COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1105

ADAPTIVE TESTING IN ARCH MODELS

Oliver Linton and Douglas G. Steigerwald

June 1995

ADAPTIVE TESTING IN ARCH MODELS

BY OLIVER B. LINTON

Department of Economics

Yale University

AND DOUGLAS G. STEIGERWALD

Department of Economics

University of California, Santa Barbara

SUMMARY

Existing specification tests for conditional heteroscedasticity are derived under the assumption that the density of the innovation, or standardized error, is Gaussian, despite the fact that many recent empirical studies provide evidence that this density is not Gaussian. We obtain specification tests for conditional heteroscedasticity 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. Monte carlo simulations indicate that asymptotic power gains are achievable in finite samples. We apply the tests to stock futures data sampled at high frequency and find evidence of conditional heteroscedasticity in the residuals from a GARCH(1,1) model, indicating that the standard (1,1) specification is not adequate.

Key words: Adaptive estimation; ARCH; Efficient; Semiparametric model; Testing.

JEL Classification Numbers: C12, C14, C22

1. Introduction

Volatility clustering is an important characteristic of financial time series and has major implications for estimation and testing in models of asset prices. To account for volatility clustering many researchers use some variant of the autoregressive conditional heteroskedasticity (ARCH) model developed by Engle (1982), see Bollerslev, Engle, and Nelson (1995) for a review. Successful application of these models requires correct specification of both the conditional mean and the conditional variance. Our interest here is in testing the specification of the conditional variance. The Gaussian likelihood based test statistics for such specifications, such as the LM, Wald, and LR, are asymptotically Chi-squared, see Engle (1984), and this remains true even when the error distribution is not Gaussian[1] as was pointed out in Bollerslev and Wooldridge (1992). However, they are suboptimal in terms of power except when the error is Gaussian. Recent empirical work questions the assumption that the innovation density is Gaussian[2].

We develop semiparametric specification tests of the conditional variance model. Our approach specifies the first two conditional moments parametrically but the innovation density is assumed only to be a member of a nonparametric family. Our work extends that of Linton (1993) and Steigerwald (1994) who study semiparametric estimators for a CH model. We show that the semiparametric test statistics are *adaptive* in the sense that they are asymptotically equivalent to the test statistics constructed from the true innovation density, i.e. the likelihood based test statistics, and hence maximize asymptotic local power[3].

Some of the specification tests we consider result in one-sided alternative hypothe-

[1] However, Newey and Steigerwald (1994) show that non-Gaussian likelihood based test statistics are not generally robust to misspecification of the innovation density.

[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] Previous work by Bera and Ng (1991) constructed test statistics based on nonparametric estimates of the score function, although their procedure is not, in general, optimal.

ses. We follow a proposal in Lee and King (1993) and construct test statistics that take advantage of this and are more powerful than two-sided test statistics. 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 against local alternatives. For a test of more than one additional parameter in the conditional variance function, we show that a semi-parametric 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. This may result in the small sample properties differing markedly from the predicted first-order asymptotic theory. We derive second-order asymptotic theory, as in Linton (1992,1994), to determine more accurately the small sample properties of our semiparametric test statistic. Specifically, we use second-order theory to determine the choice of smoothing parameter for the nonparametric density estimator. We also study a nonparametric density estimator (the multiplicative bias reduction estimator of Jones, Linton, and Nielsen (1993)) that offers reduced bias, and essentially zero bias for estimating the score function of the normal density, a canonical case here. This translates into improved performance for the semiparametric test statistic.

In Section 2 we define the hypotheses of interest and give the parametric and semi-parametric test statistics. In Section 3 we derive the first order asymptotic properties of these statistics and introduce our optimality criteria. Section 4 deals with the second order properties of the semiparametric test statistics. Section 5 contains monte carlo simulation results. Section 6 contains an application to exchange rate data.

A word on notation. Let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution and $\xrightarrow{\mathcal{P}}$ denote convergence in probability. For any function g, we use $g^{(i)}$ to denote the i'th derivative of g with respect to u, while for any vector $|x| = \left[\sum x_j^2\right]^{1/2}$ is the Euclidean norm. Finally, let $\mathbf{1}(\mathrm{A})$ be the indicator function of the set A.

2. Test Statistics

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. A common parameterization of a CH model is

$$(1) \qquad y_t = \mu_t(\gamma_0) + h_t(\gamma_0)\varepsilon_t,$$

where $\mu_t(\gamma)$ and $h_t(\gamma)$ are functions of current period information $\mathcal{F}_t = \{x_t, z_{t-1}, z_{t-2}, \ldots\}$ and a vector γ of parameters of interest, u_t is the period-t iid innovation independent of $\mathcal{F}_t$. Subscript 0 indicates true values of parameters. Now let α and σ be parameters for the location and scale of the density of ε_t. Usually, α and σ are the mean and standard deviation of ε_t respectively, and α_0 is assumed to be zero. For example, $\mu_t(\gamma) = \beta' x_t$, and

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

for an ARCH(p) model, while

$$h_t(\gamma) = [1 + \sum_{i=1}^{p} \phi_i (y_{t-i} - \beta' x_{t-i})^2 + \sum_{j=1}^{q} \rho_j h_{t-j}(\gamma)]^{\frac{1}{2}}$$

for a generalized ARCH model with order p and q (GARCH(p, q)). As these models illustrate, traditional CH parameterizations are in terms of relative scale. The parameters in ϕ are the relative scale parameters of the conditional variance, that is ϕ consists of ratios of each of the conditional variance slope parameters to the constant parameter in the conditional variance. If the conditional variance is $\delta_0 + \sum_{i=1}^{p} \delta_i (y_{t-i} - \beta' x_{t-i})^2$, then $\phi_i = \delta_i / \delta_0$. Linton (1993) shows that a semiparametric estimator of[4] $\gamma = (\beta', \phi')'$ is adaptive (i.e. is asymptotically equivalent to the maximum likelihood estimator) if the innovation density is symmetric.

Although our analysis can be applied to general CH models we restrict our attention to the important special case in which the conditional mean is correctly specified

[4] Linton sets $\sigma = e^{\frac{a}{2}}$.

to be $\mu_t(\gamma) = \beta' x_t$ and the conditional variance is ARCH(p). We derive asymptotically optimal tests for this model. In the more general GARCH(p, q) model the alternative space has a very complicated structure, see Nelson and Cao (1991), rendering it difficult to obtain asymptotically optimal tests. To obtain tests for a model in which the conditional variance is GARCH(p, q), we use the asymptotically optimal test statistics for an ARCH(p) alternative on the residuals from the estimated GARCH(p, q) model.[5]

If the Lebesgue density g of $u_t \equiv \sigma^{-1}\varepsilon_t$ were known, optimal inference about the parameters $\theta = (\sigma, \gamma')'$ would be 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 is now the residual $\sigma^{-1} h_t^{-1}(\gamma)[y_t - \mu_t(\gamma)]$, and $l_t(\theta, g)$ is the period-t conditional log-likelihood of y_t given $\mathcal{F}_t$. (The period-0 observation is considered fixed.) Define the standardized score $s_\theta(\theta, g)$, the outer product of the scores $\mathcal{I}^S_{\theta\theta}(\theta, g)$, and the observed information $\mathcal{I}^O_{\theta\theta}(\theta, g)$:

$$s_\theta(\theta, g) = T^{-1/2} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \quad ; \quad \mathcal{I}^S_{\theta\theta}(\theta, g) = T^{-1} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \quad ; \quad \mathcal{I}^O_{\theta\theta}(\theta, g) = T^{-1} \sum_{t=1}^{T} \frac{\partial^2 l_t}{\partial \theta \partial \theta'}.$$

The expected information matrix for θ, denoted $\mathcal{I}_{\theta\theta}(\theta, g)$, is defined as the probability limit of $\mathcal{I}^S_{\theta\theta}(\theta, g)$, $\mathcal{I}^O_{\theta\theta}(\theta, g)$, or indeed the standardized conditional outer product

[5] Newey and Steigerwald (1994) show that if the location of the innovation density is not zero, then quasi-maximum likelihood estimators of γ (that assume $\alpha = 0$ in (1)) are not generally consistent. Newey and Steigerwald recommend that we adopt the following parameterization

$$y_t = \mu_t(\gamma) + h_t(\gamma)(\alpha + \sigma u_t) \tag{3}$$

in which u_t is now a standardized random variable with zero location and unit scale. Because α is included in the conditional mean of y_t in (3), estimators of γ are consistent even if the location of the innovation density is not zero. In addition, Steigerwald (1994) shows that semiparametric estimators of γ in (3) satisfy a sufficient condition for adaptive estimation for general CH specifications even if the innovation density is asymmetric. The parameter α enters the conditional mean but *does not* enter the conditional variance. In an ARCH-M model, the coefficient on $h_t(\gamma_0)$ appears in both the conditional mean and the conditional variance, so α is separately identified from the coefficient on $h_t(\gamma_0)$ in an ARCH-M model.

$\mathcal{I}^C_{\theta\theta}(\theta,g) = T^{-1}\sum_{t=1}^T E[\frac{\partial l_t}{\partial\theta}\frac{\partial l_t}{\partial\theta'}|\mathcal{F}_{t-1}]$. It is instructive to examine the structure of these quantities. Define the Fisher scores for location and scale of the innovation density: $\psi_1(u) = -g(u)^{-1}\partial g(u)/\partial u$ and $\psi_2(u) = -[1+u\psi_1(u)]$, and let

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

and $v_{t-1}(\gamma) = (v_{1,t-1}(\gamma),..,v_{p,t-1}(\gamma))'$, with

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

Then

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

Thus Γ_t depends only on $\mathcal{F}_t$, while at the true parameter γ_0, $\psi(u_t)$ is mean zero and independent of $\mathcal{F}_t$. Note also that

$$\text{(4)} \qquad \text{vec}[\mathcal{I}^C_{\theta\theta}(\theta,g)] = T^{-1}\sum_{t=1}^T(\Gamma_t\otimes\Gamma_t)\text{vec}\begin{bmatrix} I_1(g) & I_{12}(g) \\ I_{12}(g) & I_2(g)\end{bmatrix} \equiv C_T(\theta)\text{vec}[I(g)],$$

where $I_1 = E[\psi_1^2(u_t)]$, $I_2 = E[\psi_2^2(u_t)]$ and $I_{12} = E[\psi_1(u_t)\psi_2(u_t)]$.

Consider the null hypothesis defined by the linear restrictions $H_0 : \; R\phi = r$, where R is a $q \times p$ matrix of full rank, r is a $q \times 1$ vector, and q is the number of restrictions. Standard likelihood based test statistics, such as Lagrange multiplier, Wald, and likelihood ratio, can be used to test H_0 against the general alternative $R\phi \neq r$. Under the null hypothesis all three test statistics are asymptotically $\chi^2(q)$, see Engle (1984). Furthermore, they maximize asymptotic power against local alternatives. However, because the conditional variance can't be negative, specification tests about this quantity more naturally have a one-sided alternative $H_A : R\phi > r$. Lee and King (1993) note that the LM, LR and Wald tests ignore the one-sided nature of the alternative (and cannot be modified simply to take account of this, when $q > 1$) and consequently suffer a loss of asymptotic power[6].

We examine two types of specification test. First, we study a test for one additional parameter in the conditional variance of an ARCH model, i.e. that the model is ARCH($p-1$) against the alternative hypothesis that the model is ARCH(p). This hypothesis, denoted T1, is

$$\text{T1:} \; H_0 : \phi_p = 0 \quad \text{vs.} \quad H_A : \phi_p > 0.$$

Our second test is for more than one additional parameter in the conditional variance of an ARCH model: i.e. that the model is ARCH($p-q$), with $q > 1$, against the alternative hypothesis that the model is ARCH(p). This hypothesis, denoted T2, is

$$\text{T2:} \; H_0 : \phi_{p-q}, .., \phi_p = 0 \quad \text{vs.} \quad H_A : \phi_{p-q}, .., \phi_{p-1} \geq 0, \phi_p > 0.$$

[6] In fact, if $\tilde{\theta}$ is constructed imposing the inequality restrictions, the LR test is not asymptotically $\chi^2(q)$, see Gourieroux et al (1982).

2.1 Parametric Test Statistics

Let $\widetilde{\theta}$ be the maximum likelihood estimator (MLE) of θ imposing the null restrictions, or a one-step Newton-Raphson approximation to it. Then define for any p by 1 vector c,

$$\tau_c = \left\{c'\widehat{\mathcal{I}}^{\phi\phi}(\widetilde{\theta}, g)c\right\}^{1/2} c' s_\phi(\widetilde{\theta}, g), \tag{5}$$

where $\widehat{\mathcal{I}}_{\theta\theta}(\theta, g)$ is any consistent estimate of $\mathcal{I}_{\theta\theta}(\theta, g)$ and $\widehat{\mathcal{I}}^{\phi\phi}$ is the corresponding element of the inverse of $\widehat{\mathcal{I}}_{\theta\theta}$. By taking $c = (0, .., 0, 1)'$, we obtain a test τ_1 of T1, while when $c = (1, .., 1)'$, we get a sum of ϕ-scores test[7] of T2, denoted τ_m. The Gaussian versions of these tests are particularly simple and have been extensively studied, see especially Lee and King (1993), Bera and Higgins (1993), Engle (1984), and Bollerslev and Wooldridge (1992). In this case, the information matrix is block diagonal between β and (σ, ϕ), while $\psi_1(u) = -u$ and $\psi_2(u) = -[u^2 - 1]$.

In the next section we show that τ_c is asymptotically standard Gaussian.

[7] As pointed out in Lee and King (1993), the LM statistic for testing a null of homoscedasticity against ARCH(p) or against GARCH(p, q) is the same because the score for the subset of the conditional variance parameters ρ equals zero under the null hypothesis of homoscedasticity.

2.2 Semiparametric Test Statistics

We now introduce semiparametric versions of our tests. We first rewrite τ_1 and τ_m using the efficient scores for ϕ relative to σ and β inside the parametric model, see Bickel et al. (1994) for discussion of the notion of efficient score and its application. This is the residual from a projection of the score for ϕ onto the scores for $\chi = (\sigma, \beta')'$:

$$\begin{aligned} \frac{\partial l_t^*}{\partial \phi} &= \frac{\partial l_t}{\partial \phi} - \mathcal{I}_{\phi\chi}\mathcal{I}_{\chi\chi}^{-1}\frac{\partial l_t}{\partial \chi} \\ &= A_t\psi_1(u_t) + B_t\psi_2(u_t), \end{aligned}$$

where A_t and B_t depend only on the past[8]. 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 these parameters, see Linton (1993) and Steigerwald (1994). Let $s_\phi^*(\theta, g) = T^{-1/2}\sum_{t=1}^T \frac{\partial l_t^*}{\partial \phi}$ and $\mathcal{I}_{\phi\phi}^*(\theta, g) = \operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^T \frac{\partial l_t^*}{\partial \phi}\frac{\partial l_t^*}{\partial \phi'}$. We first rewrite the parametric test statistics, using the efficient scores, as[9]

(6) $$\tau_c = \left\{c'\widehat{\mathcal{I}}_{\phi\phi}^*(\widetilde{\theta}, g)c\right\}^{-1/2} c's_\phi^*(\widetilde{\theta}, g).$$

We use (6) to construct a semiparametric test statistic. We replace population moments by their sample equivalents and g by a nonparametric kernel density estimator

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

[8] In the special case that g is symmetric about zero, $\mathcal{I}_{\phi\beta} = 0$, and the efficient score for ϕ has the especially simple form $\frac{\partial l_t^*}{\partial \phi} = \frac{1}{2}[v_{t-1} - E(v_{t-1})]\psi_2(u_t)$.

[9] Use of the efficient score to construct the test statistic is equivalent to assuming that $a = 0$ in the Linton parameterization. We do not estimate location and scale because these parameters are not jointly identified with the innovation density g.

where $\widetilde{u}_s$ are standardized residuals from a preliminary $T^{1/2}$ consistent procedure, while $K(\bullet)$ is a density function symmetric about zero and $h(T)$ is a bandwidth parameter. The index set $\mathcal{T}_t$ is taken here to be $\{s \neq t\}$, although in section 4 below we employ a more severe sample splitting[10]. We estimate ψ_1 and ψ_2 by $\widehat{\psi}_1(u) = -\widehat{g}^{-1}\partial\widehat{g}/\partial u(u)$ and $\widehat{\psi}_2(u) = -[u\widehat{\psi}_1(u) + 1]$, the efficient score by

$$s^*_\phi(\theta, \widehat{g}) = T^{-\frac{1}{2}} \sum_{t=1}^{T} [A_t(\theta)\widehat{\psi}_1(\widetilde{u}_t) + B_t(\theta)\widehat{\psi}_2(\widetilde{u}_t)],$$

and the efficient information $\mathcal{I}^*_{\phi\phi}(\theta, g)$ by $\widehat{\mathcal{I}}^*_{\phi\phi}(\theta, \widehat{g}) = \mathcal{I}^{*C}_{\phi\phi}(\theta, \widehat{g})$ in which $I(\widehat{g}) = T^{-1}\sum_{t=1}^{T} \widehat{\psi}(\widetilde{u}_t)\widehat{\psi}(\widetilde{u}_t$ Now let $\widehat{\theta}$ be the semiparametric estimator of θ defined in Linton (1993), i.e. $\widehat{\theta} = \widetilde{\theta} - \widehat{\mathcal{I}}^{-1}_{\theta\theta}(\widetilde{\theta}, \widehat{g})s_\theta(\widetilde{\theta}, \widehat{g})$, where $\widetilde{\theta}$ is the Gaussian QMLE. Finally, we introduce the following semiparametric test statistics

$$\widehat{\tau}_c = \left\{c'\widehat{\mathcal{I}}^*_{\phi\phi}(\widehat{\theta}, \widehat{g})c\right\}^{-1/2} c's^*_\phi(\widehat{\theta}, \widehat{g}). \tag{8}$$

Note that having obtained $\widehat{\theta}$, we could reestimate g using the semiparametric residuals $\widehat{u}_t$. However, this iteration does not affect either the first order or second order properties of the procedure.

[10] In section 4 below we discuss refinements of the kernel estimator that can be important in practice. Of course, we may prefer, for finite sample reasons, to divide by $(T-1)$ rather than T, but this makes no difference to the asymptotic arguments. Finally, it is common practice to include trimming functions in at various places in the definition of the semiparametric quantities. We don't need to do this for theoretical reasons because we assume a bounded support for g. We also eschew trimming for practical reasons, see below for more discussion.

3. First Order Asymptotic properties

3.1 Parametric Model

To derive asymptotic properties of test statistics, we first establish that a parametric ARCH model is regular in the sense that its likelihood ratio has the local asymptotically normal (LAN) approximation. Let $\theta_T = \theta_0 + \delta T^{-1/2}$ for any $\delta \in \mathbb{R}^{K+P+1}$, and let P_{T,θ_0} and P_{T,θ_T} be the probability measures of the data associated with the respective parameters. Convergences below are under P_{T,θ_0} unless otherwise stated. We make the following assumptions:

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

A2. *The density g has positive finite Fisher information for both location and scale parameters*

$$0 < \int \psi_1^2(u)g(u)du, \int \psi_2^2(u)g(u)du < \infty.$$

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

A4. *The moments $\int u^4 g(u)du$ and $\int \psi_j^4(u)g(u)du$, $j = 1, 2$, are finite.*

A5. *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 $\mathcal{I}_{\theta\theta}(\theta, g)$ is nonsingular at θ_0.*

A6. *The initial condition density $g_0(Y_0; \theta)$, where $Y_0 = (y_0, y_{-1}, .., y_{-p})$, is continuous in probability: i.e. $g_0(Y_0; \theta_T) \xrightarrow{\mathcal{P}} g_0(Y_0; \theta)$, for any $\theta_T \rightarrow \theta$.*

A7. *The regressors $\{x_t\}_{t=1}^{T}$ are weakly exogenous for θ. Furthermore, $T^{-1} \sum_{t=1}^{T} x_t x_t' \xrightarrow{\mathcal{P}} M > 0$.*

THEOREM 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-A7 are satisfied. Then*

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

and $s_\theta(\theta_0, g) \xrightarrow{\mathcal{D}} N(0, \mathcal{I}_{\theta\theta}(\theta_0, g))$, *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.*

A detailed proof of this result is given in Linton (1993) which uses some results of Swensen (1985). However, in that paper symmetry was assumed; this is not necessary.∎

REMARK: A sufficient condition for A5(c) is that h_t have bounded second moment, see Weiss (1986). However, Lumsdaine (1991) 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 (1987) allows for integrated regressors and derives the more general result of Local Asymptotic Mixed Normality in this case.

Theorem 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. For estimation we need some additional smoothness properties. We use a simple but overly strong assumption:

A8. *Both* $\int[\psi_1'(u)]^2 g(u)du$ *and* $\int[\psi_2'(u)]^2 g(u)du$ *are finite.*

THEOREM 2. *Suppose that A1-8 hold. Then,*

$$\tau_c \xrightarrow{\mathcal{D}} N(0,1), \qquad \textit{under } \theta_0. \tag{9}$$

PROOF. Follows from Linton (1993, Theorem 3). ■

An immediate consequence of Theorem 1 is "LeCam's Third Lemma" (see Bickel et al., (1994, p. 503)). This delivers to us the asymptotic distribution of scalar test statistics τ_T under a sequence of local alternatives: if

$$(\tau_T, \Lambda_T) \xrightarrow{\mathcal{D}} (\tau, \Lambda), \qquad \text{under } \theta_0,$$

where (τ, Λ) is a bivariate Gaussian random variable with mean $(\mu, \frac{-\sigma^2}{2})$ and covariance matrix $\begin{bmatrix} \eta^2 & \omega \\ \omega & \sigma^2 \end{bmatrix}$, then

$$(\tau_T, \Lambda_T) \xrightarrow{\mathcal{D}} (\tau + \omega, \Lambda + \sigma^2), \qquad \text{under } \theta_T.$$

This result can be used to derive the local power function for τ_c: combining Theorem 1 and 2 (the convergence there is also joint with Λ_T), we have

COROLLARY:

$$\tau_c \xrightarrow{\mathcal{D}} N(\{c'\mathcal{I}^{\phi\phi}(\theta_0, g)c\}^{1/2}\delta'\mathcal{I}_{\phi\phi}(\theta_0, g)c, 1), \qquad \text{under } \theta_T. \tag{10}$$

3.2 Semiparametric Statistics

We now extend Theorem 2 to the semiparametric test statistics. We make the additional assumption

A9. *The set $\mathcal{H}$ is compact, i.e. g has bounded support,*

and find

THEOREM 3. *Suppose that A1-A9 hold. Suppose also that the kernel K has bounded support and is twice continuously differentiable, and that the bandwidth sequence satisfies $h \to 0$ and $Th^4 \to \infty$. Then,*

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

PROOF: It is sufficient to establish that $s^*_\phi(\theta_T, \widehat{g}) - s^*_\phi(\theta_T, g) = o_p(1)$ and $\widehat{\mathcal{I}}^*_{\phi\phi}(\theta_T, \widehat{g}) - \widehat{\mathcal{I}}^*_{\phi\phi}(\theta_T, g) = o_p(1)$ which is established in Linton (1993). ∎

We next define our optimality criteria and show that τ_1, $\widehat{\tau}_1$, τ_m and $\widehat{\tau}_m$ are asymptotically optimal.

3.3. Optimality of Tests

Define for any test statistic τ, its critical function

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

where κ_α, with $\alpha \in (0,1)$, is a critical value, in our case determined by (9). Let $E_{T,\delta}$ denote expectation taken with respect to the measure P_{T,θ_T} of the sequence of local alternatives. Let H_0 and H_A be the δ consistent with the null and alternative hypotheses respectively.

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

$$\limsup_{T\to\infty} E_{T,\delta}\varphi_\alpha(\tau) \le \alpha, \text{ for all } \delta \in H_0,$$

while

$$\liminf_{T\to\infty} E_{T,\delta}\varphi_\alpha(\tau) \ge \alpha, \text{ for all } \delta \in H_A.$$

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

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

for any $\epsilon > 0$.

For a specification test of T1 we have

THEOREM 4. *The test statistics τ_1 and $\hat{\tau}_1$ 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 3 implies that local power is maximized by τ_1, at least when superefficient test are not allowed in the comparison.

Theorem 4 does not apply to the test of T2, 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[10].

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 $\delta_0 > 0$ such that

$$\int_{\{|\theta-\theta_0|<\delta\}\cap H_A} \pi_\tau(\theta) d\theta > \int_{\{|\theta-\theta_0|<\delta\}\cap H_A} \pi_{\tau^*}(\theta) d\theta, \qquad \forall \delta < \delta_0.$$

[10] 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, ω. Smaller values of c give higher weights to alternatives that are close to the null value and lower weights to alternatives that are distant from the null value. As c increases, the weight is transferred from alternatives that are close to the null to distant alternatives. Andrews shows that a test statistic that maximizes his criterion is a transformation of the classical solution that reflects the distance of the alternative space from the parameter estimates.

This corresponds to a locally best (i.e. maximin) in the direction $\phi_1 = .. = \phi_p$. Lee and King (1993) show that the LMMPU test for (Gaussian) T2 is based on the sum of scores, which justifies our choice (3.2). We have

THEOREM 5. *The test statistics τ_m and $\widehat{\tau}_m$ are asymptotically LMMPU.*

PROOF: This follows directly from the definition of an LMMPU test given by King and Wu (1991), and Theorems 2 and 3. ■

4. SECOND-ORDER ASYMPTOTIC PROPERTIES

4.1 Size Distortion

We now develop more refined asymptotic approximations that help us determine more accurately the small sample properties of the semiparametric test statistic. Specifically, we derive an asymptotic expansion for $\widehat{\tau}_m$ including terms to second order. By second-order, we mean terms that are asymptotically negligible with respect to $\widehat{\tau}_m - \tau_m$ and thus do not show up in the limiting distribution. In this semiparametric context, the correction terms are of larger order than $T^{-1/2}$ in probability (which is the usual magnitude for the correction factors in parametric problems). See Linton (1994) for further discussion of second order theory for semiparametric models. We calculate the first two moments of the truncated expansion and use these to better approximate the properties of the test statistic.

Many additional conditions are required for the proofs of second order properties: these are mostly unverifiable smoothness and moment conditions. We do not give a full set of primitive conditions, but indicate in the proofs what is needed. We shall restrict our attention to the case where g is known to be symmetric about zero. In this case,

$$\widehat{\tau}_m = \frac{T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\widehat{\psi}_2(\widetilde{u}_t)v_{t-i}^{*}}{[T^{-1}\sum_{t=1}^{T}\widehat{\psi}_2^2(\widetilde{u}_t)]^{1/2}[T^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}v_{t-i}^{*2}]^{1/2}}, \quad (11)$$

where $v_t^* = \tilde{u}_t^2 - T^{-1}\sum_{t=1}^T \tilde{u}_t^2$. For convenience, we use a leave-p-out kernel density estimator, i.e. the index set in (7) is $\mathcal{T}_t = \{s \neq t, .., t-p\}$. This simplifies and improves the second order approximations.

Write $\hat{\tau}_m(\hat{\theta})$, and make the Taylor expansion

$$\hat{\tau}_m(\hat{\theta}) = \hat{\tau}_m(\theta) + \frac{\partial \hat{\tau}_m}{\partial \theta}(\theta)(\hat{\theta} - \theta) + \frac{1}{2}(\hat{\theta} - \theta)' \frac{\partial^2 \hat{\tau}_m}{\partial \theta \partial \theta'}(\theta^*)(\hat{\theta} - \theta),$$

where θ^* lies between θ and $\hat{\theta}$. The second and third terms are $O_p(T^{-1/2})$ and $O_p(T^{-1})$ respectively, by similar calculations to those carried out below. For our purposes, $\hat{\tau}_m$ can be approximated by the first term which has residuals substituted by unobservable error terms,

$$(12) \qquad \hat{\tau}_m^* \equiv \hat{\tau}_m(\theta) = \frac{T^{-1/2}\sum_{i=1}^p \sum_{t=1}^T \hat{\psi}_2(u_t) v_{t-i}^{**}}{[T^{-1}\sum_{t=1}^T \hat{\psi}_2^2(u_t)]^{1/2}[T^{-1}\sum_{i=1}^p \sum_{t=1}^T v_{t-i}^{**2}]^{1/2}} \equiv \hat{\mathcal{X}}^{-1/2}\hat{\mathcal{Y}},$$

where $v_s^{**} = u_s^2 - T^{-1}\sum_{t=1}^T u_t^2$. The approximation error (in replacing $\hat{\tau}_m$ by $\hat{\tau}_m^*$) is of order $T^{-1/2}$ and does not contribute to the second order mean squared error properties of the semiparametric test statistic. See Linton (1992) for a similar result. Let $\mathcal{X} = pI_2(g)(m_4 - m_2^2)$ and $\mathcal{Y} = T^{-1/2}\sum_{i=1}^p \sum_{t=1}^T \psi_2(u_t) v_{t-i}^0$, with $v_s^0 = u_s^2 - m_2$, where $m_j = E[u_t^j]$, $j = 2, 4$. A first-order Taylor-series expansion of $\hat{\tau}_m^*$ about $\mathcal{X}^{-1/2}\mathcal{Y}$ yields

$$(13) \qquad \hat{\tau}_m^* \approx \mathcal{X}^{-1/2}\mathcal{Y} + \mathcal{X}^{-1/2}(\hat{\mathcal{Y}} - \mathcal{Y}) - \frac{1}{2}\mathcal{X}^{-3/2}\mathcal{Y}(\hat{\mathcal{X}} - \mathcal{X}),$$

where the symbol $\approx$ indicates that the remainder term can be ignored. The asymptotic distribution of the leading term, $\mathcal{X}^{-1/2}\mathcal{Y}$, was studied in section 3.1. The remaining two terms affect the variance of the test statistic to a second-order of magnitude. Let $\mu_2(K) = \int s^2 K(s) ds$, let $\nu_2(K) = \int K^2(s) ds$, and let $b_t = \frac{\mu_2}{2} u_t [\frac{g^{(3)}}{g}(u_t) - \frac{g^{(2)}}{g}\frac{g^{(1)}}{g}(u_t)]$.

THEOREM 6. *Suppose that sufficient smoothness and moment conditions hold, in addition to the conditions of the previous theorem. Let $h = O(T^{-1/7})$. Then*

$$\text{avar}[\widehat{\tau}_m] = 1 + h^4\left[\frac{E[b_t^2]}{I_2} - \left\{\frac{E[\psi_2(u_t)b_t]}{I_2}\right\}^2\right] + T^{-1}h^{-3}\nu_2(K')\frac{E[u_t^2 g^{-1}(u_t)]}{I_2} + o(T^{-4/7}),$$

as $T \to \infty$. Furthermore, both the asymptotic mean and skewness of $\widehat{\tau}_m$ are of order $T^{-1/2}$.

PROOF: See appendix. ∎

REMARK. The variance correction factors contain a term due to the bias b_t of the nonparametric score function estimate, and one due to its variance. The magnitudes of these terms depend on the underlying density and can be quite large. The magnitude of p does not affect the second order term, which is somewhat surprising. However, it does affect the $O(T^{-1})$ term in the variance, which we have not presented.

This result can be used to select the bandwidth. There are many possible optimality criteria. The approach we take is to choose h to minimize the expression for the variance of $\widehat{\tau}_m$ given above. This is equivalent to minimizing the second-order size distortion. This approach apparently neglects power, although see the simulation results below. The optimal bandwidth depends on the terms of the variance correction factors, which are unknown. One possibility is to substitute nonparametric estimates of these quantities into the bandwidth formula, see Härdle and Linton (1995) for a discussion of this "plug-in" method. We advocate using a "rule of thumb" (based on a normal "pilot") version of this bandwidth selection method for applications. 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 $b(u) = u^2\mu_2$, so

$E[b(u_t)^2] = 3\mu_2^2$ and $E[\psi_2(u_t)b(u_t)] = 2\mu_2$, and $\frac{E[b_t^2]}{I_2} - \left\{\frac{E[\psi_2(u_t)b_t]}{I_2}\right\}^2 = \frac{1}{2}\mu_2^2$. Note that $E[u_t^2 g^{-1}(u_t)] = \infty$, if g is Gaussian. We replace $E[u_t^2 g^{-1}(u_t)]$ by a consistent estimate[11] $(u_{\max}^3 - u_{\min}^3)/3$. Finally, our rule-of-thumb bandwidth is

$$\widehat{h} = \left\{\frac{\nu_2(K^{(1)})(u_{\max}^3 - u_{\min}^3)}{4\mu_2^2(K)}\right\}^{1/7} \widehat{\sigma} T^{-1/7}, \tag{14}$$

where $\widehat{\sigma}$ is an estimator of the standard deviation of the raw data and $u_{\max}$ and $u_{\min}$ are calculated from the data standardized by $\widehat{\sigma}$. This estimate asymptotically minimizes the second-order size distortion as defined above.

[11] This quantity grows at rate $(\ln T)^3$, when g has unbounded support. We recognize that in this case, theorem 6 is meaningless since the right hand side is infinite. Nevertheless, we believe that our formula makes practical sense.

4.2 Bias Reduction

The bias term b_t is nonzero for the normal density, and hence our procedure will have a bias-related variance correction term even in this canonical case, see Theorem 5. We study here a refined kernel density estimator proposed by Jones, Linton, and Nielsen (1993), hereafter JLN. The JLN estimator reduces the bias[12] of the density estimate to $O(h^4)$, from $O(h^2)$, for all g. More importantly for our purposes, it results in an essentially unbiased estimator of the score function when in fact the true density is normal. This allows a much wider bandwidth to be used and translates into gains in the second order performance of our test statistics.

Let $\widehat{g}$ be a kernel density estimator, then the JLN density estimator based on observed data $\{u_t\}_{t=1}^T$ is

$$\widetilde{g}(u) = \widehat{g}(u)\ (Th)^{-1} \sum_{t=1}^{T} \widehat{g}^{-1}(u_t) K(\frac{u - u_t}{h}). \tag{15}$$

Again, $\widetilde{g}^{(1)}(u) = \frac{d}{du}\widetilde{g}(u)$ and $\tilde{\psi}_1 = \widetilde{g}^{(1)}/\widetilde{g}$. Under the same assumptions necessary for the improved mean squared error (MSE) performance of fourth order kernels (i.e. kernels L such that $\int s^l L(s)ds = 0$, $l = 1,2,3$ and $\int L(s)ds$ and $\int s^4 L(s)ds$ are nonzero), Jones, Linton and Nielsen (1993) obtain the following asymptotic expansion for $\widetilde{g}(u)$, when $h = O(n^{-1/9})$:

$$\widetilde{g}(u) - g(u) \approx h^4 b(u) + (Th)^{-1/2} V(u) + o_p(T^{-4/9}),$$

where $b(u) = \frac{1}{4}\mu_2^2(K) g(u) \left\{\frac{g^{(2)}}{g}(u)\right\}^{(2)}$, while $V(u) = (Th)^{-1/2} \sum_{t=1}^T \left\{L(\frac{u-u_t}{h}) - E[L(\frac{u-u_t}{h})]\right\}$, with[13] $L = K * K - 2K$, is a zero mean sum of iids and is $O_p(1)$. Therefore, $\widetilde{g}(u)$ has bias an order of magnitude smaller than the usual kernel estimator. The same is true of $\widetilde{g}^{(1)}(u)$ and hence of $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$, i.e. their biases are also $O(h^4)$ com-

[12] It also guarantees a positive estimate of g everywhere unlike other bias reduction methods such as higher-order kernel density estimators.

[13] L is the fourth order kernel obtained from twicing, see Jones and Foster (1995).

pared with $O(h^2)$ for $\widehat{g}^{(1)}(u)$ and $\widehat{g}^{(1)}(u)/\widehat{g}(u)$. The reduced bias permits faster convergence rate of $n^{-8/9}$, when $h = O(n^{-1/9})$, as compared with rate $n^{-4/5}$ for the standard kernel density estimator with $h = O(n^{-1/5})$. The bias constant for $\widetilde{g}^{(1)}(u)$ is $b^{(1)}(u) = \frac{1}{4}\mu_2^2 g^{(1)}(u)\left\{\frac{g^{(3)}}{g^{(1)}}(u)\right\}^{(2)}$. Therefore, for $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$ it is $\frac{b^{(1)}}{g}(u) - \frac{g^{(1)}}{g}\frac{b}{g}(u)$. If g is the Gaussian density, $b(u) = \frac{1}{2}\mu_2^2\phi(u)$ and $b^{(1)}(u) = -\frac{1}{2}\mu_2^2 u\phi(u)$, so, remarkably, there is a cancellation and the bias of $\widetilde{g}^{(1)}(u)/\widetilde{g}(u)$ is the even better $o(h^4)$. This means that in the formula of Theorem 6, the first term on the left hand side is zero for this procedure[14]. Furthermore, we might expect some improvement in local power too. This result may be of considerable practical significance, since most empirical densities do not seem very far from the normal.

[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.

5. Finite Sample Performance

The results in Section 3 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 monte carlo simulations for samples of 100 observations. Because most financial data sets have substantially larger numbers of observations, our results provide conservative estimates of the gains achievable in practice. We find that with a sample of only 100 observations, the semiparametric test statistic $\hat{\tau}_1$ can outperform the QML test statistic τ_1.

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

***TABLE 1 HERE ***

We simulate the model (3) for an ARCH(p) specification with $f_t(\gamma) = \beta_0 + \beta_1 x_{1t}$. 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 $\sigma h_t(\gamma) u_t$. Note that for the asymmetric densities, a Gaussian QMLE is consistent only if α is included. We perform 1000 monte carlo 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

$$(16) \qquad \hat{\gamma}^i_{TQM} = \hat{\gamma}^{i-1}_{TQM} + \lambda \mathcal{I}_T^{-1}(\hat{\gamma}^{i-1}_{TQM}, g^n) s^*_\gamma(\hat{\gamma}^{i-1}_{TQM}, g^n),$$

where g^n is a Gaussian density and λ is a parameter[15] that controls the size of the updating step. The algorithm (16) is iterated until $s^*_\gamma(\hat{\gamma}^{i-1}_{TQM}, g^n)' \mathcal{I}_T^{-1}(\hat{\gamma}^{i-1}_{TQM}, g^n) s^*_\gamma(\hat{\gamma}^{i-1}_{TQM}, g^n)$ is less than .01.

The semiparametric estimator is constructed as in (16) with $\hat{g}$ used in place of a Gaussian density, where $\hat{g}$ is constructed using the residuals calculated from $\hat{\gamma}^{i-1}_T$ and $\hat{\gamma}^0_T = \hat{\gamma}^0_{TQM}$. 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 (7), with a JLN reduced-bias estimator of g, given by (15). Second, we compare the value of the smoothing parameter that minimizes second-order size distortion, given by $\hat{h}$ in (14), 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

[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.

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

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 $h_t(\gamma_0)^2 = 1 + .1(y_{t-1} - 1 + x_{1t-1})^2$ and the ARCH(2) specification is $h_t(\gamma_0)^2 = 1 + .1(y_{t-1} - 1 + x_{1t-1})^2 + .5(y_{t-2} - 1 + x_{1t-2})^2$. 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 (7) with the value of the smoothing parameter given by $\hat{h}$ in (14). The second semiparametric test statistic, denoted SP2, is constructed using the JLN estimator of g from (15) with the value of the smoothing parameter given by $\hat{h}$ in (14). The third semiparametric test statistic, denoted SP3, is constructed using the JLN estimator of g from (15) with the value of the smoothing parameter[17] given by $1.5 \cdot \hat{h}$.

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

this estimator.

[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.

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.

*** TABLE 2 HERE ***

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. For the asymmetric and platykurtic densities, the relative performance of the semiparametric test statistic improves as the departure from normality grows. If the innovation density is Gaussian, a QML test statistic correctly rejects the null hypothesis more than three-quarters of the time and the SP3 test statistic suffers a power loss of 8 percent. For the nearly Gaussian densities (A1,L1,P1), the results are similar. For the remaining leptokurtic densities, a semiparametric test statistic has a power loss of 10 percent while for the remaining platykurtic densities a semiparametric test statistic has only a slight power loss. For the remaining asymmetric densities, a semiparametric test statistic has a power gain of between 2 and 27 percent. With a sample of only 100 observations a semiparametric test statistic outperforms a QML test statistic for several densities.

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. Table 3 contains the empirical size and size-adjusted power for the QML and SP3 test statistics for a sample of 500 observations. 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 $h_t(\gamma_0)^2 = 1 + .1(y_{t-1} - 1 + x_{1t-1})^2 + .22(y_{t-2} - 1 + x_{1t-2})^2$. The second and third columns contain the empirical size for the test statistics. The fourth and fifth columns contain the size-adjusted power for the QML and SP3 test statistics, respectively. 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 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.

*** TABLE 3 HERE ***

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 (6) and (8), respectively, 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 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. If the innovation density is Gaussian, the QML test statistic correctly rejects the null hypothesis slightly less than half of the time and a semiparametric test statistic correctly rejects slightly more than half of the time. For the nearly Gaussian densities, a QML test statistic has power gains of between 7 and 9 percent. For the remaining asymmetric and leptokurtic densities the power gains are similar. For the most extreme platykurtic density, the larger smoothing parameter results in a power loss of 13 percent. With a sample of only 100 observations a semiparametric test statistic outperforms a QML test statistic for nine of the ten densities.

*** TABLE 4 HERE ***

6. Analysis of Stock Futures Data

We analyze high frequency intraday returns from stock futures with both a semiparametric and a QML test statistic. Empirical studies of intraday data underlies much of the recent modeling of market microstructures. Generally, a GARCH(1,1) specification is used to model conditional heteroscedasticity, often without formal testing of the adequacy of the specification. We estimate a GARCH(1,1) model for high frequency data and examine the residuals for evidence of additional heteroscedasticity. We find that both a QML and a semiparametric test statistic indicate the presence of higher order conditional heteroscedasticity.

Our data consist of five minute returns for the Standard and Poor's (S&P) 500 composite stock index futures contract for the period from January 2, 1986 through December 31, 1989. Each observation is constructed from "Quote Capture" information provided by the Chicago Mercantile Exchange, which specifies the time (to the nearest 10 seconds) and the price of the S&P 500 futures transaction if the price differs from the previously recorded price. Five minute returns are constructed from the last recorded price within each consecutive five minute interval. The first five minute interval of each day is dropped because this interval captures information revealed over a period that exceeds five minutes, namely the time during which the market was closed. Observations from October 15, 1987 through November 13, 1987 are also dropped because of frequent trading suspensions over these four weeks. We are left with 991 trading days; on each trading day we have 80 five minute intervals resulting in 79,280 observations.

As Andersen and Bollerslev (1994) document, intraday returns on the S&P 500 futures index have a strong cyclical component. Because cyclical components, if unaccounted for, affect parameter estimators for conditional heteroscedasticity models, we work with five minute returns that have a cyclical component removed.[18] We estimate

[18] Removal of the cyclical component is accomplished with a Fourier series approximation; details of which are provided in Andersen and Bollerslev (1994, page 20) and whom we thank for providing returns with the cycle removed.

the following MA(1)-GARCH(1,1) model typically used to estimate high frequency data

$$r_{i,t} = \beta + \varepsilon_{i,t} + \theta\varepsilon_{i,t-1},$$

where $i = 1,991$ denotes day, $t = 1,80$ denotes five minute interval, and the conditional variance process is GARCH(1,1). Semiparametric and QML parameter estimates are presented in Table 5. Standard errors, reported in parentheses below each estimate, are constructed from the outer product of the gradient form of estimating the information matrix.

*** TABLE 5 HERE ***

For each set of parameter estimates in Table 5, we construct estimates of the white noise innovations. With each estimated white noise innovation sequence, we tested a null hypothesis of no conditional heteroscedasticity against an alternative hypothesis of ARCH(2). The resulting test statistics are 2.82 for a QML-based test (with a p-value of 0.0024) and 1.87 for the semiparametric test (with a p-value of 0.0317). There is thus, quite weak evidence of ARCH effects, given the sample size.

7. CONCLUSION

The semiparametric test statistics are asymptotically optimal, dominating the widely used Gaussian test statistics according to standard criteria. They are also simple to implement and appear to behave quite well in small samples. For data that are quite far from normal, such as high frequency financial data, our procedure could be very useful.

ACKNOWLEDGEMENTS

We thank G. Andrew Karolyi for providing the five minute price series and Tim Bollerslev for providing the five minute return series. We would like to thank Steve Fox for help with some of the computations.

Appendix

A. Proof of Theorem 5. The leading term in (13), $\mathcal{X}^{-1/2}\mathcal{Y}$, is asymptotically standard normal. To study the behavior of the remaining terms in (13), we approximate $\widehat{\mathcal{Y}} - \mathcal{Y}$ and $\widehat{\mathcal{X}} - \mathcal{X}$.

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 $\overline{g}_t = E[\widehat{g}_t]$ and $\overline{g}_t^{(1)} = E[\widehat{g}_t^{(1)}]$, and write $\widehat{g}_t - g_t = \mathcal{B}_t + \mathcal{V}_t$ and $\widehat{g}_t^{(1)} - g_t^{(1)} = \mathcal{B}_t^{(1)} + \mathcal{V}_t^{(1)}$, where $\mathcal{B}_t = \overline{g}_t - g_t$ and $\mathcal{B}_t^{(1)} = \overline{g}_t^{(1)} - g_t^{(1)}$, while $\mathcal{V}_t = \widehat{g}_t - \overline{g}_t$ and $\mathcal{V}_t^{(1)} = \widehat{g}_t^{(1)} - \overline{g}_t^{(1)}$. From Silverman (1986), the conditional bias of the kernel density estimator $\mathcal{B}_t \approx h^2 g_t^{(2)} \mu_2(K)/2$ and the conditional bias of the kernel first derivative estimator $\mathcal{B}_t^{(1)} \approx h^2 g_t^{(3)} \mu_2(K)/2$, while the "stochastic" terms are such that $\mathcal{V}_t^{(1)} = O_p(T^{-1/2}h^{-3/2})$ dominates $\mathcal{V}_t = O_p(T^{-1/2}h^{-1/2})$. The following additional results are useful in the sequel:

$$\max|\widehat{g}_t - \overline{g}_t| = O_p(T^{-1/2}h^{-1}),\ \max|\widehat{g}_t^{(1)} - \overline{g}_t^{(1)}| = O_p(T^{-1/2}h^{-2}),$$
$$\max|\overline{g}_t - g_t| = O_p(h^2),\ \max|\overline{g}_t^{(1)} - g_t^{(1)}| = O_p(h^2),$$
$$\min \widehat{g}_t = O_p(1).$$

See Andrews (1993, Theorem 1) and Robinson (1987, Lemma 13). Our argument now parallels those presented in Linton (1992, 1994). We need two lemmas which are proved below.

Lemma 1. *By asymptotic expansion*

$$\widehat{\mathcal{Y}} = \mathcal{Y} + \mathcal{L} + \mathcal{Q} + O_p(T^{-3/7}),$$

where $\mathcal{L} = h^2 T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} b_t v_{t-i}^0 = O_p(h^2)$ *and* $\mathcal{Q} = T^{-1/2} \sum_{i=1}^{p} \sum_{t=1}^{T} u_t g_t^{-1} \mathcal{V}_t^{(1)} v_{t-i}^0 = O_p(T^{-1/2}h^{-3/2})$*. Furthermore,* $\mathcal{Q}$ *is mean zero and*

$$\text{var}[\mathcal{Q}] \approx T^{-1}h^{-3}p\nu_2(K^{(1)})(m_4 - m_2^2)E[u_t^2 g^{-1}(u_t)].$$

LEMMA 2. *By asymptotic expansion*

$$\widehat{\mathcal{X}} = \mathcal{X} + \mathcal{B} + O_p(T^{-1/2}),$$

where $\mathcal{B} = 2ph^2 E[\psi_2(u_t)b_t](m_4 - m_2^2) = O(h^2)$.

We are now able make our stochastic expansion. With some rearrangement of terms we arrive at the following approximation:

$$\widehat{\tau}_m = \mathcal{X}^{-1/2}\mathcal{Y} + \mathcal{X}^{-1/2}\mathcal{Q} + \mathcal{X}^{-1/2}[\mathcal{L} - \frac{1}{2}\mathcal{X}^{-1}\mathcal{Y}\mathcal{B}] + O_p(T^{-3/7}).$$

We now calculate the moments of this truncated expansion. Note that var($\mathcal{Y}$)= $\mathcal{X}$. The linear terms are straightforward. Firstly, $\text{var}[\mathcal{L}] = h^4 pE[b_t^2](m_4 - m_2^2)$ and $\text{E}[\mathcal{Y}\mathcal{L}] = h^2(m_4 - m_2^2)p\text{E}[\psi_2(u_t)b_t]$. Therefore,

$$\text{cov}[\mathcal{Y}, \mathcal{L} - \frac{1}{2}\mathcal{X}^{-1}\mathcal{Y}\mathcal{B}] = 0 \;;\; \text{var}[\mathcal{L} - \frac{1}{2}\mathcal{X}^{-1}\mathcal{Y}\mathcal{B}] = h^4 p\mathcal{X}\left[\frac{E[b_t^2]}{I_2} - \left\{\frac{E[\psi_2(u_t)b_t]}{I_2}\right\}^2\right].$$

The quadratic term has variance as stated in Lemma 1.

A.1. Proof of Lemmas.

PROOF OF LEMMA 1.

1. Taylor expansion. A geometric series expansion of $\widehat{g}_t^{(1)}/\widehat{g}_t$, see Härdle and Stoker (1989, pp992), gives

$$\widehat{\mathcal{Y}} - \mathcal{Y} = \sum_{j=1}^{r}\left\{T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\varphi_j(u_t)u_t v_{t-i}^{**}\right\} + R_T,$$

where

$$\varphi_j(u_t) = (-1)^j j![g_t^{-j}(\widehat{g}_t - g_t)^{j-1}(\widehat{g}_t^{(1)} - g_t^{(1)}) + \psi_{1t}g_t^{-j}(\widehat{g}_t - g_t)^j], \quad j = 1,..,r,$$

$$R_T = (-1)^{r+1}(r+1)!T^{-1/2}\textstyle\sum_{i=1}^{p}\sum_{t=1}^{T}[\frac{(\widehat{g}_t-g_t)^r(\widehat{g}_t^{(1)}-g_t^{(1)})}{g_t^r\widehat{g}_t} + \psi_{1t}\frac{(\widehat{g}_t-g_t)^{r+1}}{g_t^r\widehat{g}_t}]u_t v_{t-i}^{**}.$$

We first deal with the leading term

$$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\varphi_1(u_t)u_t v_{t-i}^{**}.$$

We substitute for $\widehat{g}_t - g_t$ and $\widehat{g}_t^{(1)} - g_t^{(1)}$, collect bias terms into $\mathcal{L}$, and note that $T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}u_t g_t^{-2}g_t^{(1)}\mathcal{V}_t v_{t-i}^{**} = O_p(T^{-1/2}h^{-1/2})$, by the same arguments given below in part 2 of the proof. Therefore, the leading term is as required.

The intermediate terms of the form $T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}\varphi_j(u_t)u_t v_{t-i}^{**}$, for $j = 2,..,r$, are U-statistics of order j, and are of progressively smaller order. For example,

$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}g_t^{-2}(\widehat{g}_t - g_t)(\widehat{g}_t^{(1)} - g_t^{(1)})u_t v_{t-i}^{**} =$

$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}g_t^{-2}u_t v_{t-i}^{**}\mathcal{B}_t\mathcal{B}_t^{(1)} + T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}g_t^{-2}u_t v_{t-i}^{**}E_t[\mathcal{V}_t\mathcal{V}_t^{(1)}]+$

$$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T} g_t^{-2}u_t v_{t-i}^{**}\mathcal{B}_t^{(1)}\mathcal{V}_t + T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T} g_t^{-2}u_t v_{t-i}^{**}\mathcal{B}_t\mathcal{V}_t^{(1)} +$$
$$T^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T} g_t^{-2}u_t v_{t-i}^{**}\left\{\mathcal{V}_t\mathcal{V}_t^{(1)} - E_t[\mathcal{V}_t\mathcal{V}_t^{(1)}]\right\}.$$

The first line consists of single sums of order 1 U-statistics with orders in probability $O_p(h^4)$ and $O_p(T^{-1}h^{-2})$ respectively. The second line is a U-statistic of order 2 and is $O_p(h^2T^{-1/2}h^{-1/2})$. The third line is a U-statistic of order 3 and is $O_p(T^{-1}h^{-2})$.

To deal with the remainder term we use crude bounds as in Robinson (1987). Thus

$$\begin{aligned} |R_T| &\leq (r+1)!\{\min \widehat{g}_t\}^{-1}\{\min g_t\}^{-r}T^{1/2}\{\max|\widehat{g}_t - g_t|\}^r\left\{T^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}u_t^2v_{t-i}^{**2}\right\}^{1/2} \\ &\quad\times\left[\left\{T^{-1}\sum_{t=1}^{T}\psi_{1t}^2\right\}^{1/2}\{\max|\widehat{g}_t - g_t|\} + \left\{\max|\widehat{g}_t^{(1)} - g_t^{(1)}|\right\}\right] \\ &= o_p(T^{-3/7}), \end{aligned}$$

provided $r \geq 2$.

2. Moment calculation. We now turn to the variance of $\mathcal{Q}$. We work with the case $p = 1$ for convenience. Let $x_t = u_t g^{-1}(u_t)$ and $\eta_{ts} = \theta_{ts} - E_t(\theta_{ts})$, with $\theta_{ts} = K^{(1)}(\frac{u_t - u_s}{h})$, then $\mathcal{Q} = T^{-3/2}h^{-2}\sum\sum x_t v_{t-1}^0\eta_{ts}$, where summation is over $t = 2, .., T$ and $s \neq t, t-1$. There are three types of non-zero terms in var($\mathcal{Q}$):

(1) $E(x_t^2\ v_{t-1}^{02}\eta_{ts}^2)$, $t = 2, .., T$, $s \neq t, t-1$
(2) $E(x_t v_{t-1}^0\eta_{ts}x_{t+1}v_t^0\eta_{t+1,t-1})$, $t = 2, .., T$, $s \neq t, t-1$
(3) $E(x_t v_{t-1}^0\eta_{ts}x_{s+1}v_s^0\eta_{s+1,t-1})$, $t = 2, .., T$, $s \neq t, t-1$

We first deal with (1). By straightforward calculation,

$$E_t(\theta_{ts}) = \int K^{(1)}(\tfrac{u_t - v}{h})g(v)dv = h\int K(u)g^{(1)}(u_t - hu)du \approx hg^{(1)}(u_t)$$
$$E_t(\theta_{ts}^2) = \int K^{(1)}(\tfrac{u_t - v}{h})^2 g(v)dv = h\int K^{(1)}(u)^2 g(u_t - hu)du \approx hg(u_t)\int K^{(1)}(u)^2 du$$

using integration by parts and a change of variables. Therefore, using conditioning,

$$E(x_t^2 v_{t-1}^{02}\eta_{ts}^2) = (m_4 - m_2^2)E\{u_t^2 g^{-2}(u_t)E_t[\theta_{ts}^2 - E_t^2(\theta_{ts})]\} \approx h\nu_2(K')E[u_t^2 g^{-1}(u_t)](m_4 - m_2^2), \quad \star$$

and there are $O(T^2)$ terms of this form.

We now turn to (2). Unless $s = t+1$, this expectation factors into

$$E(x_{t+1}v_{t-1}^0\eta_{t+1,t-1})E(x_t\eta_{ts}v_t^0) = 0.$$

Therefore, (2) contributes only $O(T)$ terms each of $O(h)$.

As for (3), this factors into

$$E(x_t\eta_{ts}v_s^0)E(x_{s+1}\eta_{s+1,t-1}v_{t-1}^0) = E^2(x_t\eta_{ts}v_s^0) = O(h^2),$$

since $E(x_t\eta_{ts}v_s^0) = O(h)$. There are $O(T^2)$ terms, but they are of smaller order than the (1) terms.

The result follows by multiplying $\star$ by $T^{-1}h^{-4}$. When $p > 1$, the same calculations are involved, and the answer gets multiplied by p. ∎

Proof of Lemma 2. We use the further linearization $\widehat{\psi}_2^2 - \psi_2^2 \approx 2\psi_2(\widehat{\psi}_2 - \psi_2)$, and the same arguments as given above. ∎

REFERENCES

[1] ANDERSEN, T. AND T. BOLLERSLEV (1994): "Intraday seasonality and volatility persistence in foreign exchange and equity markets," *Northwestern University Discussion Paper no. 186.*

[2] ANDREWS, D.W.K. (1993): "Nonparametric kernel estimation for semiparametric models," Forthcoming in *Econometric Theory.*

[3] ANDREWS, D.W.K. (1994): "Hypothesis testing with a restricted parameter space." *Cowles Foundation Discussion Paper no. 1060R.*

[4] 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.

[5] BERA, A. AND M. HIGGINS (1993): "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys* **7**, 305-66.

[6] BERA, A. AND P. NG (1991): "Robust tests for heteroskedasticity and autocorrelation using score function." Unpublished manuscript.

[7] BICKEL, P., KLAASSEN, C.A.J., RITOV, Y., AND J.A. WELLNER (1994): *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press: Baltimore.

[8] BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1995): "ARCH Models." In *The Handbook of Econometrics*, Vol. IV. Eds. D.F. McFadden and R.F. Engle III. North-Holland.

[9] BOLLERSLEV, T., AND J.M. WOOLDRIDGE (1992): "Quasi Maximum Likelihood Estimation and inference in dynamic models with time varying covariates." *Econometric Reviews* **11**, 143-172.

[10] BREUSCH, T.S. AND A.R. PAGAN (1979): "A simple test for heteroscedasticity and random coefficient variation." *Econometrica* **47**, 1287-1294.

[11] ENGLE, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation" *Econometrica* **50**, 987-1007.

[12] ENGLE, R. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics", in: Z. Griliches and M.D. Intriligator, eds., *The Handbook of Econometrics, Vol. II.* North Holland, Amsterdam.

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

[14] GOURIEROUX, C., HOLLY, A., AND A. MONFORT (1982): "Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constraints on regression parameters," *Econometrica* **50**, 63-80.

[15] HAJEK, J. (1972): "Local Asymptotic Minimax and Admissibility in Estimation" in *Proceedings*

of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, L. LeCam, J. Neyman, and E. Scott (eds), University of California Press.

[16] HÄRDLE, W., J. HART, J. S. MARRON, AND A. B. TSYBAKOV (1992): "Bandwidth Choice for Average Derivative Estimation," *Journal of the American Statistical Association* 87, 218-226.

[17] HÄRDLE, W., AND O.B. LINTON (1995): "Applied nonparametric methods," In *The Handbook of Econometrics*, Vol. IV. Eds, D.F. McFadden and R.F. Engle III. North Holland.

[18] HÄRDLE, W., AND T. M. STOKER (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association* 84, 986-995.

[19] HSIEH, D. (1989): "Modeling Heteroscedasticity in Daily Foreign-Exchange Rates" *Journal of Business and Economic Statistics* **7**, 307-317.

[20] JEGANATHAN, P. (1987): "Some aspects of asymptotic theory with applications to time series models," Forthcoming in *Econometric Theory.*

[21] JONES, M.C., AND P.J. FOSTER (1995): "Generalized jacknifing and higher order kernels," Forthcoming in *Journal of Nonparametric Statistics.*

[22] JONES, M.C., O.B. LINTON, AND J.P. NIELSEN (1993): "A Simple and Effective Bias Reduction Method for Density Estimation," Forthcoming in *Biometrika.*

[23] KING, M.L. AND P.X. WU (1991): "Locally optimal one-sided tests for multiparameter hypotheses,"

[24] KREISS, J. (1987): "On Adaptive Estimation in Stationary ARMA Processes," *Annals of Statistics* **15**, 112-133.

[25] KREISS, J. (1990): "Testing Linear Hypotheses in Autoregression," *Annals of Statistics* **18**, 1470-1482.

[26] LEE, J.H.H. (1991): "A Lagrange Multiplier Test for GARCH models," *Economics Letters* **37**, 265-271.

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

[28] LINTON, O.B. (1992): "Second Order Approximation in a linear regression with heteroskedasticity of unknown form." Nuffield College Discussion paper no 75. *Econometric Reviews.* Forthcoming.

[29] LINTON, O.B. (1993): "Adaptive Estimation in ARCH Models." *Econometric Theory* **9**, 539-569.

[30] LINTON, O.B. (1994): "Second Order Approximation in the Partially Linear Model," Cowles Foundation Discussion Paper no 1065. Forthcoming in *Econometrica*

[31] LUMSDAINE, R.L. (1991): "Consistency and asymptotic normality of the Quasi-Maximum Likelihood estimator in GARCH(1,1) and IGARCH(1,1) models." Manuscript, Princeton University.

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

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

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

[35] NEWEY, W. AND D. STEIGERWALD (1994): "Consistency of Quasi-Maximum Likelihood Estimators for Models with Conditional Heteroscedasticity," manuscript, University of California, Santa Barbara.

[36] ROBINSON, P.M. (1983): "Nonparametric Estimators for Time Series." *Journal of Time Series Analysis* 185-208.

[37] ROBINSON, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown form." *Econometrica* **56**, 875-891.

[38] ROUSSAS, G. (1972): *Contiguity of probability measures: some applications in statistics.* Cambridge University Press, Cambridge.

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

[40] SILVERMAN, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

[41] STEIGERWALD, D. (1994): "Efficient estimation of models with conditional heteroscedasticity." Unpublished manuscript.

[42] STRASSER, H (1985): *Mathematical theory of statistics.* de Gruyter. Berlin.

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

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

[45] WU, P.X. AND KING, M.L. (1991): "One-sided hypothesis testing in econometrics: A survey,"

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

TABLE 1. Summary Statistics of the Densities All densities are subsequently rescaled to have mean 0 and variance 1

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
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

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

Density	**QML**	**SP3**	**QML**	**SP3**
Gaussian	0.030	0.032	0.945	0.937
A1	0.026	0.028	0.920	0.934
A2	0.024	0.033	0.804	0.879
A3	0.021	0.021	0.390	0.689
L1	0.026	0.026	0.915	0.909
L2	0.024	0.028	0.844	0.838
L3	0.027	0.038	0.736	0.752
P1	0.030	0.028	0.964	0.972
P2	0.034	0.027	0.996	0.997
P3	0.046	0.000	0.995	0.975

TABLE 3. ARCH(1) vs. ARCH(2) T=500.

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

TABLE 4. Size and Size-Adjusted Power: White Noise vs. ARCH(2).

	β	θ	ϕ	ρ
QML	1.196x10-5	-.1982	.0846	.7962
	(1.324x10-6)	(.0445)	(.0094)	(.0143)
Semiparametric	1.195x10-5	-.1874	.0793	.7810
	(2.655x10-7)	(.0399)	(.0065)	(.0102)

TABLE 5. Estimates for S&P 500 Futures Data. Parameters ϕ and ρ are defined following (2). Standard Errors in Parentheses