QUASI-MAXIMUM LIKELIHOOD ESTIMATION AND INFERENCE IN DYNAMIC MODELS WITH TIME-VARYING COVARIANCES

Tim BOLLERSLEV
Department of Finance
J.L. Kellogg Graduate School of Management
Northwestern University

and

Jeffrey M. WOOLDRIDGE
Department of Economics
Massachusetts Institute of Technology

ABSTRACT

We study the properties of the quasi-maximum likelihood estimator (QMLE) and related test statistics in dynamic models that jointly parameterize conditional means and conditional covariances, when a normal log-likelihood is maximized but the assumption of normality is violated. Because the score of the normal log-likelihood has the martingale difference property when the first two conditional moments are correctly specified, the QMLE is generally consistent and has a limiting normal distribution. We provide easily computable formulas for asymptotic standard errors that are valid under nonnormality. Further, we show how robust LM tests for the adequacy of the jointly parameterized mean and variance can be computed from simple auxiliary regressions. An appealing feature of these robust inference procedures is that only first derivatives of the conditional mean and variance functions are needed. A Monte Carlo study indicates that the asymptotic results carry over to finite samples. Estimation of several AR and AR-GARCH time series models reveals that in most situations the robust test statistics compare favorably to the two standard (nonrobust) formulations of the Wald and LM tests. Also, for the GARCH models and the sample sizes analyzed here, the bias in the QMLE appears to be relatively small. An empirical application to stock return volatility illustrates the potential importance of computing robust statistics in practice.

1. INTRODUCTION

Dynamic econometric models that jointly parameterize conditional means, conditional variances, and conditional covariances are becoming increasingly popular in the analysis of economic time series. Engle's (1982a) pioneering autoregressive conditional heteroskedasticity (ARCH) model has been expanded and adapted for application in several diverse fields. One useful extension is the generalized ARCH (GARCH) model introduced by Bollerslev (1986), which allows for richer dynamics in the conditional second moments. The ARCH-in-mean (ARCH-M) model introduced by Engle, Lilien, and Robins (1987) has been successfully applied in financial economics to both univariate and multivariate dynamic asset pricing models, where conditional mean equations that contain conditional second moments arise naturally from considerations of attitudes toward risk. Bollerslev, Chou, and Kroner (1990) provide a recent survey of this literature.

As demonstrated by Pagan and Ullah (1988), certain hypotheses involving risk measures can be tested by means of instrumental variables (IV) estimation without explicitly parameterizing the relevant conditional variances and covariances. Unfortunately, the IV approach is not very helpful when interest lies in obtaining estimates of the risk premia, as the premia depend directly on the conditional second moments. A further limitation of the IV approach is that, under data generating mechanisms such as the ARCH-M model, evidently no instrumental variable estimator exists that does not require *a priori* knowledge of the ARCH-M structure (Pagan and Ullah (1988, p.99)). For these reasons, studies that jointly estimate dynamic conditional means and conditional second moments have relied heavily on maximum likelihood procedures, frequently under the assumption of conditional normality.

Taken literally, the assumption of conditional normality can be quite restrictive. The symmetry imposed under normality is difficult to justify in general, and the tails of even conditional distributions often seem to be fatter than that of the normal distribution. The extensive use of maximum likelihood under the assumption of normality is almost certainly due to its relative simplicity and the widespread familiarity with its properties under ideal conditions.

Because maximum likelihood under normality is so widely used, it is important to investigate its properties in a setting general enough to include most cases of interest to applied researchers. The purpose of this paper is to study the behavior of the quasi-maximum likelihood estimator (QMLE) and related test statistics in a general class of dynamic models when a normal log-likelihood is maximized but the normality assumption is violated. An important conclusion, developed in section 2, is that the QMLE is still consistent for the parameters of the jointly parameterized conditional mean and conditional variance. While there exists an econometric folklore suggesting that this is the case, and while special cases of this result have appeared in the literature, we know of no rigorous statement of the consistency of the QMLE for general dynamic, multivariate models.

Section 2 also derives easily computable formulas for the asymptotic standard errors that are valid under nonnormality. These formulas facilitate the construction of computationally simple Wald statistics that are valid under nonnormality, yet still optimal under normality. Section 3 derives robust, regression-based Lagrange multiplier (LM) diagnostics that can be used to check the adequacy of the specification of the first two conditional moments. Together, sections 2 and 3 contain the results and formulas needed to conduct inference about dynamic conditional means and second moments while being robust to nonnormality. An appealing feature of these results is that only first derivatives of the mean and variance functions are needed to compute all of the robust statistics.

Section 4 presents some Monte Carlo evidence on the performance of both the robust tests and the more popular nonrobust tests, while section 5 contains an empirical application to stock return volatility. Broadly speaking, the results for the robust statistics are quite encouraging, and suggest their usefulness in empirical studies.

2. CONSISTENCY AND ASYMPTOTIC NORMALITY OF THE QMLE

Let $\{(y_t,z_t)\colon t=1,2,\ldots\}$ be a sequence of observable random vectors with y_t K x 1, z_t L x 1. The vector y_t contains the "endogenous" variables and z_t contains contemporaneous "exogenous" (conditioning) variables. Let $x_t \equiv (z_t,y_{t-1},z_{t-1},\ldots,y_1,z_1)$ denote the predetermined variables. The purpose of the analysis is to estimate and test hypotheses about the conditional expectation and conditional variance of y_t given the predetermined variables x_t. If one wants to condition only on information observed before t, z_t can be excluded from x_t without altering any of the subsequent analysis. Cross section analysis is accomodated by setting $x_t \equiv z_t$ and assuming that the observations are independently distributed.

The conditional mean and variance functions are jointly parameterized by a finite dimensional vector θ:

$$\{\mu_t(x_t,\theta): \theta \in \Theta\}$$
$$\{\Omega_t(x_t,\theta): \theta \in \Theta\},$$

where Θ is a subset of $\mathbb{R}^P$ and μ_t and Ω_t are known functions of x_t and θ. In the subsequent analysis, the validity of most of the inference procedures is explicitly proven under the null hypothesis that the first two conditional moments are correctly specified. More formally, for some $\theta_o \in \Theta$,

(2.1.a) $$E(y_t|x_t) = \mu_t(x_t,\theta_o)$$

(2.1.b) $$V(y_t|x_t) = \Omega_t(x_t,\theta_o), \quad t=1,2,....$$

Sometimes one is interested in testing (2.1.a) while being robust to departures from (2.1.b). In section 3 we briefly discuss how to compute conditional mean statistics that are robust to violation of (2.1.b). This is meaningful only when it is possible to separate the conditional mean and variance functions in an appropriate sense. Also, in some of the simulations in section 4 we impose only (2.1.a) under the null hypothesis and investigate the robustness of various statistics to departures from (2.1.b).

The procedure most often used to estimate θ_o is maximization of a likelihood function that is constructed under the assumption that y_t given x_t is normally distributed with mean and variance given by (2.1). This is the approach taken here as well, but the subsequent analysis does not assume that y_t has a conditional normal distribution. Nevertheless, as stated in Theorem 2.1 below, the resulting QMLE is generally consistent for θ_o and under standard regularity conditions it is asymptotically normally distributed.

Rather than employing quasi-maximum likelihood to estimate θ_o, it is straightforward to use (2.1) to construct generalized method of moments (GMM) estimators for θ_o. The results of Chamberlain (1982), Hansen (1982), White (1982b), and Cragg (1983) can be extended to produce an instrumental variables estimator asymptotically more efficient than the QMLE under nonnormality. Further, under enough regularity conditions, it is almost certainly possible to obtain an estimator with variance that achieves the semiparametric lower bound (see Chamberlain (1987)). However, this would require nonparametric estimation of dynamic conditional third and fourth moments, as well as numerous cross product moments in a multivariate context. Because of the simplicity and undisputed popularity of QMLE, we focus on it and leave investigation of method of moments and semiparametric methods to future research.

For observation t, the quasi-conditional log-likelihood apart from a constant is

$$\ell_t(\theta;y_t,x_t) = -1/2 \log|\Omega_t(x_t,\theta)| - 1/2\,(y_t - \mu_t(x_t,\theta))'\Omega_t^{-1}(x_t,\theta)(y_t - \mu_t(x_t,\theta)). \tag{2.2}$$

Letting $\varepsilon_t(y_t,x_t,\theta) \equiv y_t - \mu_t(x_t,\theta)$ denote the K x 1 residual function and suppressing the dependence of ε_t and Ω_t on x_t and y_t yields the more concise expression

$$\ell_t(\theta) = -1/2 \log|\Omega_t(\theta)| - 1/2\,\varepsilon_t(\theta)'\Omega_t^{-1}(\theta)\varepsilon_t(\theta). \tag{2.3}$$

The QMLE $\hat{\theta}_T$ is obtained by maximizing the quasi-log likelihood function

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \ell_t(\theta). \tag{2.4}$$

If $\mu_t(x_t,\cdot)$ and $\Omega_t(x_t,\cdot)$ are differentiable on Θ for all relevant x_t, and if $\Omega_t(x_t,\theta)$ is nonsingular with probability one for all $\theta \in \Theta$, then differentiation of (2.3) yields the 1 x P score function $\mathbf{s}_t(\theta)$:

$$\mathbf{s}_t(\theta)' \equiv \nabla_\theta \ell_t(\theta)' = \nabla_\theta \mu_t(\theta)'\Omega_t^{-1}(\theta)\varepsilon_t(\theta) + 1/2\,\nabla_\theta \Omega_t(\theta)'[\Omega_t^{-1}(\theta) \otimes \Omega_t^{-1}(\theta)]\mathrm{vec}[\varepsilon_t(\theta)\varepsilon_t(\theta)' - \Omega_t(\theta)]. \tag{2.5}$$

where $\nabla_\theta \mu_t(\theta)$ is the K x P derivative of μ_t and $\nabla_\theta \Omega_t(\theta)$ is the K^2 x P derivative of $\Omega_t(\theta)$; see Appendix A for the definition of the derivative of a matrix.

If (2.1.a) holds then the true error vector is defined as $\boldsymbol{\varepsilon}_t^o \equiv \boldsymbol{\varepsilon}_t(\theta_o) = y_t - \boldsymbol{\mu}_t(x_t,\theta_o)$ and $E(\boldsymbol{\varepsilon}_t^o|x_t) = 0$. If, in addition, (2.1.b) holds then $E(\boldsymbol{\varepsilon}_t^o\boldsymbol{\varepsilon}_t^{o\prime}|x_t) = \Omega_t(x_t,\theta_o)$. It follows from (2.5) that, under correct specification of the first two conditional moments of y_t given x_t,

(2.6) $$E[\mathbf{s}_t(\theta_o)|x_t] = 0.$$

An immediate implication of (2.6) is that the score evaluated at the true parameter is a vector martingale difference sequence with respect to the σ-fields $\{\sigma(y_t,x_t): t=1,2,...\}$. This property of the score of the conditional log-likelihood is well known when the conditional density is correctly specified; see, for example, Crowder (1976), Basawa, Feigen and Heyde (1976), and Heijmans and Magnus (1986). The above analysis demonstrates that the score of a normal log-likelihood has the martingale difference property provided only that the first two moments are correctly specified. This result extends that of Weiss (1986), who considers a univariate ARMA model with ARCH errors. In cross section settings, MaCurdy (1981) and Gourieroux, Monfort and Trognon (1984) have shown that the score evaluated at the true parameter has zero expectation without the normality assumption; (2.6) is the extension to dynamic models.

In related work, Pagan and Sabau (1987) examine the robustness of the QMLE in a univariate linear model with conditional heteroskedasticity. However, their focus is on the consistency of the conditional mean parameters when the conditional variance is misspecified. They are not concerned with consistency of the mean and variance parameters under nonnormality, nor do they present limiting distribution results.

Equation (2.6) can be used in the approach of Wooldridge (1986, Chapter 3) to establish weak consistency of the QMLE. Because the regularity conditions involved with this approach are more complicated than the scope of the current paper warrants, we instead prove weak consistency by adopting the uniform law of large numbers approach of Domowitz and White (1982) (see Appendix A). Identifiability of θ_o is established by showing that θ_o maximizes $E[\mathcal{L}_T(\theta)]$. The proof of asymptotic normality of the QMLE in Appendix A does rely directly on (2.6) since a martingale central limit theorem is applied to $\{\mathbf{s}_t(\theta_o): t=1,2,...\}$.

For robust inference we also need an expression for the hessian $\mathbf{h}_t(\theta)$ of $\ell_t(\theta)$. Actually, for computations, it is useful to observe that when (2.1) holds all that is needed is $E[\mathbf{h}_t(\theta_o)|x_t]$. This matrix has a very convenient form that involves only first derivatives of the conditional mean and conditional variance functions. Define the P x P symmetric, positive semi-definite matrix $\mathbf{a}_t(\theta_o) \equiv -E[\nabla_\theta \mathbf{s}_t(\theta_o)|x_t] = E[-\mathbf{h}_t(\theta_o)|x_t]$. A straightforward calculation shows that, under (2.1.a) and (2.1.b),

$$(2.7) \qquad \mathbf{a}_t(\theta_o) = \nabla_\theta \mu_t(\theta_o)' \Omega_t^{-1}(\theta_o)\ \nabla_\theta \mu_t(\theta_o) + 1/2\ \nabla_\theta \Omega_t(\theta_o)'[\Omega_t^{-1}(\theta_o) \otimes \Omega_t^{-1}(\theta_o)]\nabla_\theta \Omega_t(\theta_o)$$

(see Kroner (1987, Lemma 1) for derivation of a similar result under normality without the conditional mean parameters). When the normality assumption holds the matrix $\mathbf{a}_t(\theta_o)$ is the conditional information matrix. However, if y_t does not have a conditional normal distribution then $V[\mathbf{s}_t(\theta_o)|x_t]$ is generally not equal to $\mathbf{a}_t(\theta_o)$ and the information matrix equality is violated. Nevertheless, it is fairly easy to carry out inference about θ_o. The proof of the following theorem is provided in Appendix A.

THEOREM 2.1: Suppose that the following conditions hold:

(i) Regularity Conditions A.1 in Appendix A;

(ii) For some $\theta_o \in \text{int}\ \Theta$ and t=1,2,...,

$$E(y_t|x_t) = \mu_t(x_t,\theta_o) \text{ and } V(y_t|x_t) = \Omega_t(x_t,\theta_o).$$

Then

$$\left(\mathbf{A}_T^{o-1}\mathbf{B}_T^o\mathbf{A}_T^{o-1}\right)^{-1/2}\sqrt{T}(\hat{\theta}_T - \theta_o) \overset{d}{\rightarrow} N(0,\mathbf{I}_P),$$

where

$$\mathbf{A}_T^o \equiv E[-\mathbf{H}_T(\theta_o)/T] = -T^{-1}\sum_{t=1}^{T} E[\mathbf{h}_t(\theta_o)] = T^{-1}\sum_{t=1}^{T} E[\mathbf{a}_t(\theta_o)]$$

and

$$\mathbf{B}_T^o \equiv V[T^{-1/2}\mathbf{s}_T(\theta_o)] = T^{-1}\sum_{t=1}^{T} E[\mathbf{s}_t(\theta_o)'\mathbf{s}_t(\theta_o)].$$

In addition,

$$\hat{\mathbf{A}}_T - \mathbf{A}_T^o \overset{p}{\rightarrow} \mathbf{0} \text{ and } \hat{\mathbf{B}}_T - \mathbf{B}_T^o \overset{p}{\rightarrow} \mathbf{0},$$

where

$$\hat{\mathbf{A}}_T \equiv T^{-1}\sum_{t=1}^{T} \mathbf{a}_t(\hat{\theta}_T) \text{ and } \hat{\mathbf{B}}_T \equiv T^{-1}\sum_{t=1}^{T} \mathbf{s}_t(\hat{\theta}_T)'\mathbf{s}_t(\hat{\theta}_T). \quad \blacksquare$$

The matrix $\hat{\mathbf{A}}_T^{-1}\hat{\mathbf{B}}_T\hat{\mathbf{A}}_T^{-1}$ is a consistent estimator of the White (1982a) robust asymptotic variance matrix of $\sqrt{T}(\hat{\theta}_T - \theta_o)$. In practice, one treats $\hat{\theta}_T$ is if it is normally distributed with "mean" θ_o and "variance" $\hat{\mathbf{A}}_T^{-1}\hat{\mathbf{B}}_T\hat{\mathbf{A}}_T^{-1}/T$. Under normality, the variance estimator can be replaced by $\hat{\mathbf{A}}_T^{-1}/T$ (Hessian form) or $\hat{\mathbf{B}}_T^{-1}/T$ (outer product of the gradient form).

The regularity conditions listed in Appendix A are somewhat abstract but widely applicable. If $\{(y_t,z_t):t=1,2,...\}$ is stationary and ergodic and μ_t and Ω_t depend on t only through x_t, then Theorem 2.1 holds under additional moment conditions. For example, the ergodic models of Nelson (1990b) can readily be handled, provided enough moments are finite to apply the law of large numbers and central limit theorem. Theorem 2.1 also potentially applies to integrated GARCH models, as Nelson (1990a) has shown that the processes generated by these models are effectively stationary and ergodic. In fact, for the GARCH(1,1) and IGARCH(1,1) models analyzed further in sections 4 and 5 below, some of the regularity conditions have already been verified by Lumsdaine (1990), who uses a slightly different approach in proving asymptotic normality of the QMLE for this particular class of models. Unfortunately, we can say little more because verification of Conditions A.1 necessarily proceeds on a case-by-case basis.

The estimators $\hat{\mathbf{A}}_T$ and $\hat{\mathbf{B}}_T$ have the convenient property of being at least positive semi-definite and usually positive definite. Moreover, they are computable entirely from the residuals $\hat{\boldsymbol{\varepsilon}}_t$, the mean and variance functions $\hat{\boldsymbol{\mu}}_t$ and $\hat{\Omega}_t$, and the first derivatives of the mean and variance functions $\nabla_\theta\hat{\boldsymbol{\mu}}_t$ and $\nabla_\theta\hat{\Omega}_t$. Thus, they do not require second derivatives of either the mean or variance functions. This is a useful simplification because estimation of these models typically relies on numerical approximations to the analytical derivatives.

With Theorem 2.1 in place it is straightforward to construct Wald statistics for testing hypotheses about $\boldsymbol{\theta}_o$. Assume that the null hypothesis can be stated as

$$H_0\colon \mathbf{c}(\theta_o) = \mathbf{0},$$

where $\mathbf{c}\colon\Theta \to \mathbb{R}^Q$ is continuously differentiable on int Θ and $Q < P$. Let $\mathbf{C}(\theta) \equiv \nabla_\theta\mathbf{c}(\theta)$ be the Q x P gradient of $\mathbf{c}$ on int(Θ). If $\theta_o \in \text{int}(\Theta)$ and rank $\mathbf{C}(\theta_o) = Q$ then, under the conditions of Theorem 2.1, the Wald statistic

$$W_T \equiv T\mathbf{c}(\hat{\theta}_T)'[\hat{\mathbf{C}}_T\hat{\mathbf{A}}_T^{-1}\hat{\mathbf{B}}_T\hat{\mathbf{A}}_T^{-1}\hat{\mathbf{C}}_T']^{-1}\mathbf{c}(\hat{\theta}_T) \tag{2.8}$$

has an asymptotic χ^2_Q distribution under H_0, where $\hat{\mathbf{C}}_T \equiv \mathbf{C}(\hat{\theta}_T)$. Again, we emphasize that the robust Wald statistic is computable entirely from first derivatives and has an asymptotic chi-square distribution whether or not the conditional normality assumption holds. Wald tests constructed from either the inverse of the Hessian or the the outer product of the gradient will not generally lead to valid inference.

3. ROBUST LAGRANGE MULTIPLIER TESTS

Because estimation of the models considered in this paper can be computationally difficult, it is useful to have diagnostics that are computable from a constrained model. In this section we derive a robust form of the Lagrange multiplier (LM) or efficient score statistic that is computable from statistics obtained after a single iteration away from the restricted model. Assume that the hypothesis of interest,

$$H_0\colon \mathbf{c}(\theta_o) = 0,$$

can be expressed as

$$H_0\colon \theta_o = \mathbf{r}(\alpha_o) \text{ for some } \alpha_o \in A,$$

where $A \subset \mathbb{R}^M$ and $M \equiv P - Q$. The function $\mathbf{r}\colon A \rightarrow \Theta$, which implicitly defines the constraints on θ, is assumed to be continuously differentiable on the interior of A, and $\alpha_o \in \text{int } A$. Note that for the LM test we require only that α_o be in the interior of its parameter space; θ_o is allowed to be on the boundary of Θ. This is especially useful in the present context, where hypotheses concerning the conditional variances and covariances necessarily impose nonnegativity restrictions. Let $\mathbf{R}(\alpha) \equiv \nabla_\alpha \mathbf{r}(\alpha)$ be the P x M gradient of $\mathbf{r}$.

The LM test is based on the gradient of the log-likelihood evaluated at the constrained QMLE. Let $\tilde{\boldsymbol{\alpha}}_T$ be the constrained QMLE of $\boldsymbol{\alpha}_o$ so that the constrained QMLE of $\boldsymbol{\theta}_o$ is $\tilde{\boldsymbol{\theta}}_T \equiv \mathbf{r}(\tilde{\boldsymbol{\alpha}}_T)$. The LM statistic is a quadratic form in the P x 1 vector

$$T^{-1/2}\mathbf{s}_T(\tilde{\boldsymbol{\theta}}_T)' \equiv T^{-1/2}\sum_{t=1}^{T} \mathbf{s}_t(\tilde{\boldsymbol{\theta}}_T)' \equiv T^{-1/2}\sum_{t=1}^{T} \tilde{\mathbf{s}}_t'.$$

Under conditional normality, the outer product of the gradient (OPG) LM statistic is obtained as TR_u^2 from the outer product regression

$$1 \text{ on } \tilde{\mathbf{s}}_t, \quad t=1,\ldots,T, \tag{3.6}$$

where R_u^2 is the uncentered r-squared. Under conditional normality, $TR_u^2 \sim \chi_Q^2$. If the conditional distribution of y_t given x_t is not normal then the limiting distribution is generally not χ_Q^2, and the nominal size can be very different from the actual size. The OPG LM statistic can also have poor finite sample properties even under normality (see Davidson and MacKinnon (1985) and section 4). Other forms of the LM statistic, in particular those based on generalized residuals, have better finite sample properties under normality but are still invalid under nonnormality. The power of the nonrobust test statistics for alternatives to the mean and variance can also be adversely affected if normality does not hold.

To derive a test of H_0 which is robust to nonnormality, we extend the univariate case considered in Wooldridge (1990, Example 3.3). First, express the (unrestricted) score in (2.5) as

$$\mathbf{s}_t(\boldsymbol{\theta})' = \begin{pmatrix} \nabla_\theta \boldsymbol{\mu}_t(\boldsymbol{\theta}) \\ \nabla_\theta \Omega_t(\boldsymbol{\theta}) \end{pmatrix}' \cdot \begin{pmatrix} \Omega_t^{-1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & [\Omega_t^{-1}(\boldsymbol{\theta}) \otimes \Omega_t^{-1}(\boldsymbol{\theta})]/2 \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) \\ \mathrm{vec}[\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})' - \Omega_t(\boldsymbol{\theta})] \end{pmatrix}. \tag{3.7}$$

Evaluating $\mathbf{s}_t$ at $\mathbf{r}(\alpha)$ yields the score with the restrictions imposed:

$$(3.8) \qquad \mathbf{s}_t(\mathbf{r}(\alpha))' = \begin{pmatrix} \nabla_\theta \mu_t(\mathbf{r}(\alpha)) \\ \nabla_\theta \Omega_t(\mathbf{r}(\alpha)) \end{pmatrix}' \cdot \begin{pmatrix} \Omega_t^{-1}(\mathbf{r}(\alpha)) & \mathbf{0} \\ \mathbf{0} & [\Omega_t^{-1}(\mathbf{r}(\alpha)) \otimes \Omega_t^{-1}(\mathbf{r}(\alpha))]/2 \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\varepsilon}_t(\mathbf{r}(\alpha)) \\ \mathrm{vec}[\boldsymbol{\varepsilon}_t(\mathbf{r}(\alpha))\boldsymbol{\varepsilon}_t(\mathbf{r}(\alpha))' - \Omega_t(\mathbf{r}(\alpha))] \end{pmatrix}$$

$$\equiv \Lambda_t(\alpha)'\Gamma_t(\alpha)^{-1}\eta_t(\alpha),$$

where $\Lambda_t(\alpha)$ is $(K+K^2) \times P$, $\Gamma_t(\alpha)$ is $(K+K^2) \times (K+K^2)$, and $\eta_t(\alpha)$ is $K+K^2 \times 1$. Note that $\eta_t(\alpha)$ is a vector of generalized residuals; in particular, $E[\eta_t(\alpha_o)|x_t] = 0$ under H_0. Let $\mathbf{m}_t(\alpha) \equiv \mu_t(\mathbf{r}(\alpha))$ and $\mathbf{W}_t(\alpha) \equiv \Omega_t(\mathbf{r}(\alpha))$ be the restricted mean and variance functions, respectively, with gradients

$$\nabla_\alpha \mathbf{m}_t(\alpha) = \nabla_\theta \mu_t(\mathbf{r}(\alpha))\mathbf{R}(\alpha)$$

$$\nabla_\alpha \mathbf{W}_t(\alpha) = \nabla_\theta \Omega_t(\mathbf{r}(\alpha))\mathbf{R}(\alpha).$$

Note that $\nabla_\alpha \mathbf{m}_t(\alpha)$ is $K \times M$ and $\nabla_\alpha \mathbf{W}_t(\alpha)$ is $K^2 \times M$. It is convenient to stack these gradients into the $K+K^2 \times M$ matrix $\Psi_t(\alpha)$:

$$\Psi_t(\alpha) \equiv \begin{pmatrix} \nabla_\alpha \mathbf{m}_t(\alpha) \\ \nabla_\alpha \mathbf{W}_t(\alpha) \end{pmatrix}.$$

The restricted residual function is $e_t(\alpha) \equiv \boldsymbol{\varepsilon}_t(\mathbf{r}(\alpha))$. Finally, let values labelled with tilde be evaluated at $\tilde{\alpha}_T$, for example, $\tilde{\Lambda}_t \equiv \Lambda_t(\tilde{\alpha}_T)$, $\tilde{\Gamma}_t \equiv \Gamma_t(\tilde{\alpha}_T)$, $\tilde{\eta}_t \equiv \eta_t(\tilde{\alpha}_T)$, and $\tilde{\Psi}_t \equiv \Psi_t(\tilde{\alpha}_T)$. Then the robust LM (RB LM) test can be computed from Wooldridge (1990, Theorem 2.1):

PROCEDURE 3.1:

(i) Compute $\tilde{\alpha}_T$, $\tilde{\mathbf{m}}_t$, $\tilde{\mathbf{W}}_t$, $\tilde{e}_t$, $\nabla_\alpha \tilde{\mathbf{m}}_t$, $\nabla_\alpha \tilde{\mathbf{W}}_t$, $\nabla_\theta \tilde{\mu}_t$, and $\nabla_\theta \tilde{\Omega}_t$.

(ii) Run the matrix regression

$$\tilde{\Gamma}_t^{-1/2}\tilde{\Lambda}_t \quad \text{on} \quad \tilde{\Gamma}_t^{-1/2}\tilde{\Psi}_t \qquad t=1,2,\ldots,T$$

and save the residuals, say $\ddot{\Lambda}_t$.

(iii) Run the OLS regression

$$1 \quad \text{on} \quad \ddot{\eta}_t' \ddot{\Lambda}_t \qquad t=1,\ldots,T,$$

where $\ddot{\eta}_t \equiv \tilde{\Gamma}_t^{-1/2}\tilde{\eta}_t$, $t=1,\ldots,T$, are the weighted generalized residuals. Under H_0, $TR_u^2 = T - SSR$ is asymptotically χ_Q^2, where SSR is the usual sum of squared residuals. ■

This form of the LM statistic has some attractive features. First, just as with the robust Wald statistic, the procedure is valid under nonnormality and loses nothing in terms of asymptotic local power if the normality assumption happens to hold. Second, it requires only the estimates from the restricted model, and there is no need to explicitly specify the constraint function **r** or its gradient. Finally, only first derivatives of the conditional mean and variance functions (evaluated at the restricted estimates) are needed for the computations.

A useful feature of both the robust Wald and LM tests is that, when the mean and variance can be appropriately separated, allowing for consistent estimation of the mean parameters under nonnormality <u>and</u> violation of (2.1.b), the conditional mean tests are in some cases robust to misspecification of $V(y_t|x_t)$. In the Wald case conditional mean tests are constructed by focusing only on elements of θ that index the conditional mean. However, we caution that consistency of the QMLE for the conditional mean parameters is <u>not</u> alone sufficient for the asymptotic covariance formula for the mean parameters, given by the appropriate block of $\hat{\mathbf{A}}_T^{-1}\hat{\mathbf{B}}_T\hat{\mathbf{A}}_T^{-1}/T$, to be valid. Consequently, the conditions of Pagan and Sabau's (1987) Theorem 5 are not sufficient for the robust Wald statistic for the conditional mean to be valid in the presence of a misspecified ARCH model. But if $\theta \equiv (\beta',\gamma')'$, where the mean parameters β and the variance parameters γ are variation free, then the robust Wald test for β_o which ignores the randomness of $\hat{\gamma}_T$ are valid under violations of (2.1.b) (and normality). Interestingly, it follows from the results of Wooldridge (1990) that RB LM is valid provided the conditional mean parameters are $\sqrt{T}$-consistent and one chooses $\tilde{\eta}_t \equiv \tilde{e}_t$, $\tilde{\Psi}_t \equiv \nabla_\alpha \tilde{m}_t$, $\tilde{\Gamma}_t \equiv \tilde{\mathbf{W}}_t$, and $\tilde{\Lambda}_t \equiv \nabla_\theta \tilde{\mu}_t$; the mean and variance parameters need not be variation-free. However, choosing $\tilde{\eta}_t = \tilde{e}_t$ could (but need not) result in a loss of asymptotic local power if the mean and variance parameters are not variation free.

In certain cases, such as ARCH-M models or Amemiya's (1973) model of heteroskedasticity, misspecification of (2.1.b) leads to inconsistency of all elements in θ_o (if a normal likelihood is maximized). By the nature of these models most hypotheses concern jointly the conditional mean and conditional variance, and robustness to variance misspecification is not meaningful. Robustness to nonnormality is obtained by applying Procedure 3.1.

As mentioned earlier, a third possibility for an LM statistic is based on the weighted generalized residuals $\tilde{\Gamma}_t^{-1/2}\tilde{\eta}_t$. Under conditional normality it can be shown that $E(\eta_t^{o\prime}\Gamma_t^{o-1}\eta_t^o|x_t) = K+K(K+1)/2$. Along with standard LM theory using $\hat{\mathbf{A}}_T$ as an estimate of the information matrix, this fact can be used to show that $[K+K(K+1)/2]TR_u^2$ from the regression

$$\text{(3.9)} \qquad \tilde{\Gamma}_t^{-1/2}\tilde{\eta}_t \quad \text{on} \quad \tilde{\Gamma}_t^{-1/2}\tilde{\Lambda}_t, \quad t=1,\ldots,T$$

is asymptotically χ_Q^2 under H_0 and normality; the regression in (3.9) is carried out by stacking the observations and using OLS (see Engle (1982b) and Kroner (1987) for related results). Because this statistic employs an estimate of the Hessian as the estimated information matrix, we subsequently call it the HE LM statistic. For computing conditional mean tests, set $\tilde{\eta}_t \equiv \tilde{e}_t$, $\tilde{\Psi}_t \equiv \nabla_\alpha \tilde{m}_t$, $\tilde{\Gamma}_t \equiv \tilde{\mathbf{W}}_t$, $\tilde{\Lambda}_t \equiv \nabla_\theta \tilde{\mu}_t$, and then the statistic is KTR_u^2. But this test is not robust to second moment misspecification. If η_t consists only of squares and cross products of the elements of $\tilde{e}_t$ then the statistic reduces to $[K(K+1)/2]TR_u^2$. As evidenced by the simulations in section 4, the HE LM statistic is typically better behaved than the OPG LM statistic. Nonetheless, it is not asymptotically robust to nonnormality.

Procedure 3.1 is computationally somewhat more difficult than (3.6) or (3.9), but not by much. It requires exactly the same quantities used in computing the HE LM statistic and in implementing efficient computational algorithms. Our view is that the additional computational burden embodied in the matrix regression of step (ii) is warranted in many situations. Unless normality is a maintained assumption, one can never be sure about the asymptotic sizes of the nonrobust tests.

A useful extension of Procedure 3.1, which allows for a variety of other specification tests, is available from the results of Wooldridge (1990). In particular, there is no need to focus on tests that can be derived from nesting models. The matrix of unrestricted gradients evaluated at the restricted estimates,

$$\tilde{\Lambda}_t \equiv \begin{pmatrix} \nabla_\theta \tilde{\mu}_t \\ \nabla_\theta \tilde{\Omega}_t \end{pmatrix},$$

can be replaced in step (ii) of the robust LM procedure by essentially any function of x_t, $\tilde{\alpha}_T$, and other nuisance parameter estimates, say $\tilde{\gamma}_T$, such that $\sqrt{T}(\tilde{\gamma}_T - \gamma_T^o) = O_p(1)$ for some nonstochastic sequence $\{\gamma_T^o\}$. This extension allows for robust, regression-based nonnested hypothesis testing -- in which case $\tilde{\gamma}_T$ would be estimates from a competing model -- as well as many other useful diagnostics. For example, the diagnostics employed by Bollerslev, Engle, and Wooldridge (1988) for evaluating a dynamic capital asset pricing model (CAPM), which involve conducting LM tests for exclusion of fitted values from competing models, can easily be "robustified" within this framework.

4. SIMULATION EXPERIMENTS

In order to investigate the finite sample performance and potential applicability of the robust inference procedures discussed above, a small simulation experiment was performed. To facilitate the presentation, all the simulated models are nested within the AR(2)-GARCH(1,2) model,

$$\begin{aligned} y_t &= \phi_{o1} y_{t-1} + \phi_{o2} y_{t-2} + \varepsilon_t \\ \omega_t^2 &= \delta_o + \alpha_{o1}\varepsilon_{t-1}^2 + \alpha_{o2}\varepsilon_{t-2}^2 + \beta_{o1}\omega_{t-1}^2, \quad t=1,...,T \\ \varepsilon_t &= \omega_t \xi_t, \quad \xi_t \text{ i.i.d. } t_\nu, \end{aligned} \tag{4.1}$$

where t_ν denotes a standard t-distribution with ν degrees of freedom.

The derivatives for the conditional mean and the conditional variance function for the AR(2)-GARCH(1,2) model are given in Appendix B. Under the assumption of conditional normal errors, i.e. $\nu = \infty$, the score and the information matrix for the quasi-log likelihood function in (2.4) follow by direct substitution from (2.5) and (2.7), and the results in section 2 allow asymptotically valid inference about the true parameter vector, θ_o, to be carried out from (2.8) based on the QMLE, $\hat{\theta}_T$. The use of this robust form of the Wald test when conducting inference in ARCH models has previously been suggested by Weiss (1984, 1986); however, no evidence on the small sample performance of the procedure is yet available. Robust LM tests for hypothesis about θ_o based on the constrained QMLE, $\tilde{\theta}_T$, can be calculated from the regressions outline in Procedure 3.1. The finite sample properties of these LM tests are also unknown.

The different parameter sets characterizing the simulated models and the sample mean for the QMLE obtained under the auxiliary assumption of conditional normality and various model specifications, together with the corresponding sample standard deviations in parentheses, are given in Table 1. The normally distributed random variables were generated by the IMSL subroutine GGNML. The t_{ν} distributed random variables were formed as $(\nu-2)$ times a $N(0,1)$ random variable divided by the square root of χ^2_{ν} variate generated by the IMSL subroutine GGAMR. All results for the AR(1) models estimated under the assumption of conditional homoskedasticity reported in Table 1 and throughout are based on 10,000 replications. Due to the computer intensive nature of the estimation, the results for the estimated AR(1)-GARCH(1,1) models are calculated from 1,000 replications only. More specifically, the QMLE for $\theta_o = (\phi_{o1},\delta_o,\alpha_{o1},\beta_{o1})$ was found through a combined grid search and a standard iterative procedure based on the Berndt, Hall, Hall, and Hausman (1974) (BHHH) algorithm. The convergence criterion was taken as an r-squared less than .001 in the BHHH updating regression. Some informal analysis suggested that very similar results would be obtained using a more stringent convergence criterion. Also, to avoid startup problems the first 100 observations were discarded in each replication.

From Table 1, the well known small sample bias in the estimates for ϕ_{o1} is generally of relatively minor order in the present context with 100 or more observations, although both the bias and the variability of the estimates tend to increase with the degree of heteroskedasticity and/or conditional leptokurtosis. Also, with $\phi_{o2} \neq 0$ the bias in the estimate for

ϕ_{o1} from the misspecified-in-mean AR(1) models are of the same order of magnitude as the true first order autocorrelation coefficient, $\phi_{o1}/(1-\phi_{o2})$. Very little evidence is available on the small sample properties of estimators and test statistics from ARCH models. To get an idea about the gain in efficiency from correctly modelling the form of the heteroskedasticity, consider the results from the AR(1)-GARCH(1,1) model with $\phi_{o1} = .5$, $\delta_o = .05$, $\alpha_{o1} = .15$, $\beta_{o1} = .8$, and T = 200. The sample mean and standard deviation for the QMLE for ϕ_{o1} across the different replications are .493 (.066) when correctly modeling both the mean and the variance functions, compared to .493 (.075) from the AR(1) model estimated under the assumption of conditional homoskedasticity. Interestingly, Table 1 also suggests that the MLE for α_{o1} is slightly upward biased, and the MLE for β_{o1} downward biased, leaving $\alpha_{o1} + \beta_{o1}$ slightly downward biased in small samples. As shown in Bollerslev (1988), the GARCH(1,1) model is readily interpreted as an ARMA(1,1) model for conditional second order moments with autoregressive parameter $\alpha_{o1} + \beta_{o1}$ and moving average parameter $-\beta_{o1}$, respectively. Therefore, in the ARMA formulation, both parameters show a bias towards zero. This is also consistent with the small sample bias in the MLE for α_{o1} in the simple ARCH(1) model reported in Engle, Hendry,and Trumble (1985). As with the conditional mean parameters, the variability in the QMLE for the conditional variance parameters increases with departures from conditional normality.

The small sample behavior of the QMLE's for ϕ_{o1} are further illustrated in Table 2, where the acceptance probabilities are listed for three different Wald statistics when testing a true null hypothesis for ϕ_{o1}. The RB statistic refers to the robust form given in equation (2.8). This form of the test is compared to the standard Hessian (HE) based Wald test, which relies on the inverse of the quasi-information matrix, $\hat{\mathbf{A}}_T^{-1}/T$, as an estimate of the variance for $\hat{\phi}_{T1}$, and the OPG test which uses the matrix $\hat{\mathbf{B}}_T^{-1}/T$. Both the HE and OPG tests are used regularly in the literature. For each model, Table 2 reports the proportion of the replications that fall below the .900, .950 and .990 fractiles in the chi-squared distribution. With conditionally homoskedastic errors the different tests are asymptotically equivalent, and from the table the actual size is very close to the nominal size for all three tests. However, with neglected conditional heteroskedasticity the RB test is clearly preferred. Both the HE and OPG estimators systematically underestimate the standard errors, resulting in the empirical size of the tests being much larger than the nominal size. Bias correcting the estimators for ϕ_{o1} does not alter any of these conclusions; the results for the bias corrected test statistics are all within .005 of the results reported in Table 2. When correctly modelling the heteroskedasticity all three t-statistics for the AR(1) parameter perform reasonably well, although the presence of conditionally leptokurtic errors introduces some minor biases for both the HE and OPG tests.

By contrast, when testing a true null hypothesis about the conditional variance parameter α_{o1}, the three covariance matrix estimators lead to very different results. From Table 3, with conditional normal errors but only 200 observations, the actual size of the RB test is much closer to the nominal size than the sizes for the HE and the OPG tests. Moreover, with conditional t_5 distributed errors it is clear that the asymptotic size of both the HE and the OPG Wald tests far exceed the nominal size. For instance, for the AR(1)-GARCH(1,1) model with 400 observations, a nominal five percent OPG test for the true value of α_{o1} results in a probability of a type I error of 21.1 percent.

Having estimated a constrained model, it is desirable to test for deviations from that specification. Table 4 reports the simulation results for the three different LM-type test statistics discussed in section 3 designed to test for additional serial correlation; specifically, $\phi_{o2} = 0$ versus $\phi_{o2} \neq 0$. The two regressions for the RB test can be deduced from Procedure 3.1. Note that the second regression from Procedure 3.1 contains perfect multicollinearity; therefore, in calculating R_u^2, only the last element in $\ddot{\eta}_t'\ddot{\Lambda}_t$ is used. This corresponds to the only nonredundant element. In particular, for the estimated homoskedastic AR(1) models, R_u^2 is simply obtained from a regression of 1 on $\tilde{e}_t\ddot{e}_{t-1}$, where $\tilde{e}_t$ is the residual from the AR(1) regression and $\ddot{e}_{t-1}$ denotes the residual from the regression of $\tilde{e}_{t-1}$ on y_{t-1}. In implementing the TR_u^2 test statistic used throughout this section we replace T with the actual number of observations in the auxiliary regression, here T-1. The HE LM test given by the regression in (3.9) that exploits block diagonality of the information matrix for AR(2) errors is readily evaluated as TR_u^2 from the regression of $\tilde{e}_t\tilde{\omega}_t^{-1}$ on $y_{t-1}\tilde{\omega}_t^{-1}$ and $\tilde{e}_{t-1}\tilde{\omega}_t^{-1}$. With conditional homoskedasticity this is equivalent to TR_u^2 from the regression of $\tilde{e}_t$ on y_{t-1} and $\tilde{e}_{t-1}$. Finally, the OPG statistic is given by TR_u^2 from the regression in (3.6) of 1 on the quasi-score evaluated at the estimates under the null, i.e. $\mathbf{s}_t(\tilde{\theta}_T)$. For the homoskedastic AR(1) models this is equivalent to a regression of 1 on $\tilde{e}_t y_{t-1}$ and $\tilde{e}_t\tilde{e}_{t-1}$. All three tests extend in a straightforward way to higher orders of serial correlation by including additional lags of $\tilde{e}_t$ or the relevant cross products in the regressions.

From Table 4 the finite sample null and asymptotic chi-square distribution are close for all three tests under ideal conditions, i.e. under homoskedasticity or correctly modelled conditional heteroskedasticity. The power properties are also similar in those situations. However, in accordance with the findings in Diebold (1986) and Domowitz and Hakkio (1988), the actual size of both the HE and the OPG tests is much higher than the nominal size with neglected heteroskedasticity, and this effect is magnified by conditional leptokurtosis. In contrast, the RB LM test is robust to heteroskedasticity.

The next set of results relates to the performance of the same three LM tests when testing for ARCH(1) errors in the homoskedastic AR(1) model, i.e. $\alpha_{o1} = 0$ versus $\alpha_{o1} > 0$. In this situation the second regression for the RB test described in Procedure 3.1 is simply equal to a regression of 1 on $(\tilde{e}_t^2 - \tilde{\delta})(\tilde{e}_{t-1}^2 - \tilde{\delta})$. The "studentized" HE version of the Breusch and Pagan (1979) and Godfrey (1978) LM statistic, as advocated by Engle (1984), Hall (1984), and Koenker (1981), is also easy to compute. Following Engle (1982b), the test for first order ARCH is given by TR_c^2 from the regression of $\tilde{e}_t^2$ on 1 and $\tilde{e}_{t-1}^2$, where R_c^2 is the centered r-squared. Finally, the OPG test for ARCH(1) takes the form TR_u^2 from the regression of 1 on $(\tilde{e}_t^2 - \tilde{\delta})$ and $\tilde{e}_{t-1}^2(\tilde{e}_t^2 - \tilde{\delta})$. Again, all three tests extend readily to checking for higher orders of ARCH by including additional lags in the auxiliary regressions. As can be seen from Table 5, the results for the RB test are less encouraging in this situation. Although the widely used HE test is conservative, it clearly outperforms both the RB and the OPG tests in terms of power. Table 5 also suggests that the RB test is not asymptotically equivalent to the HE and OPG tests against nonlocal alternatives. These results are in accordance with the findings in Engle, Hendry, and Trumble (1985), where a one sided version of the HE test is compared to a modified Wald-type test from the auxiliary HE regression. See also Bollerslev (1988) and Milhøj (1987). We should emphasize that we have not presented evidence on the behavior of the three statistics for ARCH(1) when conditional homoskedasticity holds, but the conditional fourth moment is nonconstant. In this situation the RB test will retain the appropriate size while the HE and the OPG tests will generally have the wrong size. We conjecture that this carries over to finite sample properties of the tests as well.

The final set of results in Table 6 gives the distribution of the LM test for the AR(1)-GARCH(1,1) versus AR(1)-GARCH(1,2) model, i.e. $\alpha_{o2} = 0$ versus $\alpha_{o2} > 0$. In this situation no simple form of the HE test is readily available, as the regression in (3.9) involves the derivative of the conditional variance, which takes a recursive form; see Appendix B. From the table, the actual size of both the RB and the HE test is in accordance with the nominal size, whereas the OPG test rejects far too often. The powers of the RB and HE tests are also very similar. Since the calculations required for the HE test are more involved in this situation, a related test computed as TR_u^2 from the regression of $(\tilde{e}_t^2\tilde{\omega}_t^{-2} - 1)$ on $\tilde{e}_{t-1}^2\tilde{\omega}_t^{-2}$ has often been used in practice. This simpler, residual-based diagnostic, which is equivalent to the test obtained by evaluating the derivatives in (3.9) at $\beta_{o1} = 0$ and ignoring terms for the derivatives of the conditional variance function with respect to δ_o, α_{o1}, and β_{o1}, leads to a conservative test. For instance, for $\phi_{o1} = .5$, $\delta_o = .05$, $\alpha_{o1} = .15$, $\beta_{o1} = .8$, $\nu = \infty$, and $T = 200$, the actual size corresponding to a nominal 5 % test is estimated to be 0.5 %.

The simulation results in Tables 1-6 generally support the use of the robust test statistics. Nevertheless, these simulations are limited to symmetric error distributions. It is of interest to know how the RB, HE, and OPG statistics behave when the error distribution is asymmetric. To shed some light on this issue, we carried out some additional simulations in which the conditional error distribution in (4.1) is a standardized chi-square distribution, i.e. $\varepsilon_t = \omega_t((\xi_t-1)/\sqrt{2})$, where the ξ_t are i.i.d. χ_1^2 variates. This error distribution represents a marked departure from conditional normality with the coefficients of skewness and kurtosis equal to $2\sqrt{2}$ and 12, respectively. In terms of ranking the tests, the results of the additional simulations are broadly consistent with those reported in Tables 1-6 pertaining to symetrically t-distributed errors. For instance, for the estimated AR(1)-GARCH(1,1) model with $\phi_{o1} = .5$, $\delta_o = .05$, $\alpha_{o1} = .15$, $\beta_{o1} = .8$, and T = 200, the acceptance probabilities using the nominal 5 percent Wald tests for the true null hypothesis $\phi_{o1} = .5$ (as in Table 2) are .939, .928, and .915 for the RB, HE, and OPG tests, respectively. The same three acceptance probabilities for the Wald test of $\alpha_{o1} = .15$ (as in Table 3) equal .901, .827, and .701. Note that the performance of all three tests deteriorates relative to the symmetric t-distributed case, but the robust test continues to have empirical size much closer to the nominal size. The findings for the three different LM procedures in testing for additional serial correlation or ARCH effects also tend to confirm the superiority of the robust procedures. Again, with a nominal 5 percent size the rejection probabilities for $\phi_{o2} = 0$ (as in Table 4) are .041, .049, and .118; the actual size of a 5 percent test of $\alpha_{o2} = 0$ (as in Table 6) are .038, .016, and .312, = .5, respectively. Similar patterns emerge for other fractiles and other parameter values. Further details of the experiments are available from the authors on request.

5. STOCK RETURN VOLATILITY

As noted earlier, high frequency financial time series are typically characterized by volatility clustering. To empirically illustrate the relevance of robust inference procedures in this context, we estimate a simple conditional normal GARCH(1,1) model for the monthly percentage return on the CRSP value-weighted index including dividends. The data extend from 1950.1 through 1987.12 for a total of 456 observations. The estimated equations are

$$R_t = \underset{\substack{(.190)\\ [.188]}}{1.034} + \hat{\varepsilon}_t$$

(5.1)

$$\hat{\omega}_t^2 = \underset{\substack{(.915)\\ [.491]}}{.955} + \underset{\substack{(.033)\\ [.033]}}{.089}\, \hat{\varepsilon}_{t-1}^2 + \underset{\substack{(.078)\\ [.043]}}{.864}\, \hat{\omega}_{t-1}^2 .$$

The numbers in (·) underneath the parameter estimates are the robust (RB) standard errors from (2.8), and the numbers in [·] give the standard errors calculated from the outer product of the gradient (OPG).

In accordance with previous empirical findings by Chou (1988), Nelson (1990b), Pagan and Schwert (1990), and many others, the parameter estimates in (5.1) suggest a high degree of persistence in the conditional variance. While these QMLEs were obtained under the auxiliary assumption of conditional normality, this assumption is clearly violated empirically. In particular, the sample kurtosis for the standardized residuals, $\hat{\varepsilon}_t \hat{\omega}_t^{-1}$, equals 4.633, which is highly significant at virtually any level in the corresponding asymptotic normal distribution. This violation of conditional normality in the standardized residual from estimated ARCH models is not peculiar to the current sample, but is apparent with most financial time series; see, e.g., Bollerslev (1987). Thus, as predicted by the simulation results in the previous section, the OPG standard errors tend to under-estimate the true parameter estimator uncertainty. Similarly, the RB LM tests for misspecification are generally lower than the corresponding OPG type LM tests. For instance, when testing for AR(1) disturbances the RB test statistic equals 1.156 compared with 1.377 for the OPG test, whereas the two test statistics for additional ARCH effects in the form of a GARCH(1,2) model equal 1.785 and 1.934, respectively.

Most asset pricing theories postulate a positive relationship between the expected return and risk. The exact form and significance of this risk-return tradeoff have been the subject of extensive empirical investigations; for a recent survey of this literature using the ARCH-in-mean model see Bollerslev, Chou, and Kroner (1990). Testing for inclusion of the conditional standard deviation in the conditional mean equation of (5.1) yields an RB LM statistic of 2.935, while the OPG LM statistic of 3.716 is borderline significant at the usual five percent level in the χ_1^2 distribution. As emphasized in section 3, with nonnormal errors the OPG test is not necessarily asymptotically chi-square distributed and -- as the simulations in section 4 indicate -- the marginal significance of the OPG LM statistic may be spurious. In order to perform a one-sided Wald test on the standard deviation term in the mean, we estimate the corresponding GARCH(1,1)-M model by QMLE. The result is

$$R_t = \underset{\substack{(1.509) \\ [1.369]}}{-1.690} + \underset{\substack{(.389) \\ [.340]}}{.692}\,\hat{\omega}_t + \hat{\varepsilon}_t$$

(5.2)

$$\hat{\omega}_t^2 = \underset{\substack{(.744)\\[.560]}}{1.005} + \underset{\substack{(.029)\\[.032]}}{.085}\,\hat{\varepsilon}_{t-1}^2 + \underset{\substack{(.065)\\[.046]}}{.863}\,\hat{\omega}_{t-1}^2.$$

The asymptotically justified RB t-statistic for testing the mean-standard deviation tradeoff equals 1.779, while the OPG t-statistic is notably higher at 2.035. For a one-sided test, the RB statistic provides only marginal evidence of a positive mean-variance relationship, while the OPG statistic is well above the 5% critical value in the standard normal distribution. Using a similar formulation and data set, Baillie and DeGennaro (1990) have also noted that the inference about the ARCH-in-mean parameter is sensitive to distributional assumptions. Following Bollerslev (1987), once the assumption of conditional normality is replaced with a parametrically estimated t-distribution, the ARCH-M parameter tends to become insignificant; see also Gallant, Rossi, and Tauchen (1990). We shall not pursue this example further here; it is intended to illustrate the potential importance of relying on robust inference procedures. In fact, the robust methods put forth here have already been successfully employed in modelling other financial time series by Baillie and Bollerslev (1990) and McCurdy and Morgan (1990).

6. CONCLUSION

We have shown that, for a general class of dynamic models parameterized by the first and second moments, the normal QMLE is consistent and asymptotically normal under fairly weak regularity conditions. Building on the results of Wooldridge (1990), we have offered simple formulas for the corresponding robust standard errors and robust, regression-based LM procedures. A Monte Carlo study for a set of univariate AR time series models with GARCH errors indicates that these asymptotically justified results carry over to finite samples. For the sample sizes analyzed here the biases in the QMLE are relatively minor. Nevertheless, the choice of covariance matrix estimator plays an important role when conducting inference. Wald tests based on estimates of the quasi-information matrix or the outer product of the score often lead to inference with the wrong size. In contrast, the actual size of the robust Wald test, derived from a relatively simple estimate of the White (1982a) covariance matrix estimator, is never very far from the nominal size; this is true whether or not the auxiliary assumptions hold. Moreover, in most situations the actual size and power properties of the robust LM procedure compare favorably to the more traditional LM tests. The LM test constructed from the regression of unity on the score has particularly poor finite sample properties.

Because the QMLE is not asymptotically efficient under nonnormality, future research could profitably focus on efficient method of moments estimators. Engle and Gonzalez-Rivera (1990) recently provide evidence on the loss of efficiency of the QMLE under nonnormality, and compare the QMLE to the MLE and a seminonparametric procedure. Generally -- and as supported by the simulations in section 4 -- the QMLE loses little efficiency with symmetrically t-distributed errors, but the efficiency loss can be marked under asymmetric error distributions.

Table 1

Quasi-Maximum Likelihood Estimates

$$y_t = \phi_{o1} y_{t-1} + \phi_{o2} y_{t-2} + \varepsilon_t$$
$$\omega_t^2 = \delta_o + \alpha_{o1}\varepsilon_{t-1}^2 + \alpha_{o2}\varepsilon_{t-2}^2 + \beta_{o1}\omega_{t-1}^2, \quad t=1,2,...,T$$
$$\varepsilon_t = \omega_t \xi_t, \ \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	ϕ_{o2}	δ_o	α_{o1}	α_{o2}	β_{o1}	ν	T	$\hat{\phi}_{T1}$	$\hat{\delta}_T$	$\hat{\alpha}_{T1}$	$\hat{\beta}_{T1}$
.5	.00	1.00	.00	.0	.0	∞	100	.492	.991	–	–
+	–	+	–	–	–			(.086)	(.141)		
.5	.00	1.00	.00	.0	.0	5	100	.490	.992	–	–
+	–	+	–	–	–			(.086)	(.266)		
.5	.00	1.00	.00	.0	.0	5	200	.495	.994	–	–
+	–	+	–	–	–			(.060)	(.188)		
.5	.00	.05	.15	.0	.8	5	100	.486	.936	–	–
+	–	+	–	–	–			(.105)	(1.024)		
.5	.15	1.00	.00	.0	.0	∞	100	.572	1.011	–	–
+	–	+	–	–	–			(.095)	(.147)		
.5	.15	.05	.15	.0	.8	5	100	.568	.992	–	–
+	–	+	–	–	–			(.114)	(1.391)		
.5	.00	.60	.40	.0	.0	∞	100	.484	.980	–	–
+	–	–	–	–	–			(.114)	(.270)		
.5	.00	.60	.40	.0	.0	5	100	.479	.971	–	–
+	–	–	–	–	–			(.128)	(.550)		
.5	.00	.05	.15	.0	.8	∞	200	.493	.085	.154	.750
+	–	+	+	–	+			(.066)	(.070)	(.061)	(.112)
.5	.00	.05	.15	.0	.8	5	200	.489	.078	.158	.742
+	–	+	+	–	+			(.071)	(.072)	(.075)	(.127)
.5	.00	.05	.15	.0	.8	5	400	.497	.069	.156	.762
+	–	+	+	–	+			(.052)	(.046)	(.059)	(.091)
.5	.15	.05	.15	.0	.8	∞	200	.579	.096	.160	.733
+	–	+	+	–	+			(.068)	(.089)	(.069)	(.138)

(Table 1 continued)

.5	.15	.05	.15	.0	.8	5	200	.572	.083	.162	.731
+	−	+	+	−	+			(.076)	(.066)	(.081)	(.140)
.5	.00	.10	.10	.2	.6	∞	200	.491	.103	.183	.686
+	−	+	+	−	+			(.063)	(.050)	(.067)	(.090)
.5	.00	.10	.10	.2	.6	5	200	.494	.106	.173	.666
+	−	+	+	−	+			(.070)	(.057)	(.079)	(.112)

Key: QMLE sample mean parameter estimates obtained under the assumption of conditional normality, with sample standard deviations in parenthesis. The '+' and '−' signs underneath the true coefficients refer to the model parameters being estimated and fixed at zero, respectively.

Table 2

Wald Test for $\phi_{o1} = .5$

$$y_t = \phi_{o1} y_{t-1} + \varepsilon_t$$

$$\omega_t^2 = \delta_o + \alpha_{o1}\varepsilon_{t-1}^2 + \beta_{o1}\omega_{t-1}^2, \quad t=1,2,\ldots,T$$

$$\varepsilon_t = \omega_t \xi_t, \; \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	δ_o	α_{o1}	β_{o1}	ν	T		.900	.950	.990
						RB	.889	.940	.986
.5	1.00	.00	.0	∞	100	HE	.897	.950	.991
+	+	−	−			OPG	.904	.954	.991
						RB	.885	.939	.985
.5	1.00	.00	.0	5	100	HE	.904	.951	.988
+	+	−	−			OPG	.906	.951	.984
						RB	.893	.945	.988
.5	1.00	.00	.0	5	200	HE	.905	.954	.991
+	+	−	−			OPG	.908	.953	.988
						RB	.880	.931	.980
.5	.05	.15	.8	5	100	HE	.834	.900	.965
+	+	−	−			OPG	.773	.844	.925
						RB	.871	.929	.979
.5	.60	.40	.0	∞	100	HE	.800	.865	.948
+	+	−	−			OPG	.708	.782	.883
						RB	.859	.915	.974
.5	.60	.40	.0	5	100	HE	.762	.835	.910
+	+	−	−			OPG	.663	.738	.834
						RB	.883	.952	.991
.5	.05	.15	.8	∞	200	HE	.882	.956	.991
+	+	+	+			OPG	.893	.959	.990
						RB	.891	.954	.991
.5	.05	.15	.8	5	200	HE	.879	.946	.991
+	+	+	+			OPG	.886	.942	.987
						RB	.895	.947	.987
.5	.05	.15	.8	5	400	HE	.882	.938	.988
+	+	+	+			OPG	.875	.935	.983

Key: .900, .950 and .990 give the empirical distribution based on the corresponding nominal fractiles in the chi-squared distribution. The '+' and '−' signs underneath the true parameter values refer to the parameters being estimated and fixed at zero, respectively, in the model under the null. RB denotes the robust test, HE the test based on the Hessian, and OPG the test calculated from the outer product of the gradient.

Table 3

Wald Test for $\alpha_{o1} = .15$

$$y_t = \phi_{o1} y_{t-1} + \varepsilon_t$$
$$\omega_t^2 = \delta_o + \alpha_{o1}\varepsilon_{t-1}^2 + \beta_{o1}\omega_{t-1}^2, \quad t=1,2,...,T$$
$$\varepsilon_t = \omega_t \xi_t, \ \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	δ_o	α_{o1}	β_{o1}	ν	T		.900	.950	.990
						RB	.918	.954	.983
.5	.05	.15	.8	∞	200	HE	.936	.970	.990
+	+	+	+			OPG	.955	.974	.991
						RB	.923	.952	.984
.5	.05	.15	.8	5	200	HE	.884	.932	.970
+	+	+	+			OPG	.829	.884	.944
						RB	.909	.941	.969
.5	.05	.15	.8	5	400	HE	.824	.886	.945
+	+	+	+			OPG	.699	.789	.887

Key: See Table 2.

Table 4

LM Test for $\phi_{o2} = 0$

$$y_t = \phi_{o1} y_{t-1} + \phi_{o2} y_{t-2} + \varepsilon_t$$
$$\omega_t^2 = \delta_o + \alpha_{o1} \varepsilon_{t-1}^2 + \beta_{o1} \omega_{t-1}^2, \quad t=1,2,\ldots,T$$
$$\varepsilon_t = \omega_t \xi_t, \; \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	ϕ_{o2}	δ_o	α_{o1}	β_{o1}	ν	T		.900	.950	.990
							RB	.897	.951	.992
.5	.00	1.00	.00	.0	∞	100	HE	.901	.949	.992
+	−	+	−	−			OPG	.892	.948	.991
							RB	.899	.950	.992
.5	.00	1.00	.00	.0	5	100	HE	.904	.954	.992
+	−	+	−	−			OPG	.890	.944	.991
							RB	.898	.957	.995
.5	.00	.05	.15	.8	5	100	HE	.834	.900	.967
+	−	+	−	−			OPG	.885	.949	.993
							RB	.613	.734	.905
.5	.15	1.00	.00	.0	∞	100	HE	.613	.731	.895
+	−	+	−	−			OPG	.605	.726	.902
							RB	.680	.794	.939
.5	.15	.05	.15	.8	5	100	HE	.599	.704	.865
+	−	+	−	−			OPG	.663	.779	.929
							RB	.895	.951	.991
.5	.00	.60	.40	.0	∞	100	HE	.840	.908	.972
+	−	+	−	−			OPG	.879	.940	.989
							RB	.895	.952	.993
.5	.00	.60	.40	.0	5	100	HE	.824	.897	.964
+	−	+	−	−			OPG	.865	.935	.987
							RB	.902	.953	.991
.5	.00	.05	.15	.8	∞	200	HE	.895	.951	.989
+	−	+	+	+			OPG	.878	.941	.982
							RB	.909	.958	.992
.5	.00	.05	.15	.8	5	200	HE	.886	.952	.991
+	−	+	+	+			OPG	.861	.925	.980

(Table 4 continued)

							RB	.390	.528	.775
.5	.15	.05	.15	.8	ν	200	HE	.400	.529	.759
+	−	+	+	+			OPG	.350	.481	.688
							RB	.442	.578	.803
.5	.15	.05	.15	.8	5	200	HE	.400	.516	.781
+	−	+	+	+			OPG	.389	.506	.707

Key: See Table 2.

Table 5

LM Test for $\alpha_{o1} = 0$

$$y_t = \phi_{o1} y_{t-1} + \varepsilon_t$$
$$\omega_t^2 = \delta_o + \alpha_{o1} \varepsilon_{t-1}^2, \quad t=1,2,...,T$$
$$\varepsilon_t = \omega_t \xi_t, \ \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	δ_o	α_{o1}	ν	T		.900	.950	.990
					RB	.860	.922	.986
.5	1.00	.00	∞	100	HE	.924	.965	.994
+	+	−			OPG	.828	.893	.965
					RB	.855	.931	.993
.5	1.00	.00	5	100	HE	.950	.975	.990
+	+	−			OPG	.757	.846	.945
					RB	.525	.742	.954
.5	.60	.40	∞	100	HE	.308	.387	.548
+	+	−			OPG	.347	.493	.777
					RB	.746	.891	.991
.5	.60	.40	5	100	HE	.501	.570	.702
+	+	−			OPG	.566	.703	.891

Key: See Table 2.

Table 6

LM Test for $\alpha_{o2} = 0$

$$y_t = \phi_{o1} y_{t-1} + \varepsilon_t$$
$$\omega_t^2 = \delta_o + \alpha_{o1}\varepsilon_{t-1}^2 + \alpha_{o2}\varepsilon_{t-1}^2 + \beta_{o1}\omega_{t-1}^2, \quad t=1,2,...,T$$
$$\varepsilon_t = \omega_t \xi_t, \ \xi_t \text{ i.i.d. } t_\nu$$

ϕ_{o1}	δ_o	α_{o1}	α_{o2}	β_{o1}	ν	T		.900	.950	.990
							RB	.882	.948	.992
.5	.05	.15	.0	.8	∞	200	HE	.895	.952	.993
+	+	+	−	+			OPG	.845	.906	.972
							RB	.902	.954	.995
.5	.05	.15	.0	.8	5	200	HE	.899	.957	.994
+	+	+	−	+			OPG	.770	.839	.937
							RB	.586	.722	.911
.5	.10	.10	.2	.6	∞	200	HE	.588	.692	.873
+	+	+	−	+			OPG	.518	.631	.823
							RB	.713	.822	.970
.5	.10	.10	.2	.6	5	200	HE	.734	.823	.961
+	+	+	−	+			OPG	.587	.680	.819

Key: See Table 2.

APPENDIX A

Conventions: If $\mathbf{a}(\theta)$ is an N x 1 vector depending on the P x 1 vector θ then the derivative of $\mathbf{a}$ with respect to θ, denoted $\nabla_\theta \mathbf{a}(\theta)$, is an N x P matrix. If $\mathbf{A}(\theta)$ is an N x M matrix, the derivative of $\mathbf{A}$ with respect to θ, denoted $\nabla_\theta \mathbf{A}(\theta)$, is the NM x P matrix

$$\nabla_\theta \mathbf{A}(\theta) \equiv \nabla_\theta \{\text{vec } \mathbf{A}(\theta)\}.$$

For any twice continuously differentiable L x 1 function $\mathbf{b}(\theta)$, define the second derivative of $\mathbf{b}$ to be the LP x P matrix

$$\nabla_\theta^2 \mathbf{b}(\theta) \equiv \nabla_\theta \{\nabla_\theta \mathbf{b}(\theta)\}.$$

LEMMA A.1: Let $\mathbf{A}$ be a K x K positive definite matrix. Then

$$\log |\mathbf{A}| \le \text{tr } (\mathbf{A} - \mathbf{I}_K)$$

with equality holding if and only if $\mathbf{A} = \mathbf{I}_K$.

PROOF: See Magnus and Neudecker (1988, Theorem 27).

LEMMA A.2: Let y be a K x 1 random vector with finite second moments, and let $\mu_o \equiv E(y)$, $\Sigma_o \equiv V(y)$. Assume that Σ_o is positive definite. Define the functions

$$q(y;\mu,\Sigma) \equiv \log |\Sigma| + (y - \mu)'\Sigma^{-1}(y - \mu)$$

$$\bar{q}(\mu,\Sigma) \equiv E[q(y;\mu,\Sigma)]$$

for $\mu \in \mathbb{R}^K$, Σ a positive definite K x K matrix. Then $\bar{q}$ is uniquely minimized by (μ_o,Σ_o).

PROOF: Straightforward algebra shows that

$$\bar{q}(\mu,\Sigma) = \log |\Sigma| + \text{tr } \Sigma^{-1}\Sigma_o + (\mu_o - \mu)'\Sigma^{-1}(\mu_o - \mu).$$

Therefore, $\bar{q}(\boldsymbol{\mu},\Sigma) > \bar{q}(\boldsymbol{\mu}_o,\Sigma)$ for any $\boldsymbol{\mu} \neq \boldsymbol{\mu}_o$ and any p.d. matrix Σ. It remains to show that $\bar{q}(\boldsymbol{\mu}_o,\Sigma) > \bar{q}(\boldsymbol{\mu}_o,\Sigma_o)$ for any p.d. matrix $\Sigma \neq \Sigma_o$, i.e.

$$\log|\Sigma| + \text{tr}\,\Sigma^{-1}\Sigma_o > \log|\Sigma_o| + \text{tr}\,\mathbf{I}_K$$

or

$$\log|\Sigma^{-1}\Sigma_o| < \text{tr}\,(\Sigma^{-1}\Sigma_o - \mathbf{I}_K).$$

But this follows from Lemma A.1 by setting $\mathbf{A} \equiv \Sigma^{-1/2}\Sigma_o\Sigma^{-1/2}$ and using the commutativity of the determinant and trace operators. ∎

CONDITIONS A.1:

(i) Θ is compact and has nonempty interior; $\boldsymbol{\theta}_o \in \text{int}\,\Theta$.

(ii) $\boldsymbol{\mu}_t(\cdot,\boldsymbol{\theta})$ and $\Omega_t(\cdot,\boldsymbol{\theta})$ are measurable for all $\boldsymbol{\theta} \in \Theta$, and $\boldsymbol{\mu}_t(x_t,\cdot)$ and $\Omega_t(x_t,\cdot)$ are twice continuously differentiable on int Θ for all x_t. $\Omega_t(x_t,\boldsymbol{\theta})$ is nonsingular with probability one, for all $\boldsymbol{\theta} \in \Theta$.

(iii) (a) $\{\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_o)\colon t=1,2,...\}$ satisfies the UWLLN (see Wooldridge (1990, Definition A.1)).

(b) $\boldsymbol{\theta}_o$ is the identifiably unique maximizer (see Bates and White (1985)) of

$$T^{-1}\sum_{t=1}^{T} E[\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_o)].$$

(iv) (a) $\{\mathbf{h}_t(\boldsymbol{\theta}_o)\}$ and $\{\mathbf{a}_t(\boldsymbol{\theta}_o)\}$ satisfy the WLLN.

(b) $\{\mathbf{h}_t(\boldsymbol{\theta}) - \mathbf{h}_t(\boldsymbol{\theta}_o)\}$ satisfies the UWLLN.

(c) $\{\mathbf{A}_T^o \equiv T^{-1}\sum_{t=1}^{T} E[\mathbf{a}_t(\boldsymbol{\theta}_o)]\}$ is uniformly positive definite.

(v) (a) $\{\mathbf{s}_t(\boldsymbol{\theta}_o)'\mathbf{s}_t(\boldsymbol{\theta}_o)\}$ satisfies the WLLN.

(b) $\{\mathbf{B}_T^o \equiv T^{-1}\sum_{t=1}^{T} E[\mathbf{s}_t(\boldsymbol{\theta}_o)'\mathbf{s}_t(\boldsymbol{\theta}_o)]\}$ is uniformly p.d.

(c) $\mathbf{B}_T^{o-1/2}T^{-1/2}\sum_{t=1}^{T}\mathbf{s}_t(\boldsymbol{\theta}_o)' \xrightarrow{d} N(\mathbf{0},\mathbf{I}_P)$.

(vi) (a) $\{\mathbf{a}_t(\theta) - \mathbf{a}_t(\theta_o)\}$ satisfies the UWLLN.

(b) $\{\mathbf{s}_t(\theta)'\mathbf{s}_t(\theta) - \mathbf{s}_t(\theta_o)'\mathbf{s}_t(\theta_o)\}$ satisfies the UWLLN.

PROOF OF THEOREM 2.1: First, application of Lemma A.2 demonstrates that θ_o is a maximizer of $E[\ell_t(\theta)|x_t]$ for all x_t, t=1,2,.... Consequently, θ_o is a maximizer of

$$T^{-1}\sum_{t=1}^{T} E[\ell_t(\theta) - \ell_t(\theta_o)].$$

Following standard practice, we strengthen this conclusion by assuming that θ_o is identifiability unique. This, combined with the assumption that $\{\ell_t(\theta) - \ell_t(\theta_o)\}$ satisfies the UWLLN, establishes the weak consistency of the QMLE under (i), (ii), and (iii) ((i) and (ii) are actually much stronger than needed for consistency). Next, the score is seen to be

$$\mathbf{s}_t(\theta)' = \nabla_\theta\mu_t(\theta)'\Omega_t^{-1}(\theta)\boldsymbol{\varepsilon}_t(\theta)$$
$$+ 1/2\ \nabla_\theta\Omega_t(\theta)'[\Omega_t^{-1}(\theta) \otimes \Omega_t^{-1}(\theta)]\text{vec}[\boldsymbol{\varepsilon}_t(\theta)\boldsymbol{\varepsilon}_t(\theta)' - \Omega_t(\theta)].$$

Differentiation shows that the Hessian of ℓ_t can be expressed as

$$\mathbf{h}_t(\theta) = -\mathbf{a}_t(\theta) + \mathbf{c}_t(\theta),$$

where $\mathbf{a}_t(\theta)$ is given by (2.7) and $E[\mathbf{c}_t(\theta_o)|x_t] = \mathbf{0}$. Because $\mathbf{c}_t(\theta_o)$ has mean zero it can be omitted when estimating $E[\mathbf{h}_t(\theta_o)]$. A standard mean value expansion yields

$$\sqrt{T}(\hat{\theta}_T - \theta_o) = [-\ddot{\mathbf{H}}_T]^{-1}T^{-1/2}\mathbf{S}_T(\theta_o)' \quad \text{w.p.a.1.},$$

where $\ddot{\mathbf{H}}_T$ is the Hessian of $\mathscr{L}_T/T$ evaluated at mean values. Assumption (iv) and the fact that $\mathbf{a}_t(\theta_o) + E[\mathbf{h}_t(\theta_o)|x_t] = \mathbf{0}$ imply that

$$-\ddot{\mathbf{H}}_T - \mathbf{A}_T^o \xrightarrow{p} \mathbf{0}$$

by Wooldridge (1986, Chapter 3, Lemma A.1). Combining with (v) gives

$$\sqrt{T}(\hat{\theta}_T - \theta_o) = \mathbf{A}_T^{o-1}T^{-1/2}\mathbf{S}_T(\theta_o)' + o_p(1). \tag{A.1}$$

By (A.1) and the asymptotic equivalence lemma,

$$\left(\mathbf{A}_T^{o-1}\mathbf{B}_T^{o}\mathbf{A}_T^{o-1}\right)^{-1/2}\sqrt{T}(\hat{\theta}_T - \theta_o) \overset{d}{\to} N(\mathbf{0},\mathbf{I}_P).$$

Finally, the consistency of $\hat{\mathbf{A}}_T$ for $\mathbf{A}_T^o$ and of $\hat{\mathbf{B}}_T$ for $\mathbf{B}_T^o$ follow from (iv.a), (vi.a) and (v.a), (vi.b), respectively, by applying Wooldridge (1986, Chapter 3, Lemma A.1). ■

APPENDIX B

In the notation of section 2, the AR(2)-GARCH(1,2) model in (4.1) is written as

(B.1) $$E(y_t|x_t) = \mu_t(\theta_o) = \phi_{o1}y_{t-1} + \phi_{o2}y_{t-2}$$

(B.2) $$V(y_t|x_t) = \omega_t^2(\theta_o) = \delta_o + \alpha_{o1}\varepsilon_{t-1}^2(\theta_o) + \alpha_{o2}\varepsilon_{t-2}^2(\theta_o) + \beta_{o1}\omega_{t-1}^2(\theta_o),$$

where

$$\varepsilon_t(\theta) \equiv y_t - \mu_t(\theta),$$

$\theta \equiv (\phi_1,\phi_2,\delta,\alpha_1,\alpha_2,\beta_1)$, and $\theta_o \equiv (\phi_{o1},\phi_{o2},\delta_o,\alpha_{o1},\alpha_{o2},\beta_{o1})$. Straightforward differentiation of the conditional mean in (B.1) yields the 1x6 vector

(B.3) $$\nabla_\theta\mu_t(\theta) = (y_{t-1},y_{t-2},0,0,0,0).$$

The derivative of the conditional variance function is given by the recursive formula

(B.4) $$\nabla_\theta\omega_t^2(\theta)' = \begin{bmatrix} -2\alpha_1\varepsilon_{t-1}(\theta)y_{t-2} - 2\alpha_2\varepsilon_{t-2}(\theta)y_{t-3} \\ -2\alpha_1\varepsilon_{t-1}(\theta)y_{t-3} - 2\alpha_2\varepsilon_{t-2}(\theta)y_{t-4} \\ 1 \\ \varepsilon_{t-1}^2(\theta) \\ \varepsilon_{t-2}^2(\theta) \\ \omega_{t-1}^2(\theta) \end{bmatrix} + \beta_1\nabla_\theta\omega_{t-1}^2(\theta)'.$$

The derivatives for any model nested within the AR(2)-GARCH(1,2) model can be found by simply fixing the relevant parameters at zero and deleting the corresponding redundant elements in (B.3) and (B.4).

ACKNOWLEDGEMENTS

This is a revised version of MIT Department of Economics Working Paper no. 505, entitled "Quasi-Maximum Likelihood Estimation of Dynamic Models with Time-Varying Covariances." We are grateful to Gary Chamberlain, Rob Engle, James MacKinnon, an anonymous referee, and the editor for helpful comments and suggestions.

REFERENCES

Amemiya, T. (1973), "Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation," *Journal of the American Statistical Association* 68, 928-934.

Baillie, R.T. and T. Bollerslev (1989), "The Message in Daily Exchange Rates: A Conditional Variance Tale," *Journal of Business and Economic Statistics* 7, 297-305.

Baillie, R.T. and T. Bollerslev (1990), "Intra Day and Inter Market Volatility in Foreign Exchange Rates," *Review of Economic Studies*, forthcoming.

Baillie, R.T. and R.P. DeGennaro (1990), "Stock Returns and Volatility," *Journal of Financial and Quantitative Analysis* 25, 203-214.

Basawa, I.V., P.D. Feigin, and C.C. Heyde (1976), "Asymptotic Properties of Maximum Likelihood Estimators for Stochastic Processes," *Sankhya* Series A, 38, 259-270.

Bates, C.E. and H. White (1985), "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory* 1, 151-175.

Berndt, E.R., B.H. Hall, R.E. Hall, and J.A. Hausman (1974), "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement* 4, 653-665.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics* 31, 307-328.

Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics* 69, 542-547.

Bollerslev, T. (1988), "On the Correlation Structure for the Generalized Autoregressive Heteroskedastic Model," *Journal of Time Series Analysis* 9, 121-131.

Bollerslev, T., R.Y. Chou, and K. Kroner (1990), "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, forthcoming.

Bollerslev, T., R.F. Engle, and J.M. Wooldridge (1988), "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy* 96, 116-131.

Breusch, T.R. and A.R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica* 47, 1287-1294.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics* 18, 5-46.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics* 34, 305-334.

Chou, R.Y. (1988), "Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH," *Journal of Applied Econometrics* 3, 279-294.

Cragg, J.G. (1983), "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica* 51, 751-764.

Crowder, M.J. (1976), "Maximum Likelihood Estimation with Dependent Observations," *Journal of the Royal Statistical Society* Series B, 38, 45-53.

Davidson, R. and J.G. MacKinnon (1985), "Heteroskedasticity-Robust Tests in Regression Directions," *Annales de l'INSEE* 59/60, 183-218.

Diebold, F.X. (1986), "Testing for Serial Correlation in the Presence of ARCH," *Proceedings of the American Statistical Association*, *Business and Economic Statistics Section*, 323-328.

Domowitz, I. and C.S. Hakkio (1988), "Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity," manuscript, Northwestern University Department of Economics.

Domowitz, I. and H. White (1982), "Misspecified Models with Dependent Observations," *Journal of Econometrics* 20, 35-58.

Engle, R.F. (1982a), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, 987-1007.

Engle, R.F. (1982b), "A General Approach to Lagrange Multiplier Diagnostics," *Journal of Econometrics* 20, 83-104.

Engle, R.F. (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics*, Vol. II, ed. Z. Griliches and M.D. Intriligator. Amsterdam: North Holland, 775-826.

Engle, R.F. and T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews* 5, 1-50.

Engle, R.F., D.F. Hendry, and D. Trumble (1985), "Small Sample Properties of ARCH Estimators and Tests," *Canadian Journal of Economics* 18, 66-93.

Engle, R.F., D.M. Lilien, and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica* 55, 391-407.

Engle, R.F. and G. Gonzalez-Rivera (1990), "Semiparametric ARCH Models," manuscript, UCSD Department of Economics.

Gallant, A.R., P.E. Rossi, and G. Tauchen (1990), "Stock Prices and Volatility," manuscript, Duke University Department of Economics.

Godfrey, L.G. (1978), "Testing for Multiplicative Heteroskedasticity," *Journal of Econometrics* 8, 227-236.

Gourieroux, C., A, Monfort, and A. Trognon (1984), "Pseudo Maximum Likelihood

Methods: Theory," *Econometrica* 52, 681-700.

Hall, A.D. (1984), "A Study of Various Score Tests for Heteroskedasticity in the General Linear Model," manuscript, Australian National University Department of Economics.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Heijmans, R.D. and J.R. Magnus (1986), "On the First-order Efficiency and Asymptotic Normality of Maximum Likelihood Estimators Obtained from Dependent Observations," *Statistica Neerlandica* 40.

Hsieh, D. (1989), "Modelling Heteroscedasticity in Daily Foreign-Exchange Rates," *Journal of Business and Economic Statistics* 7, 307-317.

Koenker, R. (1981), "A Note on Studentizing a Test for Heteroskedasticity," *Journal of Econometrics* 17, 107-112.

Kroner, K.F. (1987), "Estimating and Testing for Factor ARCH," manuscript, University of Arizona Department of Economics.

Lumsdaine, R.L. (1990), "Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models," manuscript, Harvard University Department of Economics.

MaCurdy, T.E. (1981), "Asymptotic Properties of Quasi-Maximum Likelihood Estimators and Test Statistics," National Bureau of Economic Research Technical Paper No. 14.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley: Chichester.

McCurdy, T.H. and I.G. Morgan (1990), "Tests for a Systematic Risk Component in Deviations from Uncovered Interest Rate Parity," manuscript, Queen's University Department of Economics.

Milhøj, A. (1987), "Autocorrelations of Squared ARCH Processes of First Order," University of Copenhagen Institute of Statistics Research Memorandum No. 100.

Nelson, D.B. (1990a), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory* 6, 318-334.

Nelson, D.B. (1990b), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, forthcoming.

Pagan, A.R. and H. Sabau (1987), "On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models," mimeo, University of Rochester Department of Economics.

Pagan, A.R. and G.W. Schwert (1990), "Alternative Models for Conditional

Stock Volatility," *Journal of Econometrics* 45, 267-290.

Pagan, A.R. and A. Ullah (1988), "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics* 3, 87-105.

Weiss, A.A. (1984), "ARMA Models with ARCH Errors," *Journal of Time Series Analysis* 5, 129-143.

Weiss, A.A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory* 2, 107-131.

White, H. (1982a), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-26.

White, H. (1982b), "Instrumental Variables Regression with Independent Observations," *Econometrica* 50, 483-499.

Wooldridge, J.M. (1986), "Asymptotic Properties of Econometric Estimators," UCSD Department of Economics Ph.D. Dissertation.

Wooldridge, J.M. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests," *Econometric Theory* 6, 17-43.