of a and is quite undersized. On the other hand, the bootstrap test has much better estimated sizes.

Tables 3 and 4 give the estimated powers for DGP3 and DGP4, respectively. From Tables 3 and 4 we observe that the $\hat{J}_n$ test based on the bootstrap test is more powerful than the test using asymptotic critical values. This is because

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Table 1. Size of DGP1

	J_n (asymptotic)										J_n (bootstrap)										
	$c = 1.0$					$c = 0.5$					$c = 1.0$			$c = 0.5$			LM				
n	1%	5%	10%	Mean	Std.	1%	5%	10%	Mean	Std.	1%	5%	10%	1%	5%	10%	1%	5%	10%	Mean	Std.
50	0.004	0.011	0.023	−0.581	0.732	0.003	0.02	0.034	−0.521	0.858	0.008	0.039	0.077	0.01	0.042	0.087	0.013	0.034	0.061	−0.29	0.951
100	0.008	0.019	0.031	−0.483	0.786	0.009	0.021	0.038	−0.411	0.868	0.014	0.047	0.093	0.014	0.049	0.092	0.01	0.033	0.062	−0.235	0.935
200	0.005	0.018	0.03	−0.447	0.766	0.009	0.029	0.048	−0.36	0.897	0.012	0.047	0.096	0.014	0.05	0.097	0.013	0.04	0.077	−0.132	0.945
500	0.008	0.024	0.035	−0.408	0.8	0.015	0.04	0.06	−0.294	0.928							0.011	0.046	0.084	−0.09	0.988
1,000	0.008	0.029	0.046	−0.351	0.832	0.01	0.028	0.051	−0.296	0.865							0.017	0.056	0.1	0.056	0.105

Table 2. Size of DGP2

	J_n (asymptotic)										J_n (bootstrap)										
	$c = 1.0$					$c = 0.5$					$c = 1.0$			$c = 0.5$			LM				
n	1%	5%	10%	Mean	Std.	1%	5%	10%	Mean	Std.	1%	5%	10%	1%	5%	10%	1%	5%	10%	Mean	Std.
50	0.004	0.015	0.025	−0.555	0.74	0.006	0.019	0.037	−0.456	0.894	0.011	0.051	0.102	0.01	0.056	0.108	0.038	0.075	0.097	−0.065	1.056
100	0.006	0.015	0.032	−0.458	0.768	0.007	0.029	0.049	−0.367	0.89	0.011	0.055	0.106	0.01	0.055	0.107	0.027	0.059	0.089	−0.084	1.01
200	0.009	0.022	0.035	−0.434	0.792	0.009	0.028	0.044	−0.357	0.883	0.012	0.054	0.103	0.099	0.052	0.104	0.025	0.062	0.099	−0.044	0.081
500	0.009	0.025	0.043	−0.398	0.814	0.011	0.03	0.052	−0.31	0.897							0.023	0.058	0.088	−0.063	1.003
1,000	0.009	0.023	0.044	−0.363	0.811	0.012	0.037	0.06	−0.292	0.906							0.022	0.06	0.099	−0.041	1.002

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

TABLE 3. Power of DGP3

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	15%	10%	1%	5%	10%	1%	5%	10%
50	0.601	0.665	0.804	0.521	0.625	0.669	0.605	0.723	0.881	0.553	0.681	0.738	0.531	0.698	0.783
100	0.875	0.911	0.934	0.844	0.902	0.921	0.96	0.99	0.994	0.854	0.921	0.95	0.838	0.93	0.96

TABLE 4. Power of DGP4

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.667	0.821	0.882	0.627	0.807	0.871	0.77	0.924	0.949	0.704	0.884	0.933	0.659	0.817	0.901
100	0.984	0.995	0.997	0.966	0.988	0.996	0.99	1	1	0.984	0.998	1	0.873	0.96	0.981

the asymptotic test is biased toward accepting the null (i.e., it is undersized). Therefore, one should expect it will suffer some power loss under the alternatives. We also observe that the $\hat{J}_n$ test is as powerful as the LM test. This result may seem surprising because usually one would expect that a correctly specified parametric LM test should be more powerful than a nonparametric test such as $\hat{J}_n$. As we have shown in Section 3, the $\hat{J}_n$ test can detect Pitman local alternatives that approach the null at the rate of $(n^{1/2}a^{p/4})^{-1} = (na^{1/2})^{-1/2}$, whereas it is well known that a parametric LM test can detect local alternatives that approach the null at a faster rate of $n^{-1/2}$. However a more powerful test against certain local alternatives does not necessarily imply that it will be more powerful against fixed alternatives in finite sample applications. Zheng (1996) also observed that a nonparametric test similar to $\hat{J}_n$ had good finite sample power performances when compared with a correctly specified *t*-test. Our simulation results show that the $\hat{J}_n$ is quite powerful in detecting the conditional heteroskedastic error processes of DGP3 and DGP4.

7.2. The Case of Generated Regressors

This section reports Monte Carlo simulation results for the $\tilde{I}_n = na^{p/2}\tilde{J}_n/\sqrt{\tilde{\Omega}}$ test when the conditional variable is not observable. We consider the following DGP's:

$$DGP5: \quad Y_t = 0.5Y_{t-1} + U_t,$$

$$DGP6: \quad Y_t = 1 + X_t + U_t,$$

$$DGP7: \quad Y_t = 0.5Y_{t-1} + U_t\eta_t,$$

$$DGP8: \quad Y_t = 0.5Y_{t-1} + U_t\epsilon_t,$$

$$DGP9: \quad Y_t = 1 + X_t + U_t\eta_t,$$

$$DGP10: \quad Y_t = 1 + X_t + U_t\epsilon_t,$$

where X_t, U_t are generated in the same ways as in DGP1 and DGP2. Thus, DGP5 and DGP6 are the null model with conditional homoskedastic errors. Here η_t follows an ARCH(1) process (Engle, 1982) with $\eta_t = \sqrt{0.5 + 0.5U_{t-1}^2}$ and ϵ_t is determined via $\epsilon_t = \exp\{0.2 + 0.8[|U_{t-1}| - E|U_{t-1}|]\}$, which is similar to an EARCH(1) error process (Nelson, 1991). Hence DGP7–DGP10 correspond to the alternative models with conditional heteroskedastic errors.

The errors are normalized so that the unconditional error variances all equal to one. We use a standard normal kernel function with smoothing parameter chosen via $a = c\hat{U}_{-1,sd}n^{-1/5}$, where $\hat{U}_{-1,sd}$ is the sample standard deviation of $\{\hat{U}_{t-1}\}_{t=2}^n$. Thus the test statistic is

$$\tilde{J}_n = \frac{1}{n^2a}\sum_t\sum_{s\neq t}(\hat{U}_t^2 - \hat{\sigma}_0^2)(\hat{U}_s^2 - \hat{\sigma}_0^2)K\left(\frac{\hat{U}_{t-1} - \hat{U}_{s-1}}{a}\right),$$

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

the standardized test is $\tilde{I}_n = na^{1/2}\tilde{J}_n/\sqrt{\tilde{\Omega}}$, which has an asymptotic standard normal distribution under H_0, and $\tilde{\Omega} = 2(n^2a)^{-1}\sum_t\sum_{s\neq t}(\hat{U}_t^2 - \hat{\sigma}_0^2)^2(\hat{U}_s^2 - \hat{\sigma}_0^2)^2K^2((\hat{U}_{t-1} - \hat{U}_{s-1})/a)$.

For comparison we also compute a LM test based on an ARCH(1) specification as suggested by Engle (1982).[7] The LM test statistic is based on testing $\alpha_1 = 0$ in a standard ARCH(1) model:

$$\hat{U}_t^2 = \alpha_0 + \alpha_1\hat{U}_{t-1}^2 + \text{error},$$

where $\hat{U}_t$ is the least squares residual.

The estimated sizes for DGP5 and DGP6 are reported in Tables 5 and 6, respectively. The results are very similar to the observed regressor cases of Tables 1 and 2; i.e., the $\tilde{J}_n$ test underestimates the nominal sizes for all cases considered, and the estimated sizes seem to converge to their nominal sizes as n increases, although at a fairly slow rate.

The parametric LM test has, not surprisingly, much better estimated sizes than the asymptotic nonparametric test $\tilde{J}_n$.

The $\tilde{J}_n$ test based on bootstrap critical values provides a much better null approximation than the asymptotic normal approximation. The estimated sizes are close to the nominal sizes for all cases, and they are much less sensitive to smoothing parameter values than the asymptotic test. The bootstrap $\tilde{J}_n$ test has better estimated sizes than that yielded by the (asymptotic) LM test.

Next we examine the power performances of the $\tilde{J}_n$ test and the LM test. Tables 7–10 give the estimated powers of the $\tilde{J}_n$ test (based on the asymptotic and the bootstrap critical values) and the LM test.

First we observe that the bootstrap $\tilde{J}_n$ test is significantly more powerful than the asymptotic $\tilde{J}_n$ test. This is due to the fact that the asymptotic $\tilde{J}_n$ test is biased toward accepting the null.

Next we observe that for both DGP7 and DGP9, for a correctly specified ARCH(1) error process, the LM test is more powerful than the nonparametric test $\tilde{J}_n$. This is because the LM test is a parametric test that is designed to detect an ARCH(1) alternative.

Third, for DGP8 and DGP10 error process, we observe that the nonparametric test $\tilde{J}_n$ is more powerful than the LM test using ARCH(1) formulation. The LM test is designed to have best power in the direction of an ARCH(1) alternative; therefore it will suffer power loss when the true alternative is not an ARCH(1) process. In contrast our nonparametric $\tilde{J}_n$ test is designed to detect *all* deviations from the null model. Therefore, it is likely that our $\tilde{J}_n$ test will be more powerful than the LM test under the alternative process that is not ARCH.

Finally, we notice that the $\tilde{J}_n$ test is more powerful against DGP8 (DGP10) than against DGP7 (DGP9). This means that DGP8 (DGP10) is further away than DGP7 (DGP9) from the null model using our distance measure, whereas the LM test is less powerful against DGP8 (DGP10) than against DGP7 (DGP9).

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

TABLE 5. Size of DGP5

	J_n (asymptotic)										J_n (bootstrap)										
	$c = 1.0$					$c = 0.5$					$c = 1.0$			$c = 0.5$			LM				
n	1%	5%	10%	Mean	Std.	1%	5%	10%	Mean	Std.	1%	5%	10%	1%	5%	10%	1%	5%	10%	Mean	Std.
50	0.002	0.011	0.016	−0.617	0.689	0.003	0.02	0.034	−0.521	0.858	0.011	0.04	0.077	0.013	0.044	0.083	0.012	0.039	0.067	−0.24	0.939
100	0.006	0.018	0.03	−0.522	0.793	0.009	0.021	0.038	−0.411	0.868	0.018	0.048	0.098	0.011	0.048	0.097	0.011	0.036	0.07	−0.165	0.943
200	0.005	0.021	0.035	−0.452	0.778	0.009	0.029	0.048	−0.36	0.897	0.014	0.051	0.102	0.012	0.052	0.101	0.018	0.051	0.095	−0.114	1
500	0.008	0.029	0.045	−0.398	0.824	0.015	0.04	0.06	−0.294	0.928							0.015	0.053	0.09	−0.055	0.999
1,000	0.007	0.027	0.041	−0.386	0.805	0.01	0.028	0.051	−0.296	0.865							0.008	0.043	0.093	−0.068	0.974

TABLE 6. Size of DGP6

	J_n (asymptotic)										J_n (bootstrap)										
	$c = 1.0$					$c = 0.5$					$c = 1.0$			$c = 0.5$			LM				
n	1%	5%	10%	Mean	Std.	1%	5%	10%	Mean	Std.	1%	5%	10%	1%	5%	10%	1%	5%	10%	Mean	Std.
50	0.001	0.011	0.025	−0.554	0.728	0.004	0.015	0.037	−0.461	0.873	0.007	0.057	0.109	0.008	0.055	0.104	0.012	0.042	0.076	−0.179	0.949
100	0.004	0.012	0.024	−0.472	0.754	0.005	0.015	0.039	−0.456	0.872	0.01	0.056	0.108	0.009	0.053	0.103	0.012	0.037	0.086	−0.079	0.971
200	0.007	0.017	0.029	−0.46	0.772	0.008	0.022	0.043	−0.367	0.883	0.009	0.054	0.105	0.012	0.049	0.098	0.015	0.05	0.091	−0.051	0.985
500	0.012	0.029	0.04	−0.394	0.791	0.01	0.031	0.053	−0.277	0.871							0.013	0.056	0.094	−0.036	0.992
1,000	0.009	0.031	0.052	−0.36	0.859	0.014	0.034	0.076	−0.285	0.939							0.012	0.057	0.101	−0.008	1.01

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

TABLE 7. Power of DGP7

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.208	0.287	0.408	0.192	0.282	0.42	0.275	0.438	0.528	0.231	0.398	0.477	0.328	0.501	0.601
100	0.517	0.622	0.683	0.472	0.581	0.675	0.577	0.738	0.803	0.512	0.666	0.745	0.666	0.812	0.87

TABLE 8. Power of DGP8

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.408	0.562	0.652	0.514	0.724	0.814	0.54	0.806	0.864	0.554	0.854	0.91	0.232	0.403	0.517
100	0.884	0.96	0.978	0.886	0.97	0.98	0.884	0.986	0.998	0.908	0.978	0.986	0.414	0.58	0.693

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

TABLE 9. Power of DGP9

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.226	0.295	0.362	0.212	0.306	0.36	0.29	0.443	0.544	0.251	0.394	0.476	0.35	0.492	0.577
100	0.538	0.634	0.701	0.48	0.605	0.662	0.595	0.751	0.82	0.528	0.685	0.758	0.671	0.824	0.874

TABLE 10. Power of DGP10

	J_n (asymptotic)						J_n (bootstrap)								
	$c = 1.0$			$c = 0.5$			$c = 1.0$			$c = 0.5$			LM		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.392	0.532	0.626	0.438	0.6	0.69	0.504	0.756	0.844	0.496	0.752	0.822	0.236	0.39	0.501
100	0.87	0.942	0.956	0.884	0.946	0.97	0.894	0.976	0.986	0.892	0.972	0.984	0.44	0.622	0.718

This shows that a parametric test can have smaller power even when the true DGP is further away from the null model.

The limited Monte Carlo results reported in this section suggest that our proposed tests ($\hat{J}_n$ or $\tilde{J}_n$) are conservative in the sense that they usually underestimate the sizes. Therefore, for small or moderate sample sizes, it is worthwhile to use extra computing time to use the bootstrap method to overcome the significant downsize problem of the $\hat{J}_n$ or the $\tilde{J}_n$ tests. When the sample size is large, one may use the asymptotic normal critical values, say, for $n \geq 1{,}000$, because the $\hat{J}_n$ and the $\tilde{J}_n$ tests are quite powerful for large sample sizes even though they are still undersized for $n = 1{,}000$.

Our simulation results suggest that the estimated sizes of the $\hat{J}_n$ and $\tilde{J}_n$ tests are not very sensitive to the choices of smoothing parameter a if one uses the bootstrap critical values. However, the estimated powers of these tests are still somewhat sensitive to the different choices of a. Therefore, an important question that has not been addressed by this paper is how to design some automatic data driven procedure so that a can be chosen optimally in the sense that the power of these tests is maximized (under both low and high frequency alternatives). The answer to this question is left for future research. At this stage, we recommend that applied researchers choose a via a rule of thumb (e.g., Härdle, 1990) such as $a_i = cx_{i,sd}\, n^{-1/(4+p)}$ for c close to one (say, $c = 1$), where $x_{i,sd}$ is the sample standard deviation of $\{X_{it}\}_{t=1}^{n}$ and X_{it} is the ith component of X_t.

NOTES

1. One example given by Hong (1993) is a linear regression model: $Y_t = 1 + X_t + U_t$ with $U_t = h(X_t)e_t$, $e_t \sim N(0,1)$. When $h(X_t) = X_t$, the test of Breusch and Pagan (1979) will have low power when regressing the squared residuals on $(1, X_t)$ because $E[X_t(U_t^2 - \sigma_0^2)] = 0$, where $\sigma_0^2 = E(U_t^2)$.

2. A referee has pointed out to us that one may weaken the moment condition on U_t by using some other distance measures in constructing a test or by introducing some transformation to our test statistic. However, the derivation is nontrivial. We shall pursue it as a future research project.

3. We owe many of the observations discussed in this section to Bruce Hansen.

4. If $\alpha_1 = 0$, we have $h_t = \alpha_0/(1 - \alpha_2)$ by the recursive structure of h_t, whereas if $\alpha_2 = 0$, it becomes an ARCH(1) process.

5. We owe this extension to the suggestion of a referee.

6. Note that when $h(X_t, \alpha) = \sigma_0^2$, a constant, $\eta_t = U_t^2 - \sigma_0^2 \equiv \epsilon_t$. Here, we use a different notation η_t to emphasize that $h(x, \alpha)$ is not a constant function.

7. Other tests such as the BDS test of Brock, Dechert, and Scheinkman (1987) can also be used to test ARCH effects.

REFERENCES

Ait-Sahalia, Y., P.J. Bickel, & T.M. Stoker (1994) Goodness-of-Fit Tests for Regression Using Kernel Methods. Manuscript, University of Chicago.

Bickel, P.L. (1978) Using residuals robustly I: Tests for heteroskedasticity and nonlinearity. *Annals of Statistics* 6, 266–291.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

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

Breusch, T.S. & A.R. Pagan (1979) A simple test for heteroskedasticity and random variation. *Econometrica* 47, 1287–1294.

Brock, W., D. Dechert, & J. Scheinkman (1987) A Test for Independence Based on the Correlation Dimension. Mimeo, University of Wisconsin-Madison.

Chen, X. & H. White. (1996) Laws of large numbers of Hilbert space-valued mixingale with applications. *Econometric Theory* 12, 284–304.

de Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

de Jong, R. (1996) The Bierens test under data dependence. *Journal of Econometrics* 72, 1–32.

Engle, R. (1982) Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom inflations. *Econometrica* 50, 987–1008.

Eubank, R.L. & W. Thomas (1993) Detecting heteroskedasticity in nonparametric regression. *Journal of the Royal Statistical Society, Series B* 55, 145–155.

Fan, Y. & Q. Li (1992) The Asymptotic Expansion for the Kernel Sum of Squared Residuals and Its Application in Hypothesis Testing. Manuscript, University of Windsor.

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate *U*-statistics of absolutely regular process with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

Fan, Y. & Q. Li (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.

Fan, Y. & O. Linton (1997) Some Higher Order Theory for a Consistent Nonparametric Model Specification Test. Discussion paper 1148, Cowles Foundation.

Glejser, H. (1969) A new test for heteroskedasticity. *Journal of the American Statistical Association* 64, 316–323.

Godfrey, L.G. (1978) Testing for multiplicative heteroskedasticity. *Journal of Econometrics* 8, 227–236.

Goldfeld, S.M. & R.E. Quandt (1965) Some tests for homoskedasticity. *Journal of the American Statistical Association* 60, 539–547.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* 14, 1–16.

Hansen, B.E. (1991) GARCH(1,1) processes are near epoch dependent. *Economics Letters* 36, 181–186.

Härdle, W. (1990) *Applied Nonparametric Regression.* New York: Cambridge University Press.

Harvey, A.C. (1976) Estimating regression models with multiplicative heteroskedasticity. *Econometrica* 44, 461–466.

Hildreth, C. & J.P. Houck (1968) Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association* 63, 584–595.

Hong, Y. (1993) Consistent Testing for Heteroskedasticity of Unknown Form. Manuscript, University of California, San Diego.

Horowitz, J.L. & W. Härdle (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.

Hsiao, C. & D. Mountain (1985) Estimating the short-run income elasticity of demand for electricity by using cross-sectional categorized data. *Journal of the American Statistical Association* 80, 259–265.

Koenker, R. & G. Bassett (1982) Robust tests for heteroskedasticity based on quantiles. *Econometrica* 50, 159–171.

Lavergne, P. & Q. Vuong (2000) Nonparametric significant tests. *Econometric Theory* 16, 576–601.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Lee, B.J. (1992) A heteroskedasticity test robust to conditional mean specification. *Econometrica* 60, 159–171.
Lee, S.W. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.
Li, Q. & S. Wang (1998) A simple consistent test for a parametric regression function. *Journal of Econometrics* 87, 145–165.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
Newey, W. & J. Powell (1987) Asymmetric Least Squares and Testing. *Econometrica* 55, 819–847.
Pagan, A.R. (1984) Econometric issues in the analysis of regression with generated regressors. *International Economic Review* 25, 221–247.
Pagan, A.R. & Y. Pak (1993) Testing for heteroskedasticity. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11, pp. 489–517. Amsterdam: Elsevier Science Publishers B.V.
Pagan, A.R. & A. Ullah (1988) The econometrics analysis of models with risk terms. *Journal of Applied Econometrics* 3, 87–105.
Powell, J.L. (1987) Semiparametric Estimation of Bivariate Latent Variable Models. Discussion paper, University of Wisconsin, Social System of Research Institute.
Szroeter, J. (1978) A class of parametric tests for heteroskedasticity in linear econometric models. *Econometrica* 46, 1311–1327.
Whang, Y.J. & D.W. Andrews (1993) Tests of specification for parametric and semiparametric models. *Journal of Econometrics* 57, 277–318.
White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 8, 452–475.
Yoshihara, K. (1976) Limiting behavior of *U*-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 35, 237–252.
Zheng, J.X. (1994) A Consistent Nonparametric Test of Heteroskedasticity. Manuscript, University of Texas at Austin.
Zheng, J.X. (1996) A Test of Heteroskedasticity Based on Kernel Estimation of Score Function. Manuscript, University of Texas at Austin.
Zheng, J.X. (1998) Consistent specification testing for conditional symmetry. *Econometric Theory* 14, 139–149.

APPENDIX A: PROOFS OF THEOREMS 3.2 AND 3.4

In this Appendix, we first give the definition of an absolutely regular process and list the conditions needed for Lemma 3.1. Then we will prove Theorems 3.2 and 3.4. Throughout the Appendix, C is a generic constant. The notation $A \sim B$ means that A has an order no larger than that of B.

Let $\{Z_t\}$ be a strictly stationary process and $\mathcal{M}_s^t(Z)$ denote the sigma algebra generated by $(Z_s, \ldots, Z_t)$ for $s \leq t$. The process $\{Z_t\}$ is called absolutely regular, if as $\tau \to \infty$,

$$\beta_\tau = \sup_{s \in N} E\left[\sup_{A \in \mathcal{M}_{s+\tau}^{\infty}} \{|P(A|\mathcal{M}_{-\infty}^{s}) - P(A)|\}\right] \to 0.$$

The following assumptions are used for Lemma 3.1.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

(A1)

(i) The process $\{Z_t\} \equiv \{\epsilon_t, X_t'\}$ is strictly stationary absolutely regular with mixing coefficient $\beta_\tau = O(\rho^\tau)$ for some $0 < \rho < 1$; with probability one, $E[\epsilon_t | \mathcal{M}_{-\infty}^t(X), \mathcal{M}_{-\infty}^{t-1}(\epsilon)] = 0$;

(ii) $E[|\epsilon_t^{4+\eta}|] < \infty$ and $E[|\epsilon_{t_1}^{i_1}\epsilon_{t_2}^{i_2}, \ldots, \epsilon_{t_l}^{i_l}|^{1+\xi}] < \infty$ for some arbitrarily small $\eta > 0$ and $\xi > 0$, where $2 \le l \le 4$ is an integer, $0 \le i_j \le 4$, and $\sum_{j=1}^l i_j \le 8$;

(iii) let $\sigma_\epsilon^2(x) = E[\epsilon_t^2 | X_t = x]$ and $\mu_{\epsilon 4}(x) = E[\epsilon_t^4 | X_t = x]$. Here $\sigma_\epsilon^2(x)$ and $\mu_{\epsilon 4}(x)$ satisfy some Lipschitz conditions: $|\sigma_\epsilon^2(x + u) - \sigma_\epsilon^2(x)| \le D(x)\|u\|$ and $|\mu_{\epsilon 4}(x + u) - \mu_{\epsilon 4}(x)| \le D(X)\|u\|$ with $E[|D(X)|^{2+\eta'}] < \infty$ for some small $\eta' > 0$.

(iv) Let $f_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ be the joint probability density function of $(X_1, X_{1+\tau_1}, \ldots, X_{1+\tau_l})$ $(1 \le l \le 3)$. Then $f_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ exists and satisfies a Lipschitz condition: $|f_{\tau_1,\ldots,\tau_l}(x_1 + u_1, x_2 + u_2, \ldots, x_l + u_l) - f_{\tau_1,\ldots,\tau_l}(x_1, x_2, \ldots, x_l)| \le D_{\tau_1,\ldots,\tau_l}(x_1, x_2, \ldots, x_l)\|u\|$, where $D_{\tau_1,\ldots,\tau_l}(\cdot,\ldots,\cdot)$ is integrable and satisfies the condition that $\int D_{\tau_1,\ldots,\tau_l}(x,\ldots,x)\|x\|^{2\xi} < M < \infty$, $\int D_{\tau_1,\ldots,\tau_l}(x_1,\ldots,x_l) f_{\tau_1,\ldots,\tau_l}(x_1,\ldots,x_l)\,dx < M < \infty$ for some $\xi > 1$.

(A2)

(i) $K(\cdot)$ is bounded and symmetric with $\int K(u)du = 1$ and $\int \|u\|^2 K(u)\,du < \infty$.

(ii) The smoothing parameter $a = O(n^{-\bar{\alpha}})$ for some $0 < \bar{\alpha} < (7/8)p$.

Now we state a lemma that will be used to prove $R_n = o_p((na^{p/2})^{-1})$.

LEMMA A.1. *Let $\{\eta_t, Z_t'\}$ be a stationary absolutely regular process with coefficient β_τ. Let the joint density function of $(Z_1, Z_{1+\tau_1}, \ldots, Z_{1+\tau_l})$ satisfy a condition as in* (C2)(iii) *with Z_t replacing X_t in* (C2)(iii), $(1 \le l \le 3)$, *β_m satisfy* (C4), $E(\eta_t | Z_t) = 0$ *a.e.*, $\max_{s \ne t} E[|\eta_t \eta_s|^\xi] < \infty$ *for some* $\xi > 4$, *and* $E[\|Z_t' Z_t\|^{\xi'}] < \infty$, *where* $\xi' = (1 - \xi^{-1})^{-1}$ *is between one and four. Also* $h \to 0$, $na^p \to \infty$ *as* $n \to \infty$. *Then*

$$A_n = (n^2 a^p)^{-1} \sum_t \sum_{s \ne t} \eta_t Z_s K_{ts} \equiv 2(n^2 a^p)^{-1} \sum_t \sum_{s > t} \eta_t Z_s K_{ts} = o_p((na^{p/2})^{-1/2}).$$

Proof.

$$E[\|A_n\|^2] = 4(n^2 a^p)^{-2} \sum_t \sum_{s>t} \sum_{t'} \sum_{s'>t'} E[\eta_t Z_s' K_{ts} \eta_{t'} Z_{s'} K_{t's'}]$$

$$= 4(n^2 a^p)^{-2} \sum_t \sum_{s>t} E[\eta_t^2 Z_s' Z_s K_{ts}^2]$$

$$+ 4(n^2 a^p)^{-2} \sum_t \sum_{s>t} \sum_{t'} \sum_{s'>t';(t',s') \ne (t,s)} E[\eta_t Z_s' K_{ts} \eta_{t'} Z_{s'} K_{t's'}]$$

$$\equiv A_{1n} + A_{2n}, \text{ say.}$$

It is easy to see that $A_{1n} = 4(n^2 a^p)^{-2} O(n^2 a^p) = O((n^2 a^p)^{-1}) = o((na^{p/2})^{-1})$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

For A_{2n}, we consider two cases. Let (a) denote the case of $\min\{|s-s'|,|s-t|,|s-t'|\} > m$ and (b) the case of $\min\{|s-s'|,|s-t|,|s-t'|\} \leq m$. Then by Lemma 1 of Yoshihara (1976), we have

$$A_{2n} = (n^2a^p)^{-2}\left\{\sum_{(a)} + \sum_{(b)}\right\} E[\eta_t Z_s' K_{ts} \eta_{t'} Z_{s'} K_{t's'}]$$

$$\leq (n^2a^p)^{-2}\{Cn^4\beta_m^{\delta/(1+\delta)} + mn^3 \max_{s\neq t, s'\neq t'} E[|\eta_t Z_s' K_{ts} \eta_{t'} Z_{s'} K_{t's'}|]\}$$

$$= (n^2a^p)^{-2}\{o(1) + n^3 O(a^{2p/\xi})\} = o((na^{p/2})^{-1})$$

for some $1 < \xi < \frac{4}{3}$, because we can choose $m = [C\log(n)]$ for a sufficiently large C. Hence $n^4\beta_m^{\delta/(1+\delta)} = o(1)$ by (C4), and

$$\max_{t\neq s, t'\neq s', (t,s)\neq(t',s')} E[\eta_t \eta_{t'} Z_s' Z_{s'} K_{ts} K_{t's'}]$$

$$\leq \max_{t,t'}\{E[|\eta_t \eta_{t'}|^{\xi'}]\}^{1/\xi'} \max_{t\neq s, t'\neq s', (t,s)\neq(t',s')}\{E[|Z_s' Z_{s'} K_{ts} K_{t's'}|^{\xi}]\}^{1/\xi}$$

$$= O(1)O(a^{2p/\xi}) = o(a^{3p/2})$$

by (C2)(iii), where $1 < \xi < \frac{4}{3}$ and $\xi' = (1-(\xi)^{-1})^{-1}$ is slightly greater than four. ■

Proof of Theorem 3.2. (i) $R_n = o_p((na^{p/2})^{-1})$.

$$R_n = (n^2a^q)^{-1}\sum_t\sum_{s\neq t} K_{ts}\{(\hat{U}_t^2 - U_t^2)(\hat{U}_s^2 - U_s^2) + (\hat{\sigma}_0^2 - \sigma_0^2)^2$$

$$+ 2(U_t^2 - \sigma_0^2)(\hat{U}_s^2 - U_s^2) - 2(U_t^2 - \sigma_0^2)(\hat{\sigma}_0^2 - \sigma_0^2)$$

$$- 2(\hat{U}_t^2 - U_t^2)(\hat{\sigma}_0^2 - \sigma_0^2)\}$$

$$\equiv R_{1n} + R_{2n} + 2R_{3n} - 2R_{4n} - 2R_{5n},$$

where the definitions of R_{jn} should be apparent. Subsequently we show that $R_{jn} = o_p((na^{p/2})^{-1})$ for $j = 1,\ldots,5$.

Proof of $R_{jn} = o_p((na^{p/2})^{-1})$ (j = 1,...,5). $\hat{U}_t = Y_t - g(W_t,\hat{\gamma}) = U_t - [g(W_t,\hat{\gamma}) - g(W_t,\gamma_0)] = U_t - (\hat{\gamma}-\gamma_0)'\nabla g(W_t,\gamma_0) - (\hat{\gamma}-\gamma_0)'\nabla^2 g(W_t,\tilde{\gamma})(\hat{\gamma}-\gamma_0)/2$, where $\tilde{\gamma}$ is between $\hat{\gamma}$ and γ_0. Using the shorthand notations $V_t \equiv \nabla g(W_t,\gamma_0)$ and $\tilde{V}_t \equiv \nabla^2 g(W_t,\tilde{\gamma})/2$, we have $\hat{U}_t = U_t - (\hat{\gamma}-\gamma_0)'V_t - (\hat{\gamma}-\gamma_0)'\tilde{V}_t(\hat{\gamma}-\gamma_0)$ and

$$\hat{U}_t^2 - U_t^2 = -(\hat{\gamma}-\gamma_0)'[V_t - \tilde{V}_t(\hat{\gamma}-\gamma_0)]$$

$$\times [2U_t - V_t'(\hat{\gamma}-\gamma_0) - (\hat{\gamma}-\gamma_0)'V_t - (\hat{\gamma}-\gamma_0)'\tilde{V}_t(\hat{\gamma}-\gamma_0)]$$

$$= -2(\hat{\gamma}-\gamma_0)'V_tU_t(\hat{\gamma}-\gamma_0)'A_t(\hat{\gamma}-\gamma_0), \qquad \textbf{(A.1)}$$

where $A_t = 2U_t\tilde{V}_t + V_tV_t' - \tilde{V}_t(\hat{\gamma}-\gamma_0)V_t' + V_t(\hat{\gamma}-\gamma_0)'\tilde{V}_t - \tilde{V}_t(\hat{\gamma}-\gamma_0)(\hat{\gamma}-\gamma_0)\tilde{V}_t$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Using (A.1), we have

$$
\begin{aligned}
R_{1n} &= 4(\hat{\gamma}-\gamma_0)'\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}V_tU_tU_sV_s'\right](\hat{\gamma}-\gamma_0)\\
&\quad - 4(\hat{\gamma}-\gamma_0)'\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}V_tU_t(\hat{\gamma}-\gamma_0)'A_s'\right](\hat{\gamma}-\gamma_0)\\
&\quad + (\hat{\gamma}-\gamma_0)'\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}A_t(\hat{\gamma}-\gamma_0)(\hat{\gamma}-\gamma_0)'A_s\right](\hat{\gamma}-\gamma_0)\\
&= O_p(n^{-1})O_p(1) = o_p((na^{p/2})^{-1})
\end{aligned}
$$

by the fact that $\hat{\gamma}-\gamma_0 = O_p(n^{-1/2})$.

Next for R_{2n}, noting that $\hat{U}_t^2 = U_t^2 + [(\hat{\gamma}-\gamma_0)'V_t]^2 + [(\hat{\gamma}-\gamma_0)'\widetilde{V}_t(\hat{\gamma}-\gamma_0)]^2 - 2U_tV_t'(\hat{\gamma}-\gamma_0) - 2U_t(\hat{\gamma}-\gamma_0)'\widetilde{V}_t(\hat{\gamma}-\gamma_0) + 2(\hat{\gamma}-\gamma_0)'V_t(\hat{\gamma}-\gamma_0)'\widetilde{V}_t(\hat{\gamma}-\gamma_0)$ and using $(\hat{\gamma}-\gamma_0) = O_p(n^{-1/2})$, one can easily show that

$$
\begin{aligned}
(\hat{\sigma}_0^2-\sigma_0^2) &= n^{-1}\sum_t(U_t^2-\sigma_0^2) - \left[2n^{-1}\sum_t U_tV_t'\right](\hat{\gamma}-\gamma_0) + O_p(n^{-1})\\
&= O_p(n^{-1/2}) + O_p(n^{-1/2})O_p(n^{-1/2}) + O_p(n^{-1}) = O_p(n^{-1/2}),
\end{aligned}
$$

where the second equality follows from Lemma 1 of Yoshihara (1976), and conditions (C1) and (C4). Thus, $(\hat{\sigma}_0^2-\sigma_0^2)^2 = O_p(n^{-1})$. Hence, $R_{2n} = O_p(n^{-1})$ because $(n^2a^{-p})^{-1}\sum_t\sum_{s\neq t}K_{ts} = O_p(1)$.

Again using equation (A.1), we have

$$
\begin{aligned}
R_{3n} &= -2(\hat{\gamma}-\gamma_0)'\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2-\sigma_0^2)V_sU_sK_{ts}\right]\\
&\quad + (\hat{\gamma}-\gamma_0)'\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2-\sigma_0^2)A_sK_{ts}\right](\hat{\gamma}-\gamma_0)\\
&= -2O_p(n^{-1/2})o_p((na^p)^{-1/2}) + O_p(n^{-1/2})O_p(1)O_p(n^{-1/2}) = o_p((na^{p/2})^{-1}).
\end{aligned}
$$

The second equality of the preceding equation follows from Lemma A.1 (with $\eta_t = U_t^2-\sigma_0^2$ and $Z_s = U_sV_s$) and $\hat{\gamma}-\gamma = O_p(n^{-1/2})$.

By Lemma A.1 (with $\eta_t = U_t^2-\sigma_0^2$ and $Z_t = 1$) and the fact that $\hat{\sigma}_0^2-\sigma_0^2 = O_p(n^{-1/2})$, it follows that

$$
\begin{aligned}
R_{4n} &= (\hat{\sigma}_0^2-\sigma_0^2)\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2-\sigma_0^2)K_{ts}\right] = (\hat{\sigma}_0^2-\sigma_0^2)o_p((na^p)^{-1/2})\\
&= O_p(n^{-1/2})o_p((na^{p/2})^{-1}) = o_p((na^{p/2})^{-1}).
\end{aligned}
$$

Similarly using equation (A.1), one can easily show that

$$
\begin{aligned}
R_{5n} &= (\hat{\sigma}_0^2-\sigma_0^2)(n^2a^p)^{-1}\sum_t\sum_{s\neq t}(\hat{U}_t^2-U_t^2)K_{ts}\\
&= (\hat{\sigma}_0^2-\sigma_0^2)(\hat{\gamma}-\gamma_0)'\\
&\quad\times\left\{-2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}U_tV_tK_{ts} + (n^2a^p)^{-1}\sum_t\sum_{s\neq t}A_tK_{ts}(\hat{\gamma}-\gamma_0)\right\}\\
&= O_p(n^{-1})\{o_p((na^p)^{-1/2}) + O_p(n^{-1/2})\} = o_p(n^{-1}) = o_p((na^{1/2})^{-1}),
\end{aligned}
$$

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

where the third equality follows from Lemma A.1, $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$, and the fact that $(n^2a^p)^{-1}\sum_t\sum_{s\neq t}A_tK_{ts} = O_p(1)$ (by Yoshihara, 1976, Lemma 1). ■

Proof of Theorem 3.2 (ii) $\hat{\Omega} = \Omega + o_p(1)$. Using the same arguments as in the proof of Theorem 3.2 (i), one can easily show that $\hat{\Omega} = 2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}\epsilon_t^2\epsilon_s^2K_{ts}^2 + o_p(1) \equiv \Omega_n + o_p(1)$. By Lemma 1 of Yoshihara (1976), we have

$$
\begin{aligned}
&|E(\Omega_n) - 2(n^2a^p)^{-1}n(n-1)\textstyle\int\int\epsilon_1^2\epsilon_2^2K^2((x_1-x_2)/a)\,dF(z_1)\,dF(z_2)| \\
&\quad \leq C(n^2a^p)^{-1}\sum_t\sum_{s\neq t}\beta_{|s-t|}^{\delta/(1+\delta)} \\
&\quad = O((na^p)^{-1}) = o_p(1),
\end{aligned}
$$

where $z_1 = (\epsilon_1, x_1')$ and $F(\cdot)$ is the distribution function of $Z = (\epsilon, X')$. Here $2a^{-p}\int\int\epsilon_1^2\epsilon_2^2K^2((x_1-x_2)/a)\,dF(z_1)\,dF(z_2) = \Omega + o(1)$. Thus, $E(\hat{\Omega}) = \Omega + o(1)$.

Finally using Lemma 1 of Yoshihara (1976) and (C4), it is straightforward to show that $\mathrm{var}(\hat{\Omega}) = o_p(1)$. Hence, $\hat{\Omega} = \Omega + o_p(1)$. ■

Proof of Theorem 3.4. The proof is similar to that of Theorem 3.2. We will only provide a sketch to save space. Let $\alpha_n = (n^{1/2}a^{p/4})^{-1}$ and define $U_{0,t}^2 = U_t^2 - \alpha_n g(X_t)$. Then we have $E(U_{0,t}^2|X_t) = \sigma_0^2$. Using the same proof as we did in the proof of Theorem 3.2, we have under H_L, $\hat{J}_n = J_n + R_n$, where $J_n = (n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^2 - \sigma_0^2)(U_s^2 - \sigma_0^2)K_{ts}$ and $R_n = \hat{J}_n - J_n$. Here $R_n = o_p((na^{p/2})^{-1})$ follows the same arguments as in the proof of Theorem 3.2.

$$
\begin{aligned}
J_n &= \frac{1}{n^2a^p}\sum_t\sum_{s\neq t}(U_{0,t}^2 - \sigma_0^2 + \alpha_n g(X_t))(U_{0,s}^2 - \sigma_0^2 + \alpha_n g(X_t))K_{ts} \\
&= \frac{1}{n^2a^p}\sum_t\sum_{s\neq t}\{(U_{0,t}^2 - \sigma_0^2)(U_{0,s}^2 - \sigma_0^2) - 2\alpha_n(U_{0,t}^2 - \sigma_0^2)g(X_s) + \alpha_n^2 g(X_t)g(X_s)\}K_{ts} \\
&\equiv B_{1n} - 2B_{2n} + B_{3n}.
\end{aligned}
$$

Here $na^{p/2}B_{1n} \to N(0,\Omega)$ in distribution by Lemma 3.1, and $na^{p/2}B_{2n} = na^{p/2}\alpha_n o_p \times ((na^{p/2})^{-1/2}) = o_p(1)$ by Lemma A.1. The expression $na^{p/2}E[B_{3n}] = (n^2a^p)^{-1} \times \sum_t\sum_{s\neq t}E[g(X_t)g(X_s)K_{ts}] = E[g^2(X_t)f(X_t)] + o(1)$, and therefore, $na^{p/2}B_{3n} = E[g^2(X_t)f(X_t)] + o_p(1) = \mu + o_p(1)$ by virtue of the law of large numbers (e.g., Chen and White, 1996).

Summarizing the preceding information we have shown that $na^{p/2}\hat{J}_n \to N(\mu,\Omega)$ in distribution under H_L. ■

APPENDIX B: PROOF OF THEOREM 4.1

The proof of Theorem 4.1 follows steps similar to those in the proof of Theorem 3.2. Therefore, we only sketch the proof here. Recall that $\mathcal{F}_n = \{(Y_t, W_t, X_t)\}_{t=1}^n$. We will use $E^*(\cdot)$ to denote $E(\cdot|\mathcal{F}_n)$

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Denote $\hat{\sigma}_0^{*2} = n^{-1}\sum_t \hat{U}_t^{*2}$, $K_{ts}^* = K((X_t^* - X_s^*)/a)$ and define $\sigma_0^{*2} = E^*(U_t^{*2}) \equiv n^{-1}\sum_t(\hat{U}_t - \bar{U})^2 = n^{-1}\sum_t \hat{U}_t^2 + O_p(n^{-1/2}) = \hat{\sigma}_0^2 + O_p(n^{-1/2})$. Adding and subtracting terms, we have

$$\begin{aligned}
\hat{J}_n^* &= (n^2a^p)^{-1}\sum_t\sum_{s\neq t}(\hat{U}_t^{*2} - \hat{\sigma}_0^{*2})(\hat{U}_s^{*2} - \hat{\sigma}_0^{*2})K_{t,s}^* \\
&= (n^2a^p)^{-1}\sum_t\sum_{s\neq t}\{[(U_t^{*2} - \sigma_0^{*2}) + (\hat{U}_t^{*2} - U_t^{*2}) - (\hat{\sigma}_0^{*2} - \sigma_0^{*2})] \\
&\qquad \times [(U_s^{*2} - \sigma_0^{*2}) + (\hat{U}_s^{*2} - U_s^{*2}) - (\hat{\sigma}_0^{*2} - \sigma_0^{*2})]\}K_{ts}^* \\
&= (n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^{*2} - \sigma_0^{*2})(U_s^{*2} - \sigma_0^{*2})K_{ts}^* + R_n^* \equiv J_n^* + R_n^*,
\end{aligned}$$

where $J_n^* = (n^2a^p)^{-1}\sum_t\sum_{s\neq t}(U_t^{*2} - \sigma_0^{*2})(U_s^{*2} - \sigma_0^{*2})K_{ts}^*$ and $R_n^* = J_n^* - \hat{J}_n^*$.

It is easy to see that $J_n^*|\mathcal{F}_n$ is a second order degenerate U-statistic (which implies $E^*(J_n^*) = 0$) because U_t^* and U_s^* are independent with each other conditional on $\mathcal{F}_n$ (for $s \neq t$).

Next we evaluate the second moment of J_n^*. As in Fan and Li (1999), we use the tilde notation to denote independent processes. For example, $\{\tilde{\eta}_t^*, \tilde{X}_t^*\}_{t=1}^n$ is an independent process and $(\tilde{\eta}_t^*, \tilde{X}_t^*)$ have the same marginal distribution as (η_t^*, X_t^*), where $\eta_t^* = U_t^{*2} - \sigma_0^{*2}$. Also, we denote $\tilde{K}_{ts}^* = K((\tilde{X}_t^* - \tilde{X}_s^*)/a)$. Using exactly the same arguments as in the proof of Theorem 2.1 of Fan and Li (1999), one can show that

$$\begin{aligned}
E^*[(na^{p/2}J_n^*)^2] &= 2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}E^*[\eta_t^{*2}\eta_s^{*2}K_{ts}^{*2}] + o_p(1) \\
&= 2(n^2a^p)^{-1}\sum_t\sum_{s\neq t}E^*[\tilde{\eta}_t^{*2}\tilde{\eta}_s^{*2}\tilde{K}_{ts}^{*2}] + o_p(1) \\
&= 2[E^*(\tilde{\eta}_t^{*2})E^*(\tilde{\eta}_s^{*2})]\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}E^*(\tilde{K}_{ts}^2)\right] + o_p(1) \\
&= 2\left[n^{-1}\sum_t(\hat{U}_t^4 - \hat{\sigma}_0^4) + o_p(1)\right]^2\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}^2 + o_p(1)\right] \\
&= 2\left[n^{-1}\sum_t(\hat{U}_t^4 - \hat{\sigma}_0^4)\right]^2\left[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}^2\right] + o_p(1) \\
&\equiv \hat{\Omega}_0 + o_p(1), \qquad \textbf{(B.1)}
\end{aligned}$$

where $\hat{\Omega}_0 = 2[n^{-1}\sum_t(\hat{U}_t^4 - \hat{\sigma}_0^4)]^2[(n^2a^p)^{-1}\sum_t\sum_{s\neq t}K_{ts}^2]$. The third equality in the preceding equation follows because $(\tilde{\eta}_t^*, \tilde{X}_t^*)$ is independent of $(\tilde{\eta}_s^*, \tilde{X}_s^*)$ for $t \neq s$ and $E^*(\tilde{\epsilon}_t^{*2}|\tilde{X}_t^*) = E^*(\tilde{\epsilon}_t^{*2})$, and the fourth equality follows because (noting that $\eta_t^* = U_t^{*2} - \sigma_0^{*2}$)

$$E^*(\eta_t^{2*}) = n^{-1}\sum_t[\hat{U}_t^4 - \hat{\sigma}_0^4] + O_p(n^{-1/2}), \qquad \textbf{(B.2)}$$

where the $O_p(n^{-1/2})$ term comes from $\bar{U} = n^{-1}\sum_t \hat{U}_t = O_p(n^{-1/2})$.

Similarly, one can show that $\hat{\Omega}^* = \hat{\Omega}_0 + o_p(1)$. Thus, $na^{p/2}J_n^*|\mathcal{F}_n$ has mean zero and variance $\hat{\Omega}_0 + o_p(1)$. Therefore, $(na^{p/2}J_n^*/\sqrt{\hat{\Omega}^*})|\mathcal{F}_n = (na^{p/2}J_n^*/\sqrt{\hat{\Omega}_0})|\mathcal{F}_n + o_p(1) \to N(0,1)$ in distribution by Lemma 3.1.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

Using $\hat{\gamma}^* - \hat{\gamma} = O_p(n^{-1/2})$ and following the same arguments as in the proof of $R_n = o_p((na^{p/2})^{-1})$ (see Appendix A), one can show that $R_n^* | \mathcal{F}_n = o_p((na^{p/2})^{-1})$. Thus, $na^{p/2}\hat{J}_n^* | \mathcal{F}_n = na^{p/2} J_n^* | \mathcal{F}_n + o_p(1) \to N(0,1)$ in distribution. This completes the proof of Theorem 4.1. ■

APPENDIX C: PROOF OF THEOREM 5.1

Recall that $X_t = (U_{t-1}, \ldots, U_{t-p})'$, $\hat{X}_t = (\hat{U}_{t-1}, \ldots, \hat{U}_{t-p})'$ and denote $Z_t = (W_{t-1}, W_{t-2}, \ldots, W_{t-p})$, which is of dimension $q \times p$. Hence, $\hat{X}_t - X_t = Z_t'(\gamma_0 - \hat{\gamma})$.

Here we give a Taylor expansion of $\hat{K}_{ts}$ that is used several times in Appendix B. By the Lipschitz condition in (C5) of Section 5 with $u = (\hat{X}_t - \hat{X}_s)/a$ and $v = (X_t - X_s)/a$, we have

$$\hat{K}_{ts} = K_{ts} + (\nabla K_{ts})'(Z_t - Z_s)'(\hat{\gamma} - \gamma_0) + \frac{1}{2!}(\hat{\gamma} - \gamma_0)'(Z_t - Z_s)\nabla^2 K_{ts}(Z_t - Z_s)'(\hat{\gamma} - \gamma_0) + \hat{D}_{3n} + \hat{D}_{4n}, \tag{C.1}$$

where $\hat{D}_{3n}$ is the third term of Taylor expansion which contains terms such as $K_i^{(3)}((X_t - X_s)/a)[(W_{t-i} - W_{s-i})'(\gamma_0 - \hat{\gamma})/a]^3/(3!)$ $(i = 1, \ldots, p)$ and third order cross derivatives, $K_i^{(j)}(v) = \{\partial^j K(u)/\partial u_i^j\}|_{u=v}$ and u_i is the ith component of u ($j \geq 1$ is an integer). Similarly, $\tilde{D}_{4n}$ is the remainder term of the Taylor expansion. It contains terms such as $K_i^{(4)}((\eta_t - \eta_s)/a)[(W_{t-i} - W_{s-i})'(\gamma_0 - \hat{\gamma})/a]^4/(4!)$, where η_t is between the line segment of $\hat{X}_t$ and X_t.

Substituting (C.1) into (6), we get

$$\begin{aligned}\tilde{J}_n &= \frac{1}{n^2 a^p}\sum_t \sum_{s \neq t} \hat{\epsilon}_t \hat{\epsilon}_s \\ &\quad \times \left\{K_{ts} + (\nabla K_{ts})' Z_t'(\hat{\gamma} - \gamma_0) + \frac{1}{2!}(\hat{\gamma} - \gamma_0)' Z_t \nabla^2 K_{ts} Z_t'(\hat{\gamma} - \gamma_0) + \hat{D}_{3n} + \hat{D}_{4n}\right\} \\ &\equiv \hat{J}_n + \tilde{J}_{1n}(\hat{\gamma} - \gamma_0) + (\hat{\gamma} - \gamma_0)'\tilde{J}_{2n}(\hat{\gamma} - \gamma_0) + \tilde{J}_{3n} + \tilde{J}_{4n},\end{aligned} \tag{C.2}$$

where $\hat{\epsilon}_t = \hat{U}_t^2 - \hat{\sigma}_0^2$ and the definition of $\tilde{J}_{in}$ $(i = 1, 2, 3, 4)$ should be obvious.

By the result of Theorem 3.2 and the fact that $(\hat{\gamma} - \gamma_0) = O_p(n^{-1/2})$, we only need to show that (i) $n^{-1/2}\tilde{J}_{1n} = o_p((na^{p/2})^{-1})$, $n^{-1}\tilde{J}_{2n} = o_p((na^{p/2})^{-1})$ $\tilde{J}_{3n} = o_p((na^{p/2})^{-1})$ and (ii) $\tilde{J}_{4n} = o_p((na^{p/2})^{-1})$.

Proof of (i): $n^{-1/2}\tilde{J}_{1n}$, $n^{-1}\tilde{J}_{2n}$, and $\tilde{J}_{3n}$ Are All $o_p((na^{p/2})^{-1})$.

(a) We first consider $n^{-1/2}\tilde{J}_{1n}$. Note that $\tilde{J}_{1n} = (1/n^2 a^p)\sum_t \sum_{s \neq t} \hat{\epsilon}_t \hat{\epsilon}_s (\nabla K_{ts})' Z_t'$ and compare it with the expression of $\hat{J}_n$ given by (3). We see that $\tilde{J}_{1n}$ can be obtained by replacing K_{ts} in $\tilde{J}_n$ by $(\nabla K_{ts})' Z_t'$. Following the same proof as in the

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press

proof of Theorem 3.2, we know that $\hat{J}_n = O_p((na^{p/2})^{-1})$. Hence, $n^{-1/2}\hat{J}_{1n} = n^{-1/2}O_p((na^{p/2})^{-1}) = o_p((na^{p/2})^{-1})$.

(b) The proof for $n^{-1}\tilde{J}_{2n}$ follows the same arguments as in (a) immediately preceding. Note that $\tilde{J}_{2n} = 1/n^2a^p \sum_t \sum_{s\neq t} \hat{\epsilon}_t\hat{\epsilon}_s Z_t\nabla^2 K_{ts} Z_t'$. Thus, $\tilde{J}_{2n}$ can be obtained from $\hat{J}_n$ by replacing K_{ts} in $\hat{J}_n$ by $Z_t\nabla^2 K_{ts} Z_t'$. By the same proof as that of Theorem 3.2 with respect to $\hat{J}_n = O_p((na^{p/2})^{-1})$, one can easily show that $n^{-1}\tilde{J}_{2n} = n^{-1}O_p((na^{p/2})^{-1}) = o_p((na^{p/2})^{-1})$.

(c) Similarly, one can show that a typical element in $\tilde{J}_{3n}$ has a term of the order $O_p(n^{-3/2})$ because $(\hat{\gamma} - \gamma_0) = O_p(n^{-1/2})$. Moreover this term can be factored out of all the summations. Hence, one gets $\tilde{J}_{3n} \sim n^{-3/2}L_{3n}$, where L_{3n} is obtained from $\tilde{J}_{3n}$ with $(\gamma_0 - \hat{\gamma})$ replaced by $\iota \equiv (1,\ldots,1)'$, a $q \times 1$ vector of ones. The definition of L_{3n} is similar to $\tilde{J}_{2n}$ but with the second order derivatives of $K_{ts}^{(2)}$ replaced by the third order derivatives $K_{st}^{(3)}$, etc. Hence, by the same arguments as given previously one can show that $n^{-3/2}\tilde{J}_{3n} = n^{-3/2}O_p((na^{p/2})^{-1}) = o_p((na^{p/2})^{-1})$.

Proof of (ii): $\tilde{J}_{4n} = o_p((na^{p/2})^{-1})$. By the Lipschitz condition in (C5), we know that $|\tilde{J}_{4n}| \leq C(n^2a^p)^{-1}\sum_t\sum_{s\neq t}|\hat{\epsilon}_t\hat{\epsilon}_s H((X_t - X_S)/a)|\cdot\|(Z_t - Z_s)'(\gamma_0 - \hat{\gamma})\|^4$. Given that $\gamma_0 - \hat{\gamma} = O_p(n^{-1/2})$, it is sufficient to show that $n^{-2}G_n \equiv n^{-2}(n^2a^p)^{-1} \times \sum_t\sum_{s\neq t}|\hat{\epsilon}_t\hat{\epsilon}_s H((X_t - X_S)/a)\|(Z_t - Z_s)'\iota\|^4 = o_p((na^{p/2})^{-1})$ (ι is a $q \times 1$ vector of ones). Obviously, the leading term of G_n is $(n^2a^p)^{-1}\sum_t\sum_{s\neq t}|\epsilon_t\epsilon_s H((X_t - X_S)/a)|\cdot\|(Z_t - Z_s)'\iota/a\|^4 \sim C(n^2a^{p+4})^{-1}\sum_t\sum_{s\neq t}U_t^2U_s^2H_{ts}\|(Z_t - Z_s)'\iota\|^4 \equiv G_{n1}$. Here $E|G_{n1}| = C(n^2a^{p+4})^{-1}O(n^2a^p) = O(a^{-4})$. Therefore, $G_n \sim G_{n1} = O_p(a^{-4})$ and $n^{-2}G_n = O_p(n^{-2}a^{-4}) = o_p((na^{p/2})^{-1})$ by the assumption that $na^{4-(p/2)} \to \infty$.

https://doi.org/10.1017/S0266466601171069 Published online by Cambridge University Press