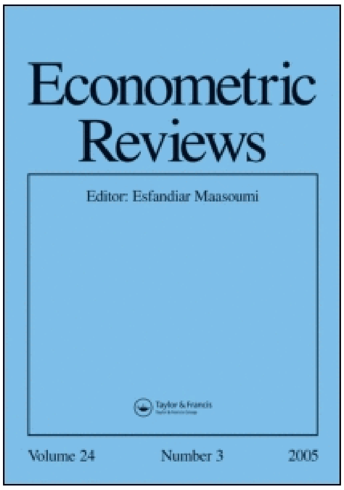

Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597248

Adaptive testing in arch models

Oliver B. Linton[a]; Douglas G. Steigerwald[a]
[a] Department of Economics, Yale University, New Haven, CT, USA

To cite this Article Linton, Oliver B. and Steigerwald, Douglas G.(2000) 'Adaptive testing in arch models', Econometric Reviews, 19: 2, 145 — 174
To link to this Article: DOI: 10.1080/07474930008800466
URL: http://dx.doi.org/10.1080/07474930008800466

ECONOMETRIC REVIEWS, 19(2), 145–174 (2000)

ADAPTIVE TESTING IN ARCH MODELS

OLIVER B. LINTON
Department of Economics
Yale University
New Haven, CT 06520
USA

DOUGLAS G. STEIGERWALD
Department of Economics
University of California
Santa Barbara, CA 93106
USA

Keywords and Phrases: adaptive testing; ARCH; conditional heteroskedasticity; semiparametric.

JEL Classification Numbers: C12, C14, C22.

ABSTRACT

Specification tests for conditional heteroskedasticity that are derived under the assumption that the density of the innovation is Gaussian may not be powerful in light of the recent empirical results that the density is not Gaussian. We obtain specification tests for conditional heteroskedasticity under the assumption that the innovation density is a member of a general family of densities. Our test statistics maximize asymptotic local power and weighted average power criteria for the general family of densities. We establish both first-order and second-order theory for our procedures. Simulations indicate that asymptotic power gains are achievable in finite samples.

1. INTRODUCTION

Volatility clustering is an important characteristic of financial time series. To account for volatility clustering, which is a term for serial correlation in the second moment of a series, researchers often estimate variants of the autoregressive conditional heteroskedasticity (ARCH) model developed by Engle (1982). Successful application of these models requires correct specification of both the conditional mean and the conditional variance. Our interest here is in developing powerful statistics for testing the specification of the conditional variance.

Much is known about the behavior of test statistics for the conditional variance that are constructed from a Gaussian likelihood.[1] Recent empirical work questions the assumption that the likelihood is Gaussian.[2] If the likelihood is not Gaussian, then test statistics based on a Gaussian likelihood are not asymptotically the most powerful test statistics. To improve power, one could use non-Gaussian test statistics. If the innovation density is correctly specified, then non-Gaussian test statistics are asymptotically most powerful. If, however, the innovation density is not correctly specified, then non-Gaussian test statistics are not asymptotically most powerful and may not be consistent.[3]

We develop test statistics for the conditional variance that do not suffer from a loss of power if the likelihood is not Gaussian. Our test statistics are semiparametric, that is we specify the first two conditional moments parametrically but assume only that the innovation density is a member of a nonparametric family. We show that the semiparametric test statistics are *adaptive* in the sense that they are asymptotically equivalent to test statistics constructed from the true likelihood. As a result, the semiparametric test statistics inherit the asymptotic properties of the correctly specified likelihood-based test statistics.

Our result may seem surprising when contrasted with previous results for estimation of conditional heteroskedasticity models. Linton (1993) and Steigerwald (1993) show that the conditional variance parameters cannot be estimated adaptively because of a lack of identification of the scale of the innovation density. Because the test statistics we construct do not depend on the scale of the innovation density, but only on the relative scale parameters that can be estimated adaptively, the test statistics are not affected by the problem of estimating the scale parameter.

Because the conditional variance must be positive, many of the specification tests we consider naturally have one-sided alternative hypotheses. For such tests we follow a proposal in Lee and King (1993) to construct test statistics that are more powerful than statistics designed for two-sided alternative hypotheses. Specifically, we show that for a test of one additional parameter in the conditional variance function, the positive square-root of a semiparametric Lagrange multiplier (LM) test statistic is consistent and maximizes asymptotic power uniformly

[1]Engle (1983) shows that if the true innovation density is Gaussian, then the Lagrange multiplier test statistic is asymptotically distributed as a Chi-square random variable. Bollerslev and Wooldridge (1992) show that the asymptotic result continues to hold even if the true innovation density is not Gaussian.

[2]Evidence that standardized errors from a CH model of asset prices do not have a Gaussian density is provided by a number of authors. For example, Baillie and Bollerslev (1989) use both an exponential power and a t density to model exchange rates, Hsieh (1989) uses several mixture densities to model exchange rates, and Nelson (1991) uses an exponential-power density to model stock prices.

[3]Results in Newey and Steigerwald (1997) imply that non-Gaussian likelihood based test statistics are not generally robust to misspecification of the innovation density.

against local alternatives.[4] For a test of more than one additional parameter in the conditional variance function, we show that a semiparametric sum of scores test statistic is consistent and maximizes asymptotic power against appropriately defined local alternatives.

Our semiparametric test statistics are constructed from a nonparametric estimator of the innovation density. The use of a nonparametric density estimator may result in the small sample properties differing markedly from the predicted first-order asymptotic theory. To determine the small sample properties of our semiparametric test statistic, we also derive second-order asymptotic theory. The second-order results allow us to determine a value of the smoothing parameter that improves the performance of semiparametric test statistics. Because it is difficult to obtain second-order asymptotic results under the weak regularity conditions we use for our first-order asymptotic results, we strengthen the regularity conditions for the second-order results. Thus the semiparametric test statistics that we study are asymptotically optimal (to first order) for a very broad class of densities and are asymptotically optimal (to second order) for a smaller class of densities.

We also examine the small sample properties of our test statistics through simulation. We find that for a test of more than one additional parameter in the conditional variance equation, the semiparametric test statistic has power gains for samples of only 100 observations.

2. Model and Hypotheses

Let $z_t = (y_t, x_t')'$, $t = 1, \ldots, T$, be the observed data, where the dependent variable y_t is a scalar, while x_t is a k by 1 vector of regressors. We consider ARCH models of the form

$$y_t = \beta' x_t + h_t(\gamma)\sigma u_t,$$

where $h_t(\gamma)$ is a function of the set of past information $\mathcal{F}_{t-1} = \{x_t, z_{t-1}, z_{t-2}, \ldots\}$ and a vector γ of parameters of interest, u_t is a period-t i.i.d. innovation with scale parameter σ and is independent of $\mathcal{F}_{t-1}$. Our parameterization is slightly different from Engle's original parameterization. In Engle's parameterization, u_t is assumed to be a Gaussian random variable with variance 1, so that $\sigma = 1$ and all the parameters of the conditional variance are identified. Because we allow the density of u_t to be a member of a general family of densities, we are only able to identify the parameters of the conditional variance up to scale. For example, in the ARCH(p) model

$$h_t^2(\gamma) = 1 + \sum_{i=1}^{p} \phi_i (y_{t-i} - \beta' x_{t-i})^2,$$

[4]Bera and Ng (1991) also construct test statistics based on nonparametric estimates of the score function, although their test statistics are based on a two-sided alternative and so do not maximize power.

where the parameters in ϕ are the relative scale parameters of the conditional variance of y_t, that is ϕ consists of ratios of each of the slope parameters to the constant parameter in the conditional variance. If the conditional variance is $\delta_0 + \sum_{i=1}^{p} \delta_i (y_{t-i} - \beta^T x_{t-i})^2$, then $\phi_i = \delta_i / \delta_0$.

We consider a test for additional parameters in the conditional variance of an ARCH model; that is, a test of the null hypothesis that the model is ARCH$(p-m)$, with $m \geq 1$:

$$H_0 : \phi_{p-m+1} = \ldots = \phi_p = 0,$$

against the one-sided alternative hypothesis that the model is ARCH(p):

$$H_A : \phi_{p-m+i} \geq 0,\ i = 1, \ldots, m \text{ with at least one strict inequality.}$$

While the alternative hypothesis we study is an ARCH model, the test has power against generalized ARCH (GARCH) models.[5] As pointed out in Lee and King (1993) the LM statistic for testing a null of homoskedasticity against ARCH(p) or against GARCH(p,q) is the same because the score for the subset of the q additional conditional variance parameters equals zero under the null hypothesis. To carry out tests for a model in which the conditional variance is GARCH(p,q), we construct the asymptotically optimal test statistics for an ARCH(p) alternative with the residuals from the estimated GARCH(p,q) model.

3. Likelihood and Test Statistics

If the Lebesgue density $g(\cdot)$ of u_t is known, then optimal inference about the parameters $\theta = (\beta', \sigma, \gamma')'$ is based on the sample log-likelihood

$$L_T(\theta, g) = \sum_{t=1}^{T} l_t(\theta, g) = -T \ln \sigma - \sum_{t=1}^{T} \ln h_t(\gamma) + \sum_{t=1}^{T} \ln g(u_t),$$

where $u_t = \sigma^{-1} h_t^{-1}(\gamma)[y_t - \beta' x_t]$, and $l_t(\theta, g)$ is the period-t conditional log-likelihood of y_t given $\mathcal{F}_{t-1}$. (The period-0 observation is considered fixed.) Define the Fisher scores for location and scale of the innovation density as $\psi_1(u) = -g^{(1)}(u)/g(u)$ and $\psi_2(u) = -[1 - u\psi_1(u)]$, where $g^{(i)}(u)$ is the i^{th} derivative of g with respect to u. Let

$$\psi(u_t(\theta)) = \begin{pmatrix} \psi_1(u_t(\theta)) \\ \psi_2(u_t(\theta)) \end{pmatrix},$$

where $\psi(u_t(\theta_0))$ is mean zero and independent of $\mathcal{F}_{t-1}$.

Of central interest are the parameters ϕ, while $\chi = (\sigma, \beta')'$ are nuisance parameters. The efficient score $s^*_\phi(\theta, g)$ and efficient information $\mathcal{J}^*_{\phi\phi}$ (see Bickel *et al.* (1994) for a discussion of the efficient score) are

[5] Nelson and Cao (1991) show that the alternative space for a GARCH model is not one-sided and instead has a very complicated structure, rendering it difficult to obtain asymptotically optimal tests.

$$s^*_\phi(\theta, g) = T^{-1/2} \sum_{t=1}^{T} \frac{\partial l^*_t}{\partial \phi} \quad ; \quad \mathcal{J}^*_{\phi\phi} = \mathcal{J}_{\phi\phi} - \mathcal{J}_{\phi\chi}\mathcal{J}^{-1}_{\chi\chi}\mathcal{J}_{\chi\phi},$$

where $\partial l^*_t/\partial\phi = \partial l_t/\partial\phi - \mathcal{J}_{\phi\chi}\mathcal{J}^{-1}_{\chi\chi}\partial l_t/\partial\chi$. The efficient scores for the relative-scale coefficients ϕ are orthogonal to the tangent space of scores for g in the semiparametric model, thus the situation is in principle adaptive (for estimation and hence for testing) for ϕ, see Linton (1993) and Steigerwald (1993). The period-t component of the efficient score is

$$\frac{\partial l^*_t}{\partial \phi}(\theta, g) = \Gamma^*_{\phi t}(\theta, g)\psi(u_t(\theta)),$$

where $\Gamma^*_{\phi t}(\theta, g)$, which is calculated from expressions contained in the appendix, depends only on $\mathcal{F}_{t-1}$. If g is symmetric about zero, then $\mathcal{J}_{\phi\beta} = 0$ and the period-t component of the efficient score is

$$\frac{\partial l^*_t}{\partial \phi} = \frac{1}{2}\left\{v_{t-1} - E(v_{t-1})\right\}\psi_2(u_t),$$

where $v_{t-1} = (v_{1,t-1}, \ldots, v_{p,t-1})'$ with $v_{i,t-1} = \frac{(y_{t-i}-\beta' x_{t-i})^2}{1+\sum_{i=1}^{p}\phi_i(y_{t-i}-\beta' x_{t-i})^2}$.[6]

Let $\widetilde{\theta}$ be the maximum likelihood estimator (MLE) of θ imposing the null restrictions, or any asymptotically equivalent estimator, and let for any p by 1 vector c:

$$\tau_c = \frac{c' s^*_\phi(\widetilde{\theta}, g)}{\left\{c'\widehat{\mathcal{J}}^*_{\phi\phi}(\widetilde{\theta}, g)c\right\}^{1/2}}, \tag{0.1}$$

where $\widehat{\mathcal{J}}_{\theta\theta}(\theta, g)$ is a consistent estimator of $\mathcal{J}_{\theta\theta}(\theta, g)$ and $\widehat{\mathcal{J}}^*_{\phi\phi}$ is the corresponding element of the efficient information estimator. A parametric sum of scores test statistic for H_0 versus H_A is τ_c where the first $p-m$ elements of c equal 0 and the last m elements of c equal 1. If g is Gaussian these test statistics are particularly simple and have been extensively studied, see especially Lee and King (1993), Bera and Higgins (1993), Engle (1983), and Bollerslev and Wooldridge (1992). As we show in Section 4, τ_c is asymptotically standard Gaussian.

The semiparametric versions of our test statistics are also based on estimating (0.1).[7] We replace population moments by their sample equivalents and $g(u)$ by a nonparametric kernel density estimator

$$\widehat{g}(u) = T^{-1}h^{-1}\sum_{s\in\mathcal{T}(t)} K\left(\frac{u - \widetilde{u}_s}{h}\right), \tag{0.2}$$

[6] If u_t is a Gaussian random variable, then $I_{12} = 0$, while $\psi_1(u) = -u$ and $\psi_2(u) = -[u^2 - 1]$.

[7] We do not estimate location and scale because these parameters are not jointly identified with the innovation density g.

where $K(\cdot)$ is a kernel function and $h(T)$ is a bandwidth parameter both satisfying conditions A9 given in the appendix.[8] The index set $\mathcal{T}_t$ is taken here to be $\{s : s \neq t\}$. Here, $\widetilde{u}_s$ are standardized residuals from a preliminary $T^{1/2}$ consistent procedure, for example the Gaussian MLE $\widetilde{\theta}_G$. Because the kernel is unbounded, we introduce the trimming rule

$$\widehat{1}_t = \begin{cases} 1 & \text{if} \quad \widehat{g}(\widetilde{u}_t) \geq d_T, \quad |\widetilde{u}_t| \leq e_T, \quad |\widehat{g}'(\widetilde{u}_t)| \leq n_T \widehat{g}(\widetilde{u}_t) \\ 0 & \text{else}, \end{cases}$$

where $n_T, e_T \to \infty, d_T \to 0$ are trimming constants that obey Assumption A8 in the appendix. We estimate $\psi_1(u)$ and $\psi_2(u)$ by $\widehat{\psi}_1(u) = -\widehat{g}^{(1)}(u)/\widehat{g}(u)$ and $\widehat{\psi}_2(u) = -[1 - u\widehat{\psi}_1(u)]$, the efficient score by

$$s^*_\phi(\theta, \widehat{g}) = T^{-\frac{1}{2}} \sum_{t=1}^{T} \Gamma^*_{\phi t}(\theta, \widehat{g}) \psi(\widetilde{u}_t) \widehat{1}_t,$$

and the efficient information by

$$\widehat{\mathcal{J}}^*_{\phi\phi}(\theta, \widehat{g}) = T^{-1} \sum_{t=1}^{T} \Gamma^*_{\phi t}(\theta, \widehat{g}) \Gamma^*_{\phi t}(\theta, \widehat{g})' I(\widehat{g}),$$

where $I(\widehat{g}) = T^{-1} \sum_{t=1}^{T} \widehat{\psi}(\widetilde{u}_t) \widehat{\psi}(\widetilde{u}_t)' \widehat{1}_t$. The semiparametric test statistic is

$$\widehat{\tau}_c = \frac{c' s^*_\phi(\widehat{\theta}, \widehat{g})}{\left\{ c' \widehat{\mathcal{J}}^*_{\phi\phi}(\widehat{\theta}, \widehat{g}) c \right\}^{1/2}}, \tag{0.3}$$

where $\hat{\theta} = \tilde{\theta}_G - \widehat{\mathcal{J}}^{-1}_{\theta\theta}(\widetilde{\theta}, \widehat{g}) T^{-1/2} s_\theta(\widetilde{\theta}, \hat{g})$ is the semiparametric estimator of θ. As we show in Section 4, $\hat{\tau}_c$ is also asymptotically standard Gaussian.

4. First-Order Asymptotic Properties

We derive the limit distributions for parametric test statistics and prove that the semiparametric test statistics defined in Section 3 are asymptotically equivalent to the parametric test statistics and so are adaptive. To allow for a meaningful asymptotic power comparison among test statistics, we consider a sequence of local alternatives

$$\theta_T = \theta_0 + \delta T^{-1/2} \quad \text{for any } \delta \in \mathbb{R}^{k+p+1}, \tag{0.4}$$

where θ_0 is the true value of θ and $\delta = (\delta_\beta, \delta_\sigma, \delta_\gamma)$. We are primarily interested in the case that $\delta_\beta, \delta_\sigma = 0$ but $\delta_\gamma \neq 0$. To derive limit distributions we establish

[8] Engle and Gonzalez-Rivera (1991) also study a semiparametric estimator for the parameters of a GARCH model.

(in the appendix) that a parametric ARCH model is regular in the sense that its likelihood ratio has the local asymptotically normal (LAN) property. We are able to relax the symmetry assumption that is typically used to establish LAN, which may be important to applications of our theory as many financial time series are characterized by asymmetry. Our general approach is to be found in a number of other papers, notably Linton (1993), Steigerwald (1993), and Jeganathan (1995), Silvapulle, Silvapulle, and Basawa (1997), and Gonzalez-Rivera and Ullah (1998).

An immediate consequence of the LAN property is "Le Cam's Third Lemma" (see Bickel *et al.* (1994, page 503)), which delivers the asymptotic distribution of scalar test statistics τ_c under a sequence of local alternatives. Let $\Lambda_T = L(\theta_T, g) - L(\theta_0, g)$ be the log-likelihood ratio and let (τ_0, Λ_0) be a bivariate Gaussian random variable with mean $(\mu, \frac{-\sigma^2}{2})$ and covariance matrix $\begin{bmatrix} \eta^2 & \omega \\ \omega & \sigma^2 \end{bmatrix}$. Under the lemma, if

$$(\tau_c, \Lambda_T) \Rightarrow (\tau_0, \Lambda_0), \qquad \text{under } \theta_0, \tag{0.5}$$

then

$$(\tau_c, \Lambda_T) \Rightarrow (\tau_0 + \omega, \Lambda_0 + \sigma^2), \qquad \text{under } \theta_T.$$

The first result contains the limit distribution of the parametric test statistic under the sequence of local alternatives, from which one can calculate its local power, and shows that the semiparametric test statistic is asymptotically equivalent to the parametric test statistic. Let $\mathcal{J}^{\phi\phi}$ be the inverse of $\mathcal{J}_{\phi\phi}$.

THEOREM 1. *Suppose that the Assumptions A1 through A7 given in the appendix are satisfied. Then, for τ_c from (0.1), we have*

$$\sup_{-\infty<x<\infty} |\Pr(\tau_c \leq x) - \Phi[x - \mu(c)]| = o(1), \tag{0.6}$$

under θ_T, where $\mu(c) = \{c^T \mathcal{J}^{\phi\phi}(\theta_0, g)c\}^{1/2} \delta^T \mathcal{J}_{\phi\phi}(\theta_0, g)c$. If, in addition, Assumption A8 holds, then

$$\widehat{\tau}_c - \tau_c = o_p(1), \qquad \text{under } \theta_T. \tag{0.7}$$

PROOF. Convergence of τ_c to a $N(0,1)$, under θ_0, and the joint convergence (0.5), which follow by arguments similar to those contained in Linton (1993, Theorem 3), together imply (0.6).

For (0.7) it suffices to establish that $s^*_\phi(\theta_T, \widehat{g}) - s^*_\phi(\theta_T, g) = o_p(1)$ and $\widehat{\mathcal{J}}^*_{\phi\phi}(\theta_T, \widehat{g}) - \widehat{\mathcal{J}}^*_{\phi\phi}(\theta_T, g) = o_p(1)$, where θ_T is the deterministic sequence defined in (0.4). These results follow by arguments similar to those used in Linton (1993). ∎

REMARK. Gonzalez-Rivera (1997) calculates $\mu(c)$ for a number of densities. See Silvapulle, Silvapulle, and Basawa (1997, Theorem 2.3) for a similar result in an i.i.d. setting.

We next define our optimality criteria and show that τ_c and $\widehat{\tau}_c$ are asymptotically optimal. The critical function of τ is

$$\varphi_\alpha(\tau) = \begin{cases} 1 & \text{if } \tau > \kappa_\alpha \\ 0 & \text{if } \tau \le \kappa_\alpha, \end{cases}$$

where κ_α, with $\alpha \in (0,1)$, is a critical value, in our case determined by (0.6). Let $E_{T,\delta}$ denote expectation taken with respect to the measure P_{T,θ_T} of the sequence of local alternatives. Let Δ_0 and Δ_A be the set of δ values consistent with the null and alternative hypotheses respectively. In our case, $\Delta_0 = \{0\}$ and $\Delta_A = \mathbb{R}^l_+$ for some $l \ge 1$.

DEFINITION. A test statistic τ is *asymptotically unbiased* if

$$\begin{aligned} \limsup_{T\to\infty} E_{T,\delta}\varphi_\alpha(\tau) &\le \alpha, \text{ for all } \delta \in \Delta_0, \text{ and} \\ \liminf_{T\to\infty} E_{T,\delta}\varphi_\alpha(\tau) &\ge \alpha, \text{ for all } \delta \in \Delta_A. \end{aligned}$$

A test statistic τ is *maximin* if it is asymptotically unbiased and if for any other asymptotically unbiased statistic τ^*, we have

$$\limsup_{T\to\infty} \inf_{|\delta|=\epsilon} E_{T,\delta}\varphi_\alpha(\tau^*) \le \limsup_{T\to\infty} \inf_{|\delta|=\epsilon} E_{T,\delta}\varphi_\alpha(\tau),$$

for any $\epsilon > 0$.

For the case in which $m = 1$ we have

THEOREM 2. *If $m = 1$, then the test statistics τ_c and $\widehat{\tau}_c$ are asymptotically maximin.*

PROOF. Follows from Strasser (1985), Theorem 82.21. ∎

This result is the equivalent of the Locally Asymptotic Minimax result for estimation, see Hajek (1972). Theorem 2 implies that, excluding superefficient test statistics, local power is maximized by τ_c.

Theorem 2 does not apply to the case of $m > 1$, because the alternative region is a proper directed subset of the full Euclidean space. In this case we consider an alternative optimality criterion. Let $\omega(\theta)$ be a measure that gives probability one to the set of possible values for θ under the alternative hypothesis, and let τ

be a level α test with power function $\pi_\tau(\theta)$. Define the weighted average power criterion

$$\Psi = \int \pi_\tau(\theta) d\omega(\theta).$$

We say that τ is Ψ-optimal if it maximizes Ψ (possibly in an asymptotic sense). Following Sengupta and Vermeire (1986) we use a weight function that is uniform over arbitrarily small (local) neighborhoods.[9]

DEFINITION. A level α test τ is *locally most mean powerful unbiased* (LMMPU) if it is asymptotically unbiased and if for any other asymptotically unbiased level α test τ^*, there exists $\eta_0 > 0$ such that

$$\int_{\{|\theta-\theta_0|<\eta\}\cap H_A} \pi_\tau(\theta) d\theta > \int_{\{|\theta-\theta_0|<\eta\}\cap H_A} \pi_{\tau^*}(\theta) d\theta, \qquad \forall \eta < \eta_0.$$

This corresponds to a locally best (i.e. maximin) in the direction $\phi_1 = \ldots = \phi_p$. Lee and King (1993) show that the LMMPU test for the case in which $m > 1$ and u_t is Gaussian is based on the sum of scores, which accords with our construction of $\hat{\tau}_c$. We have

THEOREM 3. *If $m > 1$, then the test statistics τ_c and $\widehat{\tau}_c$ are asymptotically LMMPU.*

PROOF: This follows directly from the definition of an LMMPU test given by King and Wu (1991), and Theorem 1. ∎

REMARK: Although standard likelihood-based test statistics, such as the Lagrange multiplier, likelihood ratio, and Wald, maximize asymptotic power against two-sided local alternatives, they do not maximize asymptotic power against one-sided local alternatives. Further, the standard likelihood-based test statistics cannot be modified simply to take account of the one-sided alternative, H_A, if $m > 1$.

5. SECOND-ORDER ASYMPTOTIC PROPERTIES

5.1 Size Distortion

Theorems 1, 2 and 3 guarantee that $\hat{\tau}_c$ is asymptotically optimal. Of course, these results may hold only for very large sample sizes. To provide more insight

[9]Andrews (1994) uses a multivariate truncated normal distribution function for ω that is indexed by c, where c scales the covariance matrix of the weight function.

into the finite sample behavior of $\hat{\tau}_c$, we derive an asymptotic expansion of $\hat{\tau}_c$ that contains an error of smaller order than the error in (0.7), which is known only to be $o(1)$. The expansion is a second-order expansion, because the additional terms that are included are asymptotically negligible with respect to $\hat{\tau}_c - \tau_c$ and thus do not show up in the limiting distribution. In parametric problems, the correction terms are of order $T^{-1/2}$ in probability, see Rothenberg (1984), while in this semiparametric context, the correction terms are of order strictly larger than $T^{-1/2}$ in probability. Furthermore, the precise magnitude of the second-order terms depends on the bandwidth.

Mathematical derivation of the second-order term for $\hat{\tau}_c$ is demanding in the general framework under which the first-order theory is derived. Specifically, it is quite difficult to derive the expansion if $g(\cdot)$ is asymmetric and if trimming parameters are used to construct $\hat{g}(\cdot)$. To overcome these difficulties, we derive the second-order term under the assumption that $g(\cdot)$ is symmetric about zero and that $g(\cdot)$ has finite support and is bounded away from zero everywhere on that support, which obviates the need to trim $\hat{g}(\cdot)$. If the additional assumptions hold the modified test statistic is second-order optimal, if the additional assumptions do not hold the modified test statistic is first-order optimal. To determine how the modified test statistic performs in finite samples if the additional assumptions do not hold, we perform simulations.

If $g(\cdot)$ is symmetric about zero, then the semiparametric test statistic is

$$\hat{\tau}_c = \frac{T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\widehat{\psi}_2(\widetilde{u}_t)v_{t-i}^*}{\left\{T^{-1}\sum_{t=1}^{T}\widehat{\psi}_2^2(\widetilde{u}_t)\right\}^{1/2}\left\{T^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}v_{t-i}^{*2}\right\}^{1/2}},$$

where $v_t^* = \widetilde{u}_t^2 - T^{-1}\sum_{t=1}^{T}\widetilde{u}_t^2$ in which $\widetilde{u}_t = \widetilde{\sigma}^{-1}h_t^{-1}(\widetilde{\gamma})\left(y_t - \widetilde{\beta}' x_t\right)$. For convenience, we use a leave-p-out kernel density estimator, so that the index set in (0.2) is $\mathcal{T}_t = \{s : s \neq t, \ldots, t-p\}$.

The second-order expansion (full derivation is contained in the appendix) is

$$\hat{\tau}_c - \tau_c = h^2 A_1 + \frac{A_2}{T^{1/2}h^{3/2}} + o_p(\max\{h^2, T^{-1/2}h^{-3/2}\}) \equiv C + \mathfrak{R},$$

where the random variables A_1 and A_2 have zero mean and are uncorrelated with τ_c. The optimal bandwidth rate, in terms of minimizing the order of magnitude of the correction term C, balances the two terms and is $h = O(T^{-1/7})$ for which the remainder term $\mathfrak{R}$ is $o_p(T^{-2/7})$.

The second-order expansion yields an optimal bandwidth rate, but does not yield an optimal bandwidth value. To determine the optimal value of the bandwidth, we select the value of h that minimizes the variance of C. The following theorem gives the variance of C.

THEOREM 4. *If Assumptions A1-A9 hold, then under* H_0, as $T \to \infty$,

$$\operatorname{Var}(C) = h^4\left[\frac{E[u_t^2 b_t^2]}{I_2} - \left\{\frac{E[\psi_2(u_t)u_t b_t]}{I_2}\right\}^2\right] + T^{-1}h^{-3}\nu_2(K^{(1)})\frac{E[u_t^2 g^{-1}(u_t)]}{I_2}, \tag{0.8}$$

where $I_2 = E[\psi_2^2(u_t)]$, $\mu_2(K) = \int s^2 K(s)ds$, $\nu_2(K) = \int K^2(s)ds$, *and*
$b_t = \frac{1}{2}\mu_2(K)\left\{\frac{g^{(3)}}{g}(u_t) - \frac{g^{(2)}}{g}\frac{g^{(1)}}{g}(u_t)\right\}$.

PROOF: See appendix. ∎

REMARK. The first term on the right-hand side of (0.8) is a function of b_t, which is the bias of the nonparametric estimator of the score function.[10] The second term in the right-hand side of (0.8) is the variance of A_2. Because A_2 is a weighted degenerate U-statistic, from recent work by Fan & Li (1996) and Hjellvik, Yao, and Tjøstheim (1996) we expect that C is asymptotically normal with mean zero and variance given by Theorem 4. Because b_t and $\psi_2(u_t)$ are functions of $g(\cdot)$, the magnitude of the variance of C depends on the underlying density.[11] From Theorem 4 we see that the variance of C is of order $T^{-4/7}$ and strictly positive (unless I_2 is very small), which suggests that the empirical size will exceed the nominal size if critical values from the asymptotic Gaussian distribution are used.

We now turn to the question of bandwidth selection. Because (0.8) is derived under the null hypothesis, selecting h to minimize (0.8) is analogous to selecting h to minimize the second-order size distortion.[12] The optimal bandwidth from Theorem 4 is a function of $g(\cdot)$ and the first three derivatives of $g(\cdot)$, all of which are unknown. One could substitute nonparametric estimators of these quantities into the bandwidth formula, but nonparametric estimators of the second and third derivatives of $g(\cdot)$ can be badly behaved even with moderate sized samples. To overcome the difficulty of nonparametrically estimating the higher derivatives of $g(\cdot)$, we use a rule-of-thumb bandwidth selection method pioneered by Silverman (1986). To control second-order size distortion it is not appropriate to use cross validation to select h. Under cross validation, h is of order $T^{-1/5}$, which does not minimize the order of C.

[10]To see that b_t is the bias, substitute the standard formulas for the bias of nonparametric estimators of $g(\cdot)$ and $g^{(1)}(\cdot)$ into the approximation

$$\frac{\widehat{g}^{(1)}}{\widehat{g}}(u) \approx \frac{\widehat{g}^{(1)}(u) - g^{(1)}(u)}{g(u)} - \frac{g^{(1)}(u)}{g(u)}\frac{\widehat{g}^{(1)}(u) - g^{(1)}(u)}{g(u)}.$$

[11]The magnitude of p does not affect the second-order term, although it does affect the order T^{-1} term in the variance, which we have not calculated.

[12]The second-order size distortion is also a function of the bias and skewness of C. We do not focus on the bias and skewness of C because they do not depend on h. Because the bias and skewness of C are of larger order (order $T^{-1/2}$) they could increase or decrease the size of the test statistic.

The method works as follows. We calculate the various quantities in (0.8) for the standard Gaussian density and thereby determine an optimal bandwidth. Because the resulting bandwidth value is second-order optimal if u_t is a truncated Gaussian random variable, the bandwidth value should be approximately second-order optimal under small departures from a truncated Gaussian distribution. If the innovation density is a standard Gaussian density, then $g^{(1)}(u) = -ug(u)$, $g^{(2)}(u) = (u^2-1)g(u)$, and $g^{(3)}(u) = -(u^3-3u)g(u)$. Thus, $\psi_2(u) = (u^2-1)$, $I_2 = 2$, and $b(u) = u\mu_2(K)$, so $E[u_t^2 b(u_t)^2] = 3\mu_2^2(K)$, $E[u_t\psi_2(u_t)b(u_t)] = 2\mu_2(K)$, $I_2^{-1}E(u_t^2 b_t^2) - \left\{I_2^{-1}E[u_t\psi_2(u_t)b_t]\right\}^2 = \frac{1}{2}\mu_2^2(K)$. We replace $E[u_t^2 g^{-1}(u_t)]$ by a consistent estimator $(u_{\max}^3 - u_{\min}^3)/3$. The rule-of-thumb bandwidth is

$$\widehat{h}_{rot} = \left\{\frac{\nu_2(K^{(1)})(u_{\max}^3 - u_{\min}^3)}{4\mu_2^2(K)}\right\}^{1/7} \widehat{\sigma}T^{-1/7}, \tag{0.9}$$

where $\widehat{\sigma}$ is an estimator of the standard deviation of the residual $\widetilde{u}_t$, while $u_{\max}$ and $u_{\min}$ are the maximum and minimum respectively of $\{\widetilde{u}_t/\widehat{\sigma}; t = 1, \ldots, T\}$. For the Gaussian kernel, $\mu_2(K) = 1$ and $\nu_2(K^{(1)}) = 1/4\sqrt{\pi}$.[13] The test statistic $\widehat{\tau}_c$ that uses $\widehat{h}$ in (0.2) asymptotically minimizes the second-order size distortion, as defined above, for the Gaussian density.

We recognize that if $g(\cdot)$ has unbounded support, the right-hand side of the asymptotic variance formula of Theorem 4 is infinite. Even so, the bandwidth constant estimate is not totally unfounded; it too will increase without bound thus reflecting this reality. For example, with Gaussian data $u_{\max} - u_{\min}$ will grow at rate $\ln T$ and so the estimated bandwidth will be larger in magnitude than order $T^{-1/7}$. By contrast, methods based on the integrated mean square error of either $\hat{g}(\cdot)$ or $\hat{g}^{(1)}(\cdot)$, such as cross validation, result in a bandwidth magnitude that is the same whether or not the support for $g(\cdot)$ is bounded. As a result, cross-validation results in a bandwidth that is too small if the support for $g(\cdot)$ is unbounded.

5.2 Bias Reduction

The bias term b_t in Theorem 4 is nonzero for the Gaussian density, and hence our procedure has a bias-related variance correction term even in this canonical case. To reduce bias, we study here the effect of replacing the standard kernel density estimator by the estimator proposed by Jones, Linton, and Nielsen (1995), hereafter JLN. The JLN estimator reduces the bias of the density estimator to order h^4, from order h^2, for all $g(\cdot)$ possessing four continuous derivatives. It also guarantees a positive estimate of $g(\cdot)$ everywhere, unlike other bias reduction methods such as higher-order kernel density estimators. In practice, the reduction in bias allows a wider bandwidth to be used, which translates into gains in the second-order performance of our test statistics for at least some region of the

[13] Härdle & Linton (1994) give the magnitude of ν_2 and μ_2 for several kernels.

(functional) parameter space.

The JLN density estimator based on observed data $\{u_t\}_{t=1}^T$ is

$$\widetilde{g}(u) = \widehat{g}(u)\ (Th)^{-1} \sum_{t=1}^{T} \widehat{g}^{-1}(u_t) K\left(\frac{u-u_t}{h}\right), \tag{0.10}$$

where $\widehat{g}(\cdot)$ is defined in (0.2) and $\widetilde{\psi}_1(\cdot) = \widetilde{g}^{(1)}(\cdot)/\widetilde{g}(\cdot)$. As JLN show, the biases of $\widetilde{g}^{(1)}(u)$ and hence of $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$ are also order h^4 compared with order h^2 for $\widehat{g}^{(1)}(u)$ and $\widehat{g}^{(1)}(u)/\widehat{g}(u)$ (provided that $g(\cdot)$ possesses two additional continuous derivatives). The reduced bias permits a faster mean-square error rate of $n^{-8/9}$ as compared with the best possible rate of $n^{-4/5}$ for the standard kernel density estimator. This translates into a second-order variance correction to $\widehat{\tau}_c$ of order $n^{-8/11}$, provided $h = O(n^{-1/9})$, which is an improvement over the order $n^{-4/7}$ correction in Theorem 4. The bias constant for $\widetilde{g}^{(1)}(u)$ is $\mathfrak{b}'(u) = \frac{1}{4}\mu_2^2(K) g^{(1)}(u)\{\frac{g^{(3)}}{g^{(1)}}(u)\}^{(2)}$; therefore, the bias constant for $\frac{\widetilde{g}^{(1)}(u)}{\widetilde{g}(u)}$ is

$$\frac{\mathfrak{b}'(u)}{g(u)} - \frac{g^{(1)}(u)}{g(u)}\frac{\mathfrak{b}(u)}{g(u)},$$

with $\mathfrak{b}(u) = \frac{1}{4}\mu_2^2(K) g(u)\{\frac{g^{(2)}}{g}(u)\}^{(2)}$, which is the relevant quantity appearing in Theorem 4. If $g(\cdot)$ is the Gaussian density, then $\mathfrak{b}(u) = \frac{1}{2}\mu_2^2(K)\phi(u)$ and $\mathfrak{b}'(u) = -\frac{1}{2}\mu_2^2(K) u\phi(u)$. In this case, there is, remarkably, a cancellation, and the bias of $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$ is the even better $o(h^4)$. This means that if we use the JLN estimate in place of the standard kernel density estimate when constructing $\widehat{\tau}_c$ we shall have a second-order correction to the variance that can be made to be smaller than $n^{-8/11}$ in magnitude, i.e. improves on the magnitudes in Theorem 4.[14] This improved performance should also hold for any statistic depending on $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$.

6. Finite Sample Performance

The results in Section 4 indicate that with a large number of observations, a semiparametric test statistic outperforms a quasi-maximum likelihood test statistic. The question is, how well does a semiparametric test statistic perform with a small number of observations? To shed light on the issue, we run simulations for samples of 100 and 500 observations. Because most financial data sets have substantially larger numbers of observations, our results provide conservative estimates of the gains achievable in practice.

A semiparametric test statistic is asymptotically more powerful than a QML test statistic if the innovation density is non-Gaussian. Therefore, in the simulations we conduct the true innovation density is either asymmetric, leptokurtic, or

[14] As before, replacing unobserved errors u_t by root-T consistently estimated residuals $\widetilde{u}_t$ makes no difference to the first-order properties of $\widetilde{g}(u)$ and hence the second-order properties of the semiparametric test statistic.

platykurtic. Asymmetric innovation densities that have more mass concentrated in the tails result in marginal densities for y_t that capture the large number of outliers and the asymmetric pattern in certain exchange rate series. The specific asymmetric densities that we consider are log-normal densities that are constructed from a Gaussian density with variance that takes values (0.01, 0.10, 1.00). Leptokurtic innovation densities result in marginal densities for y_t that capture both the large number of outliers and the shape of many daily exchange rate series. The specific leptokurtic densities that we consider are t densities with 30, 8, and 5 degrees of freedom, respectively. Platykurtic innovation densities result in marginal densities for y_t that capture the large number of outliers and the effect of the random arrival of information that characterize many asset return series. The specific platykurtic densities that we consider are bimodal symmetric mixtures of Gaussian random variables with means that take values $(\pm 1, \pm 2, \pm 10)$. In the tables summarizing the results each of the densities is denoted by a capital letter for **A**symmetric, **L**eptokurtic, or **P**latykurtic together with a number 1, 2, or 3, where a larger number corresponds to a larger departure from a Gaussian density. Thus L2 denotes the t density with 8 degrees of freedom. Summary statistics for the densities are in Table 1. To allow for sensible comparisons across the different densities, all variables drawn from the densities in Table 1 are subsequently rescaled to have mean 0 and variance 1.

We simulate an ARCH(p) specification with $\mu_t(\gamma) = \beta_0 + \beta_1 x_{1t}$. For the conditional mean we set $\beta_0 = 1$, $\beta_1 = -1$, and take x_{1t} to be i.i.d. Gaussian (0,1) and independent of $\sigma h_t(\gamma) u_t$. We perform 1000 simulations.

The test statistics are constructed from the semiparametric and QML estimators, which are constructed using the method of scoring. Specifically, the QMLE is constructed as

$$\widehat{\gamma}^{i}_{TQM} = \widehat{\gamma}^{i-1}_{TQM} + \lambda \mathcal{J}_T^{-1}(\widehat{\gamma}^{i-1}_{TQM}, g^n) s^*_{\gamma}(\widehat{\gamma}^{i-1}_{TQM}, g^n), \tag{0.11}$$

where g^n is a Gaussian density and λ is a parameter that controls the size of the updating step.[15] We iterate (0.11) until $s^*_{\gamma}(\widehat{\gamma}^{i-1}_{TQM}, g^n)^T \mathcal{J}_T^{-1}(\widehat{\gamma}^{i-1}_{TQM}, g^n) s^*_{\gamma}(\widehat{\gamma}^{i-1}_{TQM}, g^n)$ is less than 0.01.

The semiparametric estimator is constructed as in (0.11) with $\widehat{g}$ used in place of a Gaussian density, where $\widehat{g}$ is constructed using the residuals calculated from $\widehat{\gamma}^{i-1}_T$ and $\widehat{\gamma}^0_T = \widehat{\gamma}^0_{TQM}$. The nonparametric estimator of g is constructed with the quartic kernel $K(u) = \frac{15}{16}[1-u^2]^2 \mathbf{1}(|u| \leq 1)$.

We study two important issues for practical implementation of semiparametric test statistics. First, we compare a standard nonparametric estimator of g, given by (0.2), with a JLN reduced-bias estimator of g, given by (0.10). Second, we

[15] At the beginning of each iteration, λ equals 1. If the value of $\hat{\gamma}^i_{TQM}$ does not increase the log-likelihood, then λ is set to $\frac{1}{2}$. If the resulting value of $\hat{\gamma}^i_{TQM}$ does not increase the log-likelihood the process is repeated, shrinking λ by a factor of 2 each time until a step is found that increases the log-likelihood.

TABLE 1
Density Summary Statistics

Name	Construction	Mean	Variance	Skewness	Kurtosis
Asymmetric 1	$\exp(z)$ where z is $N(0, 0.01)$	1.01	0.01	0.30	3.16
Asymmetric 2	$\exp(z)$ where z is $N(0, 0.10)$	1.05	0.12	1.01	4.86
Asymmetric 3	$\exp(z)$ where z is $N(0, 1.00)$	1.65	4.67	6.18	113.94
Leptokurtic 1	$t(30)$	0.00	1.07	0.00	3.20
Leptokurtic 2	$t(8)$	0.00	1.33	0.00	4.50
Leptokurtic 3	$t(5)$	0.00	1.67	0.00	9.00
Platykurtic 1	$.5[N(-1,1) + N(1,1)]$	0.00	2.00	0.00	2.50
Platykurtic 2	$.5[N(-2,1) + N(2,1)]$	0.00	5.00	0.00	1.72
Platykurtic 3	$.5[N(-10,1) + N(10,1)]$	0.00	101.00	0.00	1.04

compare the value of the smoothing parameter that minimizes second-order size distortion, given by $\hat{h}$ in (0.9), with other values of the smoothing parameter. In particular, because the JLN density estimator has reduced bias, we can use a smoothing parameter that is larger than $\hat{h}$ in forming the JLN density estimator.[16] To determine the value of the smoothing parameter that maximizes the size-adjusted power of a semiparametric test statistic we examine the values $h = c \cdot \hat{h}$, where c takes values (0.5,1.0,1.5,2.0).

The first testing problem that we consider is the univariate testing problem. Specifically, we study the test of the null hypothesis that the model is ARCH(1) against the alternative hypothesis that the model is ARCH (2). The ARCH(1) specification is $h_t(\gamma_0)^2 = 1 + 0.1(y_{t-1} - 1 + x_{1t-1})^2$ and the ARCH(2) specification is $h_t(\gamma_0)^2 = 1 + 0.1(y_{t-1} - 1 + x_{1t-1})^2 + 0.5(y_{t-2} - 1 + x_{1t-2})^2$. Thus we test a null model with only weak ARCH effects against an alternative model with substantially more ARCH effects.

In Table 2 we compare the positive square-root of the Lagrange multiplier test statistic constructed from a Gaussian QMLE, denoted QML, with three semiparametric test statistics, for a sample of 100 observations. The first semiparametric test statistic, denoted SP1, is constructed using the standard nonparametric estimator of g from (0.2) with the value of the smoothing parameter given by $\hat{h}$ in (0.9). The second semiparametric test statistic, denoted SP2, is constructed using the JLN estimator of g from (0.10) with the value of the smoothing parameter given by $\hat{h}$ in (0.9). The third semiparametric test statistic, denoted SP3, is constructed using the JLN estimator of g from (0.10) with the value of the smoothing

[16] Because the standard density estimator does not offer reduced bias, we restrict attention to $\hat{h}$ for this estimator.

TABLE 2

Size and Size-Adjusted Power: ARCH(1) vs. ARCH(2) T=100

Density	QML	SP1	SP2	SP3
Size				
Gaussian	0.027	0.030	0.038	0.030
A1	0.025	0.040	0.025	0.020
A2	0.023	0.049	0.029	0.028
A3	0.025	0.035	0.045	0.019
L1	0.016	0.043	0.027	0.020
L2	0.024	0.035	0.037	0.034
L3	0.021	0.035	0.033	0.032
P1	0.021	0.030	0.026	0.023
P2	0.032	0.024	0.033	0.033
P3	0.028	0.000	0.008	0.000
Size-Adjusted Power				
Gaussian	0.782	0.678	0.644	0.704
A1	0.784	0.659	0.616	0.704
A2	0.649	0.621	0.603	0.669
A3	0.325	0.579	0.647	0.599
L1	0.753	0.637	0.593	0.683
L2	0.700	0.560	0.509	0.598
L3	0.593	0.459	0.396	0.507
P1	0.858	0.782	0.757	0.821
P2	0.962	0.933	0.949	0.937
P3	0.957	0.974	0.999	0.932

parameter given by $1.5 \cdot \hat{h}$.[17] Use of the optimal bandwidth derived from second-order theory is important. If the simple rule-of-thumb value $h = T^{-\frac{1}{5}}$ is used to construct a semiparametric test statistic, the power is generally reduced by a factor of 2.

[17]Results for other values of c, namely 0.5 and 2.0, are not separately reported. Reducing the smoothing parameter, $c = 0.5$, reduced the size-adjusted power of a semiparametric test statistic for every density. Increasing the smoothing parameter further, $c = 2.0$, increased the size-adjusted power for the leptokurtic densities but reduced, and in some cases greatly reduced, the size-adjusted power for the remaining densities.

The upper panel contains the empirical size of the test statistics for a test with a nominal size of five percent. The lower panel contains the size-adjusted power for each of the test statistics. To construct the size-adjusted power for each test statistic, we use critical values that correspond to an empirical size of five percent if the empirical size exceeds five percent and use nominal five percent critical values otherwise. Within a panel, each row of the table corresponds to a different innovation density and each column corresponds to a different semiparametric test statistic. (Tables 3 and 4 are constructed similarly.) The third through fifth columns, headed by SP1, SP2, and SP3, respectively, contain the empirical sizes for the positive square root of the Lagrange multiplier test statistic constructed from each of the nonparametric density estimators described above. For each density all four test statistics have empirical size that is below nominal size.

To compare size-adjusted power, we begin with the semiparametric test statistics. In comparing SP1 with SP2, we see that for seven of the ten densities the standard density estimator delivers a higher size-adjusted power than the JLN density estimator if both use the same value of the smoothing parameter. Only for three of the densities with the greatest departures from normality does the SP2 test statistic outperform the SP1 test statistic, and in two of these cases the power gain is slight. The real advantage in using the JLN density estimator comes from the ability to increase the value of the smoothing parameter. For nine of the ten densities the JLN estimator with the increased smoothing parameter outperforms the standard density estimator and for seven of the ten densities SP3 outperforms SP2. Again, the three densities where SP2 has highest power represent extreme departures from normality. In comparing the size-adjusted power of the QML test statistic with the preferred semiparametric test statistic SP3, we see that for eight of the ten densities the QML test statistic has higher power. In general, the relative performance of the semiparametric test statistic improves as the departure from normality grows. It appears that a sample size in excess of 100 observations is needed to capture the efficiency gains of a semiparametric test statistic.

Because a sample of 100 observations is fairly small, we compare a QML test statistic with the preferred semiparametric test statistic for a larger sample of 500 observations in Table 3.
Because the sample size is increased, the alternative hypothesis must be changed to keep the power below 1. As explained in previous sections, the magnitude of the alternative hypothesis shrinks at rate $T^{1/2}$, so the ARCH(2) specification is $h_t(\gamma_0)^2 = 1 + 0.1(y_{t-1} - 1 + x_{1t-1})^2 + 0.22(y_{t-2} - 1 + x_{1t-2})^2$. The second and third columns contain the empirical size for the QML and SP3 test statistics, respectively. The fourth and fifth columns contain the size-adjusted power for the test statistics. For each density, both test statistics have empirical size that is below nominal size and for seven of the ten densities the size distortion (the difference between the empirical size and the nominal size) of SP3 is reduced as the sample size increases. In comparing size-adjusted power, we see that for eight of

TABLE 3

ARCH(1) vs. ARCH(2) T=500

Density	QML	SP3	QML	SP3
Gaussian	0.030	0.030	0.945	0.937
A1	0.026	0.027	0.920	0.934
A2	0.024	0.035	0.804	0.879
A3	0.021	0.027	0.390	0.689
L1	0.026	0.025	0.915	0.909
L2	0.024	0.030	0.844	0.838
L3	0.027	0.035	0.736	0.752
P1	0.030	0.029	0.964	0.972
P2	0.034	0.025	0.996	0.997
P3	0.046	0.000	0.995	0.975

the ten densities the two test statistics are virtually identical and for the remaining two densities SP3 has higher power. Because the two densities for which SP3 has higher power are asymmetric, our simulations indicate that for univariate testing problems, the most substantial gains from a semiparametric estimator occur with asymmetric densities.

The second testing problem that we consider is the multivariate testing problem. Specifically, we study the test of the null hypothesis that the model is white noise against the alternative hypothesis that the model is ARCH(2). The ARCH(2) specification is the same specification used in the univariate testing problem.

In Table 4 we compare a multivariate QML test statistic, constructed from a Gaussian QMLE, with three multivariate semiparametric test statistics, constructed from each of the nonparametric density estimators described above, for a sample of 100 observations. Each of the test statistics is formed as a sum of scores, given by (0.3) with c a vector of ones.

The upper panel contains the empirical size of the test statistics for a test with a nominal size of five percent. The lower panel contains the size-adjusted power for each of the test statistics. For each density all test statistics have empirical size that is below nominal size. To compare size-adjusted power, which again is simply raw power, we begin with the semiparametric test statistics. In comparing SP1 with SP2, we see a sharp contrast with the univariate results. For each of the ten densities, the JLN density estimator delivers a higher size-adjusted power than the standard density estimator if both use the same value of the smoothing parameter. Once again, increasing the value of the smoothing parameter can increase the size-adjusted power of a semiparametric test statistic that uses the JLN estimator. For nine of the ten densities the JLN estimator with the increased

TABLE 4

Size and Size-Adjusted Power: White Noise vs. ARCH(2)

Density	QML	SP1	SP2	SP3
Size				
Gaussian	0.015	0.014	0.014	0.012
A1	0.013	0.013	0.014	0.014
A2	0.009	0.013	0.016	0.016
A3	0.002	0.009	0.020	0.010
L1	0.017	0.019	0.018	0.019
L2	0.012	0.010	0.010	0.016
L3	0.001	0.011	0.011	0.010
P1	0.017	0.013	0.017	0.017
P2	0.027	0.011	0.020	0.015
P3	0.027	0.000	0.000	0.000
Size-Adjusted Power				
Gaussian	0.425	0.435	0.467	0.525
A1	0.435	0.440	0.468	0.532
A2	0.333	0.414	0.469	0.496
A3	0.072	0.314	0.410	0.374
L1	0.452	0.447	0.470	0.535
L2	0.362	0.403	0.421	0.493
L3	0.265	0.319	0.321	0.395
P1	0.522	0.517	0.544	0.593
P2	0.687	0.709	0.768	0.737
P3	0.679	0.440	0.824	0.312

smoothing parameter outperforms a standard density estimator and for seven of the ten densities SP3 outperforms SP2. In comparing the size-adjusted power of the QML test statistic with the preferred semiparametric test statistic, SP3, we see that for nine of the ten densities the semiparametric test statistic has higher power. In contrast to the univariate testing result, with a sample of only 100 observations a multivariate semiparametric test statistic outperforms a QML test statistic for nine of the ten densities.

7. Conclusion

The semiparametric test statistics are asymptotically optimal, dominating the widely used Gaussian test statistics according to standard criteria. The simulations provide several interesting results on the finite sample performance of our proposed semiparametric test statistic. First, even though the second-order optimal rule-of-thumb bandwidth is derived under the assumption that the innovation density is symmetric (and bounded), the method performs very well if the innovation density is asymmetric (and unbounded). Second, even with a sample of 500 observations, the semiparametric test delivers power gains only for asymmetric innovations for a univariate test. Third, even with a sample of only 100 observations, the semiparametric test statistic delivers power gains on the order of 10 percent for a multiparameter test.

Acknowledgments

We thank Steve Fox for help with the computations. We acknowledge financial support from the National Science Foundation.

References

[1] Andrews, D. (1994) Hypothesis testing with a restricted parameter space. Cowles Foundation Discussion Paper 1060R. Yale University.

[2] Andrews, D. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11: 560-598.

[3] Baillie, R. and T. Bollerslev. (1989) The message in daily exchange rates: a conditional-variance tale. *Journal of Business and Economic Statistics* 7: 297-305.

[4] Bera, A. and M. Higgins. (1993) ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7: 305-66.

[5] Bera, A. and P. Ng. (1991) Robust tests for heteroskedasticity and autocorrelation using score function. Economics Department Manuscript. University of Illinois.

[6] Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner. (1994) *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

[7] Bollerslev, T., and J. Wooldridge. (1992) Quasi maximum likelihood

estimation and inference in dynamic models with time varying covariates. *Econometric Reviews* 11: 143-172.

[8] ENGLE, R. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987-1007.

[9] ENGLE, R. (1983) Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in *Handbook of Econometrics, Volume II,* Z. Griliches and M. Intriligator (editors). New York: North Holland.

[10] ENGLE, R. AND G. GONZALEZ-RIVERA. (1991) Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9: 345-360.

[11] FAN, Y. AND Q. LI. (1996) Central limit theorems for degenerate U-statistics of absolutely regular processes with applications to model specification testing. Economics Department Manuscript 1996-8. Guelph University.

[12] GONZALEZ-RIVERA, G. (1993) A note on adaptation in GARCH models. *Econometric Reviews* 16: 55-68.

[13] GONZALEZ-RIVERA, G. AND A. ULLAH. (1998) Rao's score test with non-parametric density estimators. *Journal of Statistical Planning and Inference*, forthcoming.

[14] HAJEK, J. (1972) Local asymptotic minimax and admissibility in estimation, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, L. Le Cam, J. Neyman, and E. Scott (editors), Berkeley: University of California Press.

[15] HÄRDLE, W., AND O. LINTON. (1994) Applied nonparametric methods, in *Handbook of Econometrics*, Volume IV, D. McFadden and R. Engle (editors). New York: North Holland.

[16] HÄRDLE, W., AND T. STOKER. (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84: 986-995.

[17] HJELLVIK, V., Q. YAO, AND D. TJØSTHEIM. (1996) Linearity testing using local polynomial approximation. *Journal of Planning and Statistical Inference*, forthcoming.

[18] HSIEH, D. (1989) Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics* 7: 307-317.

[19] JEGANATHAN, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11: 818-887.

[20] JONES, M., O. LINTON, AND J. NIELSEN. (1995) A simple and effective bias reduction method for density estimation. *Biometrika* 82: 93-100.

[21] KING, M. AND P. WU. (1991) Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* 16: 131-156.

[22] LE CAM, L. (1960) Locally asymptotically normal families of distributions. *University of California Publications in Statistics* 3: 37-98.

[23] LEE, J. AND M. KING. (1993) A locally most mean powerful score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics* 11: 17-27.

[24] LINTON, O. (1993) Adaptive estimation in ARCH models. *Econometric Theory* 9: 539-569.

[25] LINTON, O. (1995) Second order approximation in the partially linear model. *Econometrica* 63: 1079-1113.

[26] LUMSDAINE, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. *Econometrica* 64: 575-596.

[27] NELSON, D. (1990) Stationarity and persistence in the GARCH(1,1) models. *Econometric Theory* 6: 318-334.

[28] NELSON, D. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.

[29] NELSON, D. AND C. CAO. (1991) Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10: 229-235.

[30] NEWEY, W. AND D. STEIGERWALD. (1997) Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroscedasticity models. *Econometrica* 65: 587-599.

[31] ROBINSON, P. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4: 185-208.

[32] ROBINSON, P. (1987) Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* 56: 875-891.

[33] ROTHENBERG, T. (1984) Approximation the distributions of econometric estimators and test statistics, in *Handbook of Econometrics, Volume II*, Z. Griliches and M. Intriligator (editors). New York: North-Holland.

[34] ROUSSAS, G. (1972) *Contiguity of Probability Measures*. Cambridge: Cambridge University Press.

[35] SENGUPTA, A., AND L. VERMEIRE. (1986) Locally optimal tests for multiparameter hypotheses. *Journal of the American Statistical Society* 81: 819-825.

[36] SILVAPULLE, M., P. SILVAPULLE, AND I. BASAWA. (1997) On adaptive tests. *Proceedings of the Australasian Econometric Society, Melbourne Unversity Volume 2*, 480-499.

[37] SILVERMAN, B. (1986) *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

[38] STEIGERWALD, D. (1993) Efficient estimation of models with conditional heteroscedasticity. Economics Department Manuscript 1993-5. University of California, Santa Barbara.

[39] STRASSER, H. (1985) *Mathematical Theory of Statistics.* Berlin: de Gruyter.

[40] SWENSEN, A. (1985) The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* 16: 54-70.

[41] WEISS, A. (1986) Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2: 107-131.

APPENDIX

To calculate the quantity $\Gamma^*_{\phi t}(\theta, g)$, we first define $\Gamma_t(\theta, g)$. Let

$$w_{t-1}(\gamma) = -\frac{\sum_{i=1}^p \phi_i x_{t-i}(y_{t-i} - \beta' x_{t-i})}{1 + \sum_{i=1}^p \phi_i (y_{t-i} - \beta' x_{t-i})^2} \quad .$$

Then

$$\begin{bmatrix} \frac{\partial l_t}{\partial \beta} \\ \frac{\partial l_t}{\partial \sigma} \\ \frac{\partial l_t}{\partial \phi} \end{bmatrix} = \begin{pmatrix} \sigma^{-1} h_t(\gamma)^{-1} x_t & w_{t-1}(\gamma) \\ 0 & \sigma^{-1} \\ 0 & v_{t-1}(\gamma)/2 \end{pmatrix} \begin{pmatrix} \psi_1(u_t(\theta)) \\ \psi_2(u_t(\theta)) \end{pmatrix} = \Gamma_t(\theta)\psi(u_t(\theta)),$$

where Γ_t depends only on $\mathcal{F}_{t-1}$ and $\psi(u_t(\theta_0))$ is mean zero and independent of $\mathcal{F}_{t-1}$. To construct the efficient score for ϕ, we use the information matrix for θ. Although there are a number of alternative asymptotically equivalent versions of the information matrix, we find the conditional information $\mathcal{J}_{\theta\theta}(\theta, g) = T^{-1} \sum_{t=1}^T E[\frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} | \mathcal{F}_{t-1}]$ particularly convenient because of the following representation

$$\text{vec}[\mathcal{J}_{\theta\theta}(\theta, g)] = T^{-1} \sum_{t=1}^T \{\Gamma_t(\theta) \otimes \Gamma_t(\theta)\} \text{vec} \begin{bmatrix} I_1(g) & I_{12}(g) \\ I_{12}(g) & I_2(g) \end{bmatrix} \equiv C_T(\theta)\text{vec}[I(g)],$$

where $I_1 = E[\psi_1^2(u_t)]$, $I_2 = E[\psi_2^2(u_t)]$ and $I_{12} = E[\psi_1(u_t)\psi_2(u_t)]$. The efficient score is constructed from the relevant elements of the information matrix.

For any vector x with element j denoted x_j, let $|x|$ be the Euclidean norm of x, that is $|x| = \left[\sum_j x_j^2\right]^{\frac{1}{2}}$. We make the following assumptions:

A1. *The random variables $u_1, \ldots, u_T$ are i.i.d. with absolutely continuous Lebesgue density g, and there exists a contiguous set $\mathcal{H} \subseteq \mathbb{R}$ on which $g(u) > 0$ and $\int_{\mathcal{H}} g(u)du = 1$.*

A2. *The moments $\int u^4 g(u)du$ and $\int \psi_j^4(u)g(u)du$, $j = 1, 2$, are finite.*

A3. *The density g is twice boundedly continuously differentiable.*

A4. *The parameter space Θ is an open subset of $\mathbb{R}^{k+p+1}$ that satisfies various restrictions such that*

(a) *The process $\{h_t^2\}_{t=1}^{\infty}$ is bounded below by a constant $\underline{h} > 0$.*

(b) *The process $\{h_t^2\}_{t=1}^{\infty}$ is strictly stationary and ergodic.*

(c) *The information matrix $\mathcal{J}_{\theta\theta}(\theta, g)$ is nonsingular at θ_0.*

(d) *The quantity Eh_t^4 is finite for all t.*

A5. *The initial condition density $g_0(Y_0; \theta)$, where $Y_0 = (y_0, y_{-1}, \ldots, y_{-p})$, is continuous in probability: i.e. $g_0(Y_0; \theta_T) \xrightarrow{P} g_0(Y_0; \theta)$,* for any $\theta_T \to \theta$.

A6. *The regressors $\{x_t\}_{t=1}^{T}$ are weakly exogenous for θ and $T^{-1}\sum_{t=1}^{T} x_t x_t^T \xrightarrow{P} M$, where M is a positive definite matrix.*

A7. *Both $\int[\psi_1^{(1)}(u)]^2 g(u)du$ and $\int[\psi_2^{(1)}(u)]^2 g(u)du$ are finite.*

A8. *The kernel K has bounded support and is twice continuously differentiable. The bandwidth sequence satisfies: $h(T), d_T \to 0$, $e_T, n_T \to \infty$, $h(T)n_T \to 0$, and $Th(T)^3 n_T^{-2} e_T^{-2} \to \infty$.*

A9. *The density function g has bounded support. Both g and K are many times continuously differentiable, K nonvanishingly so.*

REMARK: A sufficient condition for A4(c) is that h_t have bounded second moment, see Weiss (1986). However, Lumsdaine (1996) weakened this condition somewhat, and allows for processes with total roots exceeding one. The conditions on the regressors can be relaxed in various directions: for example, Swensen (1985) allows for deterministic trends in the regressors, while Jeganathan (1995) allows for integrated regressors and derives the more general result of Local Asymptotic

Mixed Normality in this case. Assumption A9 is made to simplify algebra for the second-order asymptotic theory and is not necessary for the first-order asymptotic theory.

LEMMA 1 ***(Local Asymptotic Normality).*** *Let $\Lambda_T = L(\theta_T, g) - L(\theta_0, g)$ be the log-likelihood ratio and suppose that assumptions A1-A6 are satisfied. Then*

$$\Lambda_T - \delta^T s_\theta(\theta_0, g) + \frac{1}{2}\delta^T \mathcal{J}_{\theta\theta}(\theta_0, g)\delta \xrightarrow{P} 0,$$

and $s_\theta(\theta_0, g) \Rightarrow N(0, \mathcal{J}_{\theta\theta}(\theta_0, g))$, where convergence is under the probability measure induced by θ_0. Furthermore, the probability measures P_{T,θ_0} and P_{T,θ_T} are mutually contiguous in the sense of Roussas (1972, Definition 2.1, p7): i.e. $P_{T,\theta_0}(A) \to 0$ if and only if $P_{T,\theta_T}(A) \to 0$, for any event A.

PROOF. Swensen (1985) lists six conditions that together imply the LAN property. Linton (1993) verifies the six conditions for a parametric ARCH model under the assumption that g is symmetric. The first five conditions follow directly, as the verification contained in Linton does not rely on symmetry. The sixth condition, which we verify under asymmetry, is

$$E[-\frac{1}{2}T^{-\frac{1}{2}}\delta^T\Gamma_t(\gamma_0)\psi(u_t)|\mathcal{F}_{t-1}] = 0. \tag{.12}$$

Because u_t is an i.i.d. random variable, (.12) follows if

$$E\psi_1(u_t) + E\psi_2(u_t) = 0.$$

Integration by parts for $E\psi_1(u_t)$ reveals

$$E\psi_1(u_t) = \int \frac{g'(u)}{g(u)}g(u)du = \int g'(u)du = g(u)|_{-\infty}^{\infty} - \int g(u)\cdot 0 = 0,$$

where the last equality follows from the fact that absolute continuity of g implies $\lim_{u\to\infty} g(u) = \lim_{u\to-\infty} g(u) = 0$. Integration by parts for $E\psi_2(u_t)$ reveals

$$\begin{aligned} E\psi_2(u_t) &= \int \left[\frac{g'(u)}{g(u)}u + 1\right] g(u)du = \int ug'(u)du + \int g(u)du \\ &= ug(u)\ |_{-\infty}^{\infty} - \int g(u)du + \int g(u)du = 0, \end{aligned}$$

where the last equality follows from the fact that boundedness of Eu implies $\lim_{u\to\infty} ug(u) = \lim_{u\to-\infty} ug(u) = 0$. ∎

REMARK: Lemma 1 provides the key local regularity result needed to establish the asymptotic distribution of the parametric test statistics. Our tests are constructed from residuals; the significance of the contiguity property is that it enables us to proceed, in many respects, as if the true unobservable errors were used instead. This is of considerable help when working with the nonparametric estimates.

PROOF OF THEOREM 4. Let o_p(small) denote $o_p(\max\{h^2, T^{-1/2}h^{-3/2}\})$. Write the dependence of $\widehat{\tau}_c$ on $\widetilde{\beta}$ explicitly, and make the Taylor expansion

$$\widehat{\tau}_c(\widetilde{\beta}) = \widehat{\tau}_c(\beta_0) + \frac{\partial \widehat{\tau}_c}{\partial \beta}(\beta_0^*)(\widetilde{\beta} - \beta_0),$$

where β^* lies between β_0 and $\widetilde{\beta}$. The second term on the right hand side is $O_p(T^{-1/2})$, which can be verified by direct but lengthy calculation as in Linton (1995). Essentially, the parametric error caused by estimation of β_0 is of smaller order than the error due to estimating g. Therefore, $\widehat{\tau}_c(\widetilde{\beta})$ can be approximated by $\widehat{\tau}_c(\beta_0)$, which implies that the residuals are replaced by unobservable error terms

$$\widehat{\tau}_c^{**} \equiv \widehat{\tau}_c(\beta_0) = \frac{T^{-1/2}\sum_{i=1}^p \sum_{t=1}^T \widehat{\psi}_2(u_t) v_{t-i}^{**}}{\left\{T^{-1}\sum_{t=1}^T \widehat{\psi}_2^2(u_t)\right\}^{1/2} \left\{T^{-1}\sum_{i=1}^p \sum_{t=1}^T v_{t-i}^{**2}\right\}^{1/2}},$$

where $v_s^{**} = u_s^2 - T^{-1}\sum_{t=1}^T u_t^2$. We replace $v_s^{**} \Big/ \left\{T^{-1}\sum_{i=1}^p \sum_{t=1}^T v_{t-i}^{**2}\right\}^{1/2}$ with its asymptotic equivalent $\ddot{v}_s = (u_s^2 - m_2) \Big/ \{p(m_4 - m_2^2)\}^{1/2}$, where $m_j = E(u_t^j)$, arriving at

$$\ddot{\tau}_2 = \frac{T^{-1/2}\sum_{i=1}^p \sum_{t=1}^T \widehat{\psi}_2(u_t)\ddot{v}_{t-i}}{\left\{T^{-1}\sum_{t=1}^T \widehat{\psi}_2^2(u_t)\right\}^{1/2}} \equiv \frac{\widehat{\mathcal{Y}}}{\widehat{\mathcal{X}}^{1/2}}.$$

The approximation error (in replacing $\widehat{\tau}_c(\beta_0)$ by $\ddot{\tau}_c$) is of order $T^{-1/2}$. We make a two term Taylor expansion of $\ddot{\tau}_c$ about $I_2^{-1/2}\widehat{\mathcal{Y}}$, to give

$$\ddot{\tau}_c = I_2^{-1/2}\widehat{\mathcal{Y}} - \frac{1}{2}I_2^{-3/2}\widehat{\mathcal{Y}}(\widehat{\mathcal{X}} - I_2) + o_p(\widehat{\mathcal{X}} - I_2), \tag{.13}$$

where we show below that $\widehat{\mathcal{X}} - I_2 = O_p(h^2)$. Let $\mathcal{Y} = T^{-1/2}\sum_{i=1}^p \sum_{t=1}^T \psi_2(u_t)\ddot{v}_{t-i}$, and write $\widehat{\mathcal{Y}} = \mathcal{Y} + \widehat{\mathcal{Y}} - \mathcal{Y}$. That the leading term $I_2^{-1/2}\mathcal{Y}$ is asymptotically standard Gaussian is shown in Section 4.1. We replace $\widehat{\mathcal{Y}} - \mathcal{Y}$ and $\widehat{\mathcal{X}} - I_2$ by further approximations given in Lemmas 2 and 3 below.

Before that we introduce additional notation. We use subscript t to denote evaluation at u_t, e.g. $g_t = g(u_t)$, $\psi_{1t} = \psi_1(u_t)$, and $\widehat{g}_t = \widehat{g}(u_t)$, and E_t to denote expectation conditional on u_t. Let also $\overline{g}_t = E_t(\widehat{g}_t)$ and $\overline{g}_t^{(1)} = E_t(\widehat{g}_t^{(1)})$, and write $\widehat{g}_t - g_t = B_t + V_t$ and $\widehat{g}_t^{(1)} - g_t^{(1)} = B_t^{(1)} + V_t^{(1)}$, where $B_t = \overline{g}_t - g_t$ and $B_t^{(1)} = \overline{g}_t^{(1)} - g_t^{(1)}$, while $V_t = \widehat{g}_t - \overline{g}_t$ and $V_t^{(1)} = \widehat{g}_t^{(1)} - \overline{g}_t^{(1)}$. From Silverman (1986), the conditional bias of the kernel density estimator is $B_t = h^2 g_t^{(2)}\mu_2(K)/2 + h^4 g_t^{(4)}\mu_4(K)/4! + o(h^4)$ and the conditional bias of the kernel first derivative estimator is $B_t^{(1)} \approx h^2 g_t^{(3)}\mu_2(K)/2 + h^4 g_t^{(5)}\mu_4(K)/4! + o(h^4)$, while the "stochastic" terms are such that $V_t' = O_p(T^{-1/2}h^{-3/2})$ dominates $V_t = O_p(T^{-1/2}h^{-1/2})$. Note that the asymptotic bias of $\widehat{g}_t^{(1)}/\widehat{g}_t$ is to first order, $b_t = g_t^{-1}\{B_t^{(1)} + \psi_{1t}B_t\}$. In what follows we use:

$$\begin{aligned}\max|\widehat{g}_t - \overline{g}_t| &= O_p(T^{-1/2}h^{-1}), \max|\widehat{g}_t^{(1)} - \overline{g}_t^{(1)}| = O_p(T^{-1/2}h^{-2}),\\ \max|\overline{g}_t - g_t| &= O_p(h^2), \max|\overline{g}_t^{(1)} - g_t^{(1)}| = O_p(h^2),\end{aligned}$$

where max and min are both taken over $1 \le t \le T$. For proof of these results see Andrews (1995, Theorem 1) and Robinson (1987, Lemma 13). The additional conditions required for the proofs of second-order properties are mostly unverifiable smoothness and moment conditions which we shall assume hold. Our argument now parallels those presented in Linton (1995). We need the following two lemmas which are proved below.

LEMMA 2. *Assuming the moment exists, $E(\widehat{\mathcal{Y}}) = 0$ because of the independence of $\widehat{\psi}_2(u_t)$ and $\ddot{v}_{t-i}$. Also, by asymptotic expansion*

$$\widehat{\mathcal{Y}} = \mathcal{Y} - \mathcal{L}_1 - \mathcal{Q}_1 + o_p(\text{small}) \tag{.14}$$

where $\mathcal{Y} = T^{-1/2}\sum_{i=1}^p\sum_{t=1}^T \psi_{2t}\ddot{v}_{t-i} = O_p(1)$, $\mathcal{L}_1 = T^{-1/2}\sum_{i=1}^p\sum_{t=1}^T u_t\ddot{v}_{t-i}g_t^{-1}\{B_t^{(1)} + \psi_{1t}B_t\} = O_p(h^2)$, and $\mathcal{Q}_1 = T^{-1/2}\sum_{i=1}^p\sum_{t=1}^T u_t\ddot{v}_{t-i}g_t^{-1}V_t^{(1)} = O_p(T^{-1/2}h^{-3/2})$.

LEMMA 3. *By asymptotic expansion*

$$\widehat{\mathcal{X}} = I_2 + \mathcal{X} - 2\mathcal{B}_1 + o_p(\text{small}), \tag{.15}$$

where $\mathcal{X} = T^{-1}\sum_{t=1}^T(\psi_{2t}^2 - I_2) = O(T^{-1/2})$ and $\mathcal{B}_1 = E[\psi_{2t}u_t g_t^{-1}\{B_t^{(1)} + \psi_{1t}B_t\}] = O(h^2)$.

Substituting (.14) and (.15) into (.13) we arrive at the following approximation:

$$\ddot{\tau}_2 = I_2^{-1/2}\mathcal{Y} - I_2^{-1/2}\mathcal{Q}_1 - I_2^{-1/2}\left\{\mathcal{L}_1 - I_2^{-1}\mathcal{B}_1\mathcal{Y}\right\} + o_p(\text{small}), \tag{.16}$$

where the random variables $\mathcal{Y}$, $\mathcal{Q}_1$, and $\mathcal{L}_1$ satisfy: (1) $\text{var}(\mathcal{Y}) = I_2$; (2) $\text{cov}(\mathcal{Y},\mathcal{Q}_1) = 0$; (3) $\text{cov}(\mathcal{Y},\mathcal{L}_1) = E[\psi_{2t}u_t g_t^{-1}(B_t^{(1)} + \psi_{1t}B_t)] = \mathcal{B}_1$; (4) $\text{cov}(\mathcal{L}_1,\mathcal{Q}_1) = 0$; (5) $\text{var}(\mathcal{L}_1) = E[u_t^2 g_t^{-2}(B_t^{(1)} + \psi_{1t}B_t)^2] = \mathcal{B}_3$; (6) $\text{var}(\mathcal{Q}_1) = T^{-1}h^{-3}\nu_2(K^{(1)})E(u_t^2 g_t^{-1}) = \mathcal{B}_4$.

Proof of (1)-(5) is obvious; see below for a proof of (6). Substituting into (.16),

$$\text{cov}(\mathcal{Y}, \mathcal{L}_1 - I_2^{-1}\mathcal{Y}\mathcal{B}_1) = 0, \ \text{var}[\mathcal{L}_1 - I_2^{-1}\mathcal{Y}\mathcal{B}_1] = \mathcal{B}_3 - I_2^{-1}\mathcal{B}_1^2,$$

so that $\text{var}(C) = I_2^{-1}\mathcal{B}_4 + I_2^{-1}\{\mathcal{B}_3 - I_2^{-1}\mathcal{B}_1^2\} + o_p(\text{small})$, as required. ∎

PROOF OF LEMMAS. First we calculate the variance of $\mathcal{Q}_1$, which is a degenerate weighted U-statistic, see Fan and Li (1996). Write $V_t^{(1)} = \sum_{s \in \mathcal{T}(t)} \eta_{ts}$, where $\eta_{ts} = T^{-1}h^{-2}\left\{K_{ts}^{(1)} - E_t(K_{ts}^{(1)})\right\}$ with $K_{ts}^{(1)} = K^{(1)}\left(\frac{u_t - u_s}{h}\right)$, and $m_t = u_t g_t^{-1} \sum_{i=1}^{p} \ddot{v}_{t-i}$, so that

$$\mathcal{Q}_1 = T^{-1/2}\sum_{t=1}^{T} m_t V_t^{(1)} = T^{-1/2}\sum_{t=1}^{T}\sum_{s \in \mathcal{T}(t)} m_t \eta_{ts},$$

where m_t satisfies $E_{t-j}(m_t) = 0, \quad j = 0, 1, \ldots, p$, and m_t is independent of $V_t^{(1)}$. Therefore,

$$\text{var}(\mathcal{Q}_1) = T^{-1}\sum_{t=1}^{T} E(m_t^2)E(V_t^{(1)2}) + T^{-1}\sum_{t=1}^{T}\sum_{\substack{t=1 \\ r \neq t}}^{T} E(m_t m_r V_t^{(1)} V_r^{(1)}).$$

We claim that the double sum is of smaller order. Consider $t = r + p + k$, with $k \geq 1$. The typical terms in $E(m_{r+p+k} m_r V_{r+p+k}^{(1)} V_r^{(1)})$ are:

$$\begin{aligned}
E(m_{r+p+k} m_r \eta_{r+p+k,s} \eta_{rs}) &= E\left[E_{r,r+p+k,s}(m_{r+p+k} m_r \eta_{r+p+k,s} \eta_{rs})\right] = 0 \\
E(m_{r+p+k} m_r \eta_{r+p+k,r} \eta_{r,r+p+k}) &= E\left[E_{r,r+p+k}(m_{r+p+k} m_r \eta_{r+p+k,r} \eta_{r,r+p+k})\right] = 0 \\
E(m_{r+p+k} m_r \eta_{r+p+k,s} \eta_{rv}) &= 0.
\end{aligned}$$

Now consider $t = r + 1$. If $s = v$, then we must have $s > r + 1, s < r - p$, and

$$E(m_{r+1} m_r \eta_{r+1,s} \eta_{r,v}) = E\left[E_{s,r,r+1}(m_{r+1} m_r)\eta_{r+1,s}\eta_{r,s}\right] = 0,$$

because $E(\ddot{v}_{t-p}) = 0$.

PROOF OF LEMMA 2. Write

$$\widehat{\psi}_{2t} - \psi_{2t} = -u_t\left\{\frac{\widehat{g}_t^{(1)}}{\widehat{g}_t} - \frac{g_t^{(1)}}{g_t}\right\}$$

and make a geometric series expansion of $\widehat{g}_t^{(1)}/\widehat{g}_t$, see Härdle and Stoker (1989, pp992), to give

$$\widehat{\mathcal{Y}} - \mathcal{Y} = \sum_{j=1}^{r}\left\{T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\varphi_j(u_t)u_t\ddot{v}_{t-i}\right\} + \mathcal{R}_2 = \sum_{j=1}^{r} d_j + \mathcal{R}_2,$$

where

$$\varphi_j(u_t) = (-1)^j g_t^{-j} \left\{ (\widehat{g}_t - g_t)^{j-1}(\widehat{g}_t^{(1)} - g_t^{(1)}) + \psi_{1t}(\widehat{g}_t - g_t)^j \right\}, \quad j = 1, \ldots, r,$$

and $\mathcal{R}_2$ equals

$$(-1)^{r+1}T^{-1/2} \textstyle\sum_{i=1}^{p} \sum_{t=1}^{T} g_t^{-r}\widehat{g}_t^{-1} \left\{ (\widehat{g}_t - g_t)^r(\widehat{g}_t^{(1)} - g_t^{(1)}) + \psi_{1t}(\widehat{g}_t - g_t)^{r+1} \right\} u_t\ddot{v}_{t-i}.$$

Our proof has two parts: first the leading terms d_1 and d_2 and finally the remainder term $\mathcal{R}_2$.

(1) We first deal with the leading term d_1. Substituting for $\widehat{g}_t - g_t$ and $\widehat{g}_t^{(1)} - g_t^{(1)}$, we have

$$d_1 = -T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-1} \left\{ (B_t^{(1)} + V_t^{(1)}) + \psi_{1t}(B_t + V_t) \right\} = -\left(\mathcal{L}_1 + \mathcal{Q}_1 + \mathcal{Q}_2\right),$$

where

$$\begin{aligned}
\mathcal{L}_1 &= T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-1} \left\{ B_t^{(1)} + \psi_{1t}B_t \right\} = O_p(h^2) \\
\mathcal{Q}_1 &= T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t g_t^{-1} V_t^{(1)} \ddot{v}_{t-i} = O_p(T^{-1/2}h^{-3/2}) \\
\mathcal{Q}_2 &= T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t \psi_{1t} g_t^{-1} V_t \ddot{v}_{t-i} = O_p(T^{-1/2}h^{-1/2}).
\end{aligned}$$

(2) We next examine the term d_2. Substituting again for $\widehat{g}_t - g_t$ and $\widehat{g}_t^{(1)} - g_t^{(1)}$, we have

$$d_2 = T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-2} \left\{ (B_t^{(1)} + V_t^{(1)})(B_t + V_t) + \psi_{1t}(B_t + V_t)^2 \right\}.$$

Collecting terms we have

$$\begin{aligned}
d_2 \;=\;& T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-2} \left\{ B_t B_t^{(1)} + \psi_{1t}B_t^2 \right\} + \\
& T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}\psi_{1t}g_t^{-1}E_t(V_t^2) + T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-2}E_t(V_t^{(1)}V_t) + \\
& 2T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}\psi_{1t}g_t^{-1}B_t V_t + T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-2}B_t^{(1)}V_t + \\
& T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t\ddot{v}_{t-i}g_t^{-2}B_t V_t^{(1)} +
\end{aligned}$$

$$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}u_t\ddot{v}_{t-i}g_t^{-2}\left\{V_t^{(1)}V_t - E_t(V_t^{(1)}V_t)\right\} +$$
$$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}u_t\ddot{v}_{t-i}\psi_{1t}g_t^{-1}\left\{V_t^2 - E(V_t^2)\right\}.$$

The first line consists of single sums of order $O_p(h^4)$. The second line consists of single sums of orders in probability $O_p(T^{-1}h^{-1})$ and $O_p(T^{-1}h^{-2})$, respectively. The third line is a U-statistic of order 2 and is $O_p(h^2T^{-1/2}h^{-1/2})$. The fourth line is a U-statistic of order 2 and is $O_p(h^2T^{-1/2}h^{-3/2})$. Both these lines are uncorrelated with $\mathcal{Y}$. The fifth line is a U-statistic of order 3 and is $O_p(T^{-1}h^{-2})$, while the sixth line is a U-statistic of order 3 and is $O_p(T^{-1}h^{-1})$.

(3) To deal with the remainder term $\mathcal{R}_2$ we use crude bounds as in Robinson (1987). By the Cauchy-Schwarz inequality $|\mathcal{R}_2|$ is less than or equal to

$$\{(r+1)!\}^{-1}\{\min\widehat{g}_t\}^{-1}\{\min g_t\}^{-r}T^{1/2}\{\max|\widehat{g}_t - g_t|\}^r\left\{T^{-1}\textstyle\sum_{i=1}^{p}\sum_{t=1}^{T}u_t^2\ddot{v}_{t-i}^2\right\}^{1/2}$$
$$\times\left[\left\{T^{-1}\textstyle\sum_{t=1}^{T}\psi_{1t}^2\right\}^{1/2}\{\max|\widehat{g}_t - g_t|\} + \left\{\max|\widehat{g}_t^{(1)} - g_t^{(1)}|\right\}\right],$$

which is $o_p(T^{-2/7})$, provided $r \geq 2$. ∎

PROOF OF LEMMA 3. Write $\widehat{\psi}_{2t}^2 - \psi_{2t}^2 = 2\psi_{2t}(\widehat{\psi}_{2t} - \psi_{2t}) + (\widehat{\psi}_{2t} - \psi_{2t})^2$, and make the same geometric series expansion for $\widehat{\psi}_{2t} - \psi_{2t}$, to obtain

$$\widehat{\mathcal{X}} - \mathcal{X} = \sum_{j=1}^{r}\left\{T^{-1}\sum_{t=1}^{T}2\psi_{2t}\varphi_j(u_t)u_t\right\} + \left[\sum_{j=1}^{r}\left\{T^{-1}\sum_{t=1}^{T}\varphi_j(u_t)u_t\right\}\right]^2 + o_p(\text{small}),$$

where $\varphi_j(\cdot)$ is as defined above. The leading terms are

$$-2\psi_{2t}u_tg_t^{-1}\left\{(B_t^{(1)} + V_t^{(1)}) + \psi_{1t}(B_t + V_t)\right\}$$
$$+2\psi_{2t}u_tg_t^{-2}\left\{(B_t^{(1)} + V_t^{(1)})(B_t + V_t) + \psi_{1t}(B_t + V_t)^2\right\}$$
$$+u_t^2g_t^{-2}\left\{(B_t^{(1)} + V_t^{(1)}) + \psi_{1t}(B_t + V_t)\right\}^2.$$

We only need collect the bias terms, because the stochastic terms are $O_p(n^{-1/2})$ or smaller, and get multiplied by $\mathcal{Y}$ in (.13). We have

$$\begin{aligned}\widehat{\psi}_{2t}^2 - \psi_{2t}^2 \approx\ & -2\psi_{2t}u_tg_t^{-1}\{B_t^{(1)} + \psi_{1t}B_t\} + 2\psi_{2t}u_tg_t^{-2}\{B_t^{(1)}B_t + \psi_{1t}B_t^2\}\\ & +u_t^2g_t^{-2}\{B_t^{(1)} + \psi_{1t}B_t\}^2 + u_t^2g_t^{-2}E_t\{(V_t^{(1)})^2\}.\end{aligned}$$

All but the first term on the right hand side contributes $o_p(\text{small})$.

∎