

Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models

Author(s): Whitney K. Newey and Douglas G. Steigerwald

Source: *Econometrica*, Vol. 65, No. 3 (May, 1997), pp. 587-599

Published by: [The Econometric Society](#)

Stable URL: <http://www.jstor.org/stable/2171754>

Accessed: 16/06/2011 13:09

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=econosoc>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*.

ASYMPTOTIC BIAS FOR QUASI-MAXIMUM-LIKELIHOOD ESTIMATORS IN CONDITIONAL HETEROSKEDASTICITY MODELS

BY WHITNEY K. NEWEY AND DOUGLAS G. STEIGERWALD

Virtually all applications of time-varying conditional variance models use a quasi-maximum-likelihood estimator (QMLE). Consistency of a QMLE requires an identification condition that the quasi-log-likelihood have a unique maximum at the true conditional mean and relative scale parameters. We show that the identification condition holds for a non-Gaussian QMLE if the conditional mean is identically zero or if a symmetry condition is satisfied. Without symmetry, an additional parameter, for the location of the innovation density, must be added for identification. We calculate the efficiency loss from adding such a parameter under symmetry, when the parameter is not needed. We also show that there is no efficiency loss for the conditional variance parameters of a GARCH process.

KEYWORDS: Conditional heteroskedasticity, consistency, quasi-maximum-likelihood.

1. INTRODUCTION

VIRTUALLY ALL EMPIRICAL STUDIES that assume a time-varying conditional variance, termed conditional heteroskedasticity (CH), also use a quasi-maximum-likelihood estimator (QMLE). If the likelihood is assumed to be Gaussian, the QMLE is known to be consistent under correct specification of both the conditional mean and the conditional variance.¹ If the likelihood is assumed to be non-Gaussian, as has become increasingly common in applied work, less is known about the consistency of a QMLE.² We present conditions for the expected conditional log-likelihood to be maximized at the true conditional mean and relative scale parameters, which is a crucial condition for consistency of a QMLE, and study asymptotic efficiency of a QMLE.

We focus on maximization of the expected log-likelihood at the true parameter value because this is the essential identification condition for consistency. We show that if both the assumed innovation density and the true innovation density are unimodal and symmetric around zero (hereafter termed the symmetry condition), then the expected conditional log-likelihood is maximized at the

¹Bollerslev and Wooldridge (1992) show consistency for the Gaussian QMLE for a wide range of CH models but require that the unconditional variance be finite. Lumsdaine (1991), Lee (1991), and Lee and Hansen (1991) show consistency for the Gaussian QMLE for the GARCH(1,1) model without requiring that the unconditional variance be finite.

²Several recent studies use non-Gaussian likelihoods and find that a non-Gaussian likelihood is needed to account for the large number of outliers in the variable under study. For example, Bollerslev (1987) uses a t distribution to model exchange rates and stock returns, Baillie and Bollerslev (1989) use both t and exponential-power distributions to model exchange rates, Hsieh (1989) uses several other distributions to model exchange rates, and Nelson (1991) uses the exponential-power distribution to model stock prices.

true conditional mean and relative scale parameters, so that the identification condition is satisfied. We also show that if one additional parameter is included, the identification condition for consistency is satisfied even if the symmetry condition does not hold. The additional parameter is a location parameter for the innovation density, and is also interpretable as a constant parameter in the heteroskedasticity-corrected equation. Finally, we show that if the conditional mean is identically zero, then neither the symmetry condition nor an additional parameter are needed to establish that the identification condition is satisfied.

Evidence of asymmetry is commonly found in studies of financial variables. Hsieh (1988) studies the statistical properties of daily exchange rates and finds statistically significant negative skewness for each of the five exchange rates that he studies. French, Schwert, and Stambaugh (1987) study the relation between the mean and variance of monthly stock returns and find significant skewness for each of the stock returns that they study.

The addition of a location parameter makes a QMLE robust to asymmetry, but may lead to an efficiency loss if the true innovation density is really symmetric. We derive an expression for the increase in the asymptotic variance of a QMLE from the additional location parameter under symmetry. We also show that if the information matrix is block diagonal between the parameters of the conditional mean and the parameters of the conditional variance, then there is no efficiency loss for a QMLE of the conditional variance parameters. For example, a QMLE for the conditional variance parameters in a GARCH model constructed under the assumption that the innovation has a t density, suffers no efficiency loss from the additional location parameter.

To understand why identification differs for a Gaussian and a non-Gaussian QMLE, we focus on location and scale parameters. For a density $f_Y(y)$ the natural location parameter μ and scale parameter σ are those that maximize $E[-\ln \sigma + \ln f_Y((y - \mu)/\sigma)]$. If f_Y is a Gaussian density, then μ is the mean of Y and σ is the standard deviation of Y . For this reason, if the conditional mean and conditional variance are correctly specified, and if all the parameters of the conditional mean and conditional variance are identified, then the identification condition for consistency is satisfied for a Gaussian QMLE. If f_Y is not a Gaussian density, then the mean and variance are not the natural location and scale parameters. As a result, correct specification of the conditional mean and conditional variance is not sufficient to ensure that the identification condition for consistency is satisfied. To illustrate, for a double exponential density, where $f_Y(y) = \exp(-|y|)/2$, the natural location parameter is the median rather than the mean. In general, different densities have different natural location and scale parameters. Thus for a non-Gaussian QMLE it is important to focus on a model where there are interesting components that are invariant to location and scale. One such model, which we study below, is a location and scale shift of independent and identically distributed innovations.

To describe the models we consider, suppose that the data consist of observations $z_t = (y_t, x_t)'$, $t = 1, \dots, T$, where the period- t variable of interest y_t is a function of period- t regressors x_t . Let $\mathcal{S}_t = (z_{t-1}, z_{t-2}, \dots)$ and let $f_t(\gamma)$ and

$h_t(\gamma)$ denote functions of γ and \mathcal{S}_t , where dependence on \mathcal{S}_t is suppressed for notational convenience. Also, let u_t be the period- t innovation that is identically distributed and independent of \mathcal{S}_t , and let σ be the scale parameter for the density of u_t . One way to describe many CH models that are estimated is

$$(1.1) \quad y_t = f_t(\gamma_0) + \sigma_0 h_t(\gamma_0) u_t.$$

Throughout, true values of parameters are indicated by subscript 0. For example, a first-order autoregressive model with conditional variance that depends linearly on the lagged squared residual (termed an AR(1)-ARCH(1)), has the form of (1.1) with $\gamma = (\beta_1, \beta_2, \phi)$ and

$$f_t(\gamma) = \beta_1 + \beta_2 y_{t-1},$$

$$h_t(\gamma) = \left[1 + \phi (y_{t-1} - \beta_1 - \beta_2 y_{t-2})^2 \right]^{\frac{1}{2}}.$$

An AR(1)-ARCH(1) model illustrates a feature that is implicit in (1.1), which is parameterization in terms of relative scale. The parameter ϕ in the equation for $h_t(\gamma)$ is the relative scale parameter of the conditional variance; that is, ϕ is the ratio of the slope parameter in the conditional variance to the constant parameter in the conditional variance. If the conditional variance is $\delta_1 + \delta_2 (y_{t-1} - \beta_1 - \beta_2 y_{t-2})^2$, then $\phi = \delta_2 / \delta_1$.

Many CH models of the form in (1.1) are estimated by specifying a family of density functions for the innovation u_t , denoted $g(u, \eta)$ where η is a vector of shape parameters. (Throughout, we treat η as a vector of parameters that are estimated). For example, a common specification is to assume that $g(u, \eta)$ is the t density, with η indexing the degrees of freedom. Given an assumed value of g , the QMLE is the value of $\theta = (\gamma', \sigma, \eta')$ that maximizes $L_T(\theta) = T^{-1} \sum_{t=1}^T l_t(\theta)$ where $l_t(\theta)$, the period- t conditional log-likelihood of y_t given \mathcal{S}_t , is

$$(1.2) \quad l_t(\theta) = -\ln \sigma - \ln h_t(\gamma) + \ln g\left([h_t(\gamma)\sigma]^{-1}[y_t - f_t(\gamma)], \eta\right).$$

Equation (1.1) nests models used by: (i) Bollerslev (1987), in which $h_t(\gamma)$ is a generalized autoregressive CH (GARCH) specification and g is a t density; (ii) Baillie and Bollerslev (1989), in which $h_t(\gamma)$ is a GARCH specification and g is either t or exponential power; and (iii) Hsieh (1989), in which $h_t(\gamma)$ is either a GARCH or exponential-GARCH specification and g is either t , exponential power, Gaussian-Poisson mixture, or Gaussian-log-Gaussian mixture. The class of models given by (1.1) also nests models in which the conditional variance is included as a regressor, as in the ARCH-M specification of Engle, Lilien, and Robins (1987).

As discussed below, if $f_t(\gamma_0) = 0$ the identification condition for consistency is satisfied for a QMLE of γ from maximizing $L_T(\theta)$. If $f_t(\gamma_0) \neq 0$, then the identification condition is satisfied if u_t is symmetrically distributed around zero and if the assumed density, $g(u, \eta)$, is symmetric around zero (an even function). If the symmetry condition does not hold, the identification condition for consistency is satisfied by adding a parameter for the location of the distribution of u_t .

The additional parameter is introduced by modifying (1.1) as

$$(1.3) \quad y_t = f_t(\gamma_0) + h_t(\gamma_0)(\alpha_0 + \sigma_0 u_t),$$

where α_0 is the location of the innovation distribution.³ Adding the parameter α_0 has the effect of including $h_t(\gamma_0)$ among the regressors in (1.1). Returning to the AR(1)-ARCH(1) specification, inclusion of the additional parameter yields

$$y_t = \beta_{01} + \beta_{02}y_{t-1} + \left[1 + \phi_0(y_{t-1} - \beta_{01} - \beta_{02}y_{t-2})^2\right]^{\frac{1}{2}}(\alpha_0 + \sigma_0 u_t).$$

Adding the parameter α_0 is distinct from adding the conditional variance as a regressor, as is done in the ARCH-M specification. Consider adding the parameter α_0 to an AR(1)-ARCH(1)-M specification

$$y_t = \beta_{01} + \beta_{02}y_{t-1} + \beta_{03}h_t^2 + \left[1 + \phi_0(y_{t-1} - \beta_{01} - \beta_{02}y_{t-2} - \beta_{03}h_{t-1}^2)\right]^{\frac{1}{2}}(\alpha_0 + \sigma_0 u_t).$$

The parameter β_{03} enters both the conditional mean and conditional variance while, as the example makes clear, the additional parameter α_0 enters the conditional mean but *does not* enter the conditional variance.

With the parameter α included, the QMLE of $\theta = (\gamma', \alpha, \sigma, \eta)'$ is obtained by maximizing $L_T(\theta)$ with

$$(1.4) \quad l_t(\theta) = -\ln \sigma - \ln h_t(\gamma) + \ln g\left([\sigma h_t(\gamma)]^{-1}[y_t - f_t(\gamma) - \alpha h_t(\gamma)], \eta\right).$$

The identification condition for consistency for a QMLE for γ constructed from (1.4), denoted $\hat{\gamma}$, is satisfied even if the symmetry condition does not hold. For example, if g is a t density with η degrees of freedom, then $\hat{\gamma}$ is consistent even if u_t is not symmetrically distributed around zero.

To provide intuition for the result, consider the case in which g is symmetric and non-Gaussian. Recall that our goal is to consistently estimate the parameters of the conditional mean and the relative scale parameters of the conditional variance. As noted in the introduction, if g is not Gaussian, then the mean is generally not the natural location parameter of the assumed density, so correct specification of the conditional mean is not sufficient for consistent estimation of the conditional mean parameters. The identification problem stems from the fact that the conditional mean is not the natural location parameter of the assumed density. If the true density is symmetric, then the mean, median, and mode coincide, so that there is no discrepancy between the conditional mean and the natural location parameter for g . If the true density is asymmetric, however, the discrepancy between the conditional mean and the natural location parameter for g results in a failure to identify the parameters of the conditional mean. The additional parameter α in (1.4) accounts for the discrepancy between

³An alternative way to introduce the additional parameter is to retain (1.1) and to include the location parameter for the innovation density in η .

the conditional mean and the natural location parameter for each possible non-Gaussian QMLE. Note that the problem arises independently of the magnitude of the natural location parameter. Even if the location of the true innovation density is zero, that is if $E[u_t] = 0$ and $\alpha_0 = 0$ in (1.3), which implies that $f_t(\gamma_0)$ is the conditional mean, the additional parameter is needed to account for the discrepancy between the conditional mean and the natural location parameter for g .

As (1.3) makes clear, the presence of a constant in $f_t(\gamma_0)$ does not account for a nonzero location parameter for u_t . In (1.3), $h_t(\gamma_0)$ remains proportional to the conditional standard deviation of y_t , preserving the CH structure of (1.1), but the conditional location of y_t is $f_t(\gamma_0) + \alpha_0 h_t(\gamma_0)$, where the additional term $\alpha_0 h_t(\gamma_0)$ is time varying.⁴ The inclusion of a constant cannot capture the presence of the time-varying term.

To test the restriction $\alpha = 0$ that is typically imposed when constructing a QMLE, a simple score test of the restriction that $\alpha = 0$ can be based on the approach in Wooldridge (1990) that allows for misspecification. The resulting statistic provides a simple way of testing whether the additional parameter α is needed to ensure that a QMLE is consistent.

The previous discussion focuses on consistent estimation of γ . In many instances, such as testing for integration in a GARCH model, the scale or location parameters are also parameters of interest. Although a QMLE of the scale and location parameters is generally inconsistent, there are simple consistent estimators of location and scale given a consistent estimate of γ . To estimate the scale or location parameters corresponding to a particular density g , form $\hat{u}_t = (y_t - f_t(\hat{\gamma}))/h_t(\hat{\gamma})$ and then maximize

$$\hat{L}(\alpha, \sigma) = \sum_{t=1}^n \ln \{g([\hat{u}_t - \alpha]/\sigma, \hat{\eta})/\sigma\}.$$

For example, if g is a standard normal density, the resulting $\hat{\alpha}$ and $\hat{\sigma}$ are the mean and standard deviation, respectively, of \hat{u}_t .

2. IDENTIFICATION

We focus on conditions for the expected log-likelihood to have a unique maximum at the true parameter value. In each case the expected log-likelihood is $\bar{L}(\theta) = E[l_t(\theta)]$, so the conclusion of our results is that $\bar{L}(\theta)$ is maximized at γ_0 . The first result is for the case in which the additional location parameter does not need to be included; that is, either the symmetry condition holds or $f_t(\gamma_0) = 0$. We make the following four assumptions.

⁴To retain interpretation of $f_t(\gamma_0)$ as the conditional mean of y_t , which may be important because the conditional mean is easily linked to expected returns, one must assume that $\alpha_0 + \sigma_0 E[u_t] = 0$ in (1.3).

ASSUMPTION 2.1: $E[l_t(\theta)] < \infty$ for all $\gamma \in \Gamma$, $\eta \in \mathcal{N}$, and $\sigma > 0$.

The sets Γ and \mathcal{N} are feasible sets for the parameters γ and η respectively.

ASSUMPTION 2.2: The function $h_t(\gamma_0) > 0$, and if $\gamma \neq \gamma_0$ then either $h_t(\gamma)/h_t(\gamma_0)$ is not constant or $f_t(\gamma) \neq f_t(\gamma_0)$.

Assumption 2.2 is an identification condition for γ , which is the natural extension of identification conditions for regression models to a CH model. The condition that $h_t(\gamma)/h_t(\gamma_0)$ is not constant means that $h_t(\gamma)$ and $h_t(\gamma_0)$ are not permitted to be constant scale multiples of each other, so a constant scale parameter is excluded from $h_t(\gamma)$.

ASSUMPTION 2.3: The innovation u_t is symmetrically distributed around zero with unimodal density $k(u)$ satisfying $k(u_1) \leq k(u_2)$ for $|u_1| \geq |u_2|$. For each η , $g(u, \eta)$ is symmetric around zero and $g(u_1, \eta) < g(u_2, \eta)$ for $|u_1| > |u_2|$.

Assumption 2.3 imposes symmetry and unimodality on the true innovation density, $k(u)$, and on the assumed innovation density $g(u, \eta)$. Assumption 2.3 also requires that $g(u, \eta)$ be strictly decreasing as u moves away from zero. Many densities satisfy Assumption 2.3, including the t density.

ASSUMPTION 2.4: The function $Q(\sigma, \eta) = -\ln \sigma + E[\ln g(\sigma_0 u_t / \sigma, \eta)]$ has a unique maximum at some $\bar{\sigma}$ and $\bar{\eta}$ over $\sigma > 0$ and $\eta \in \mathcal{N}$.

Assumption 2.4 is a generic identification condition for the scale and shape parameters of the density g . Although it is difficult to specify more primitive conditions for Assumption 2.4, Newey (1986) provides insight. If the true innovation density is a member of the assumed parametric family, then Assumption 2.4 follows if the true innovation density is a one-to-one function of the parameters σ and η (that is, by identification of the parameters (σ, η) for the true innovation density). If the true innovation density is not a member of the assumed parametric family, then, for most innovation densities, Assumption 2.4 follows from identification of the parameters (σ, η) for the true innovation density.

The first result shows that $\bar{L}(\theta)$ is uniquely maximized at γ_0 under either symmetry or $f_t(\gamma_0) = 0$ and the other conditions given above.

THEOREM 1: If Assumptions' 2.1, 2.2, and 2.4 are satisfied and either Assumption 2.3 is satisfied or $f_t(\gamma_0) = 0$, then for $l_t(\theta)$ from (1.2), the expected log-likelihood $\bar{L}(\theta) = E[l_t(\theta)]$ has a unique maximum at some $\bar{\theta}$ with $\bar{\gamma} = \gamma_0$.

PROOF: See Appendix B.

If the location parameter α is added to the model, as in (1.3), it is important to strengthen slightly the identification condition of Assumption 2.2. The following assumption does this.

ASSUMPTION 2.5: *The function $h_t(\gamma_0) > 0$, and if $\gamma \neq \gamma_0$ then either $h_t(\gamma)/h_t(\gamma_0)$ or $[f_t(\gamma) - f_t(\gamma_0)]/h_t(\gamma_0)$ is not constant.*

The requirement that $[f_t(\gamma) - f_t(\gamma_0)]/h_t(\gamma_0)$ is not constant if $\gamma \neq \gamma_0$ implies that one cannot have an additive constant parameter in $f_t(\gamma)$ if $h_t(\gamma_0)$ is constant.

The next assumption is a different version of the identification condition of Assumption 2.4 that explicitly includes a location parameter for the innovation. Including a location parameter for the innovation was not needed in the previous case, because symmetry and unimodality automatically lead to the expected log-likelihood being maximized at $\alpha = 0$.

ASSUMPTION 2.6: *The function $Q(\alpha, \sigma, \eta) = -\ln \sigma + E[\ln g((\sigma_0 u_t + \alpha)/\sigma, \eta)]$ has a unique maximum in $(\alpha, \sigma, \eta)'$.*

With the new assumptions we obtain the following result for the case where the symmetry condition does not hold.

THEOREM 2: *If Assumptions 2.1, 2.5, and 2.6 are satisfied, then for $l_t(\theta)$ from (1.4), the expected log-likelihood $\bar{L}(\theta) = E[l_t(\theta)]$ has a unique maximum at some $\bar{\theta}$ with $\bar{\gamma} = \gamma_0$.*

PROOF: See Appendix B.

Theorem 2 shows that inclusion of the location parameter α ensures that the identification condition for consistency of the QMLE $\hat{\gamma}$ is satisfied even if the symmetry condition does not hold. It is also of interest to know if the converse of Theorem 2 holds, that is, whether *exclusion* of the location parameter for the innovation density implies that a QMLE is *inconsistent*. We show that the converse of Theorem 2 does hold if $E[h_t(\gamma_0)^{-1} \partial f_t(\gamma_0) / \partial \gamma] \neq 0$, in the sense that for each sufficiently regular quasi-likelihood there is a distribution where the first order conditions for maximization at γ_0 are not satisfied. We describe this result later, because it is convenient to first analyze efficiency of a QMLE.

3. ASYMPTOTIC EFFICIENCY

To measure the asymptotic efficiency loss from including the location parameter if the symmetry condition holds, we compare the asymptotic variance of a QMLE obtained by maximizing $L_T(\theta)$ with $l_t(\theta)$ from (1.2) with the asymptotic variance of $\hat{\gamma}$ (recall $\hat{\gamma}$ is a QMLE obtained by maximizing $L_T(\theta)$ with $l_t(\theta)$ from (1.4)). For such a comparison to be sensible both estimators must be consistent, so we assume that the symmetry condition holds. To simplify our results we also assume that the innovation density is correctly specified, so it is easier to understand the effect of including α . Under suitable conditions,

$\sqrt{T}(\hat{\theta} - \bar{\theta}) \rightarrow^d N(0, H^{-1}\Sigma H^{-1})$, where H is the limit quantity for the second derivative of the average log-likelihood and Σ is the covariance matrix from the limit distribution of the first derivative of the average log-likelihood. In particular, correct specification of the innovation density implies that $H = -\Sigma$ (the information matrix equality holds), so the asymptotic variance of a QMLE is Σ^{-1} with $\Sigma = \text{Var}(\partial l_t(\theta)/\partial\theta)$.

To derive a formula for the asymptotic efficiency loss from including α if the symmetry condition holds, let V be the asymptotic variance of a QMLE of γ in the case where α is not estimated. Let V_α be the asymptotic variance of a QMLE of γ in the case where α is estimated. Let $s(u) = g(u, \bar{\eta})^{-1} \partial g(u, \bar{\eta}) / \partial u$, $r(u) = 1 + s(u)u$, $h_t = h_t(\gamma_0)$, $A_t = -h_t^{-1} \partial h_t(\gamma_0) / \partial \gamma$, $B_t = -h_t^{-1} \bar{\sigma}^{-1} \partial f_t(\gamma_0) / \partial \gamma$, and $U_t = (\sigma_0 u_t + \alpha_0 - \bar{\alpha}) / \bar{\sigma}$. Let $\bar{B} = E[B_t]$ and $J = E[s(u_t)^2]$.

THEOREM 3: *The asymptotic efficiency loss from including the location parameter, α , is*

$$V_\alpha - V = V \bar{B} \bar{B}' V / (J^{-1} - \bar{B}' V \bar{B}).$$

PROOF: See Appendix B.

The difference of the asymptotic variance matrices for a QMLE of γ if α is included and if α is excluded is a rank one matrix, which tends to be smaller if \bar{B} is closer to zero.

Perhaps the most frequency constructed QMLE is based on an assumed Gaussian innovation density for a GARCH specification. For a GARCH specification, the asymptotic variance of a Gaussian QMLE under incorrect density specification is also quite simple. Engle and Gonzalez-Rivera (1991) show that for a Gaussian QMLE with a GARCH specification $V = H^{-1} \Sigma H^{-1}$ with $H = (2/(\kappa - 1)) \Sigma$, where κ is the kurtosis of the true innovation density. Therefore, the difference of the asymptotic variance matrices of a misspecified Gaussian QMLE for a GARCH model if α is included and if α is excluded is the rank one matrix

$$\left(\frac{(\kappa - 1)^2}{4} \right)^2 \Sigma^{-1} \bar{B} \bar{B}' \Sigma^{-1} / \left(J^{-1} - \bar{B}' \frac{(\kappa - 1)^2}{4} \Sigma^{-1} \bar{B} \right).$$

There is one important case where there is no asymptotic efficiency loss for estimation of a subset of the parameters of γ . Suppose that γ is partitioned as $\gamma = (\beta', \phi')'$, and that $f_t(\gamma) = f_t(\beta)$ depends only on β , so that ϕ are parameters of the conditional variance. (An example is the ARCH model introduced in Section 1.) Suppose also that

$$(3.1) \quad \text{Cov}(h_t^{-1} \partial h_t(\gamma_0) / \partial \beta, h_t^{-1} \partial h_t(\gamma_0) / \partial \phi) = 0.$$

Condition (3.1) is equivalent to block diagonality of the information matrix between β and ϕ , because $E[B_t B_t']$ is block diagonal by the assumption that ϕ does not appear in $f_t(\beta)$.

COROLLARY: If $f_t(\gamma) = f_t(\beta)$ depends only on β , and $\text{Cov}(h_t^{-1}\partial h_t(\gamma_0)/\partial\beta, h_t^{-1}\partial h_t(\gamma_0)/\partial\phi) = 0$, then there is no asymptotic efficiency loss for a QMLE of ϕ from including the location parameter, α .

PROOF: The block diagonality holds for both the case where α is estimated and where α is not estimated. Furthermore, the block of the information matrix corresponding to ϕ is not affected by the addition of α , so that the asymptotic variance of $\hat{\phi}$ remains the same whether α is included or not. Thus, if (3.1) holds, inclusion of α has no effect on the asymptotic variance of $\hat{\phi}$. *Q.E.D.*

As is well known (3.1) holds for ARCH and GARCH models (for example, Theorem 4 of Engle (1982) for ARCH models). For the exponential-GARCH model (3.1) does not hold, so the asymptotic efficiency loss from including α is given in Theorem 3.

4. INCONSISTENCY WITHOUT A LOCATION PARAMETER

It is straightforward to show that if a location parameter is not included and $E[B_t] \neq 0$, then there are distributions for which a QMLE of γ is inconsistent. For simplicity, we focus on the case where the assumed innovation density is symmetric, so we use the calculations from Section 3. Also, we consider a location shift of the assumed innovation density, obtained by varying α away from zero. Let $E_\alpha[\cdot]$ denote expectation if the true model is that given in (1.3) with $\alpha_0 = \alpha$ and the other parameters equal to their true values, and let $l_t(\theta)$ be the conditional log-likelihood (1.2) where α is *excluded*.

THEOREM 4: *If the symmetry condition holds and $\gamma(\alpha)$ is twice differentiable in α and satisfies the theorem of the maximum, then*

$$\partial\gamma(\alpha)/\partial\alpha|_{\alpha=0} = V^{-1}E[S_t s(u_t)/\sigma_0] = \sigma_0^{-1}V^{-1}E[B_t]E[s(u_t)^2].$$

PROOF: See Appendix B.

Therefore, if $E[B_t] \neq 0$ then $\partial\gamma(\alpha)/\partial\alpha|_{\alpha=0} \neq 0$, implying that a small enough change of α away from zero changes γ . Because $\gamma(0) = \gamma_0$, this means that for a small enough value of α , a QMLE constructed from (1.2) does not satisfy the identification condition. Thus if $E[B_t] \neq 0$, then for any innovation density that is unimodal and symmetric around zero there is another density, consisting of a location shift of the original innovation density (note that the location shift creates a density that is not symmetric around zero), for which a QMLE constructed from (1.2) is inconsistent.

5. CONCLUSION

Consistency of a QMLE for the parameters of interest requires that the expected quasi-log likelihood have a unique maximum at the true value of the parameters of interest. We show that this identification condition holds for a non-Gaussian QMLE of relative scale parameters if either: (i) the conditional mean is identically zero; or (ii) both the assumed and the true innovation densities are symmetric around zero. We also show that if the conditional mean is not identically zero and the innovation density is asymmetric, then an additional parameter is needed to ensure that the identification condition holds. The additional parameter accounts for the location of the innovation density. These results may help clarify the consistency properties of QMLE's for ARCH models, as well as providing an approach to obtaining consistency under asymmetry. In future work, it would be interesting to explore the practical implications of these results, such as the degree of inconsistency arising from asymmetry.

Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02139, U.S.A.

and

Department of Economics, University of California, Santa Barbara, CA 93106, U.S.A.

Manuscript received May, 1994; final revision received July, 1996.

APPENDIX A

In this appendix we give an identification result that is used to prove Theorem 2.

LEMMA A: *If u has density $k(u)$ that is an even function with $k(u) \geq k(u + v)$ for all $u, v \geq 0$, and $\rho(u)$ is an even function with $\int \rho(u + \alpha)k(u)dy < \infty$ for all α and $\rho(u) > \rho(u + v)$ for all $u \geq 0, v > 0$, then $\int \rho(u + \alpha)k(u)du$ has a unique maximum at $\alpha = 0$.*

PROOF: Note that $\int \rho(u - \alpha)k(u)du = \int \rho(-t - \alpha)k(-t)dt = \int \rho(u + \alpha)k(u)du$ by the change of variables $t = -u$ and by symmetry (i.e. $\rho(u)$ and $k(u)$ even functions), so that it suffices to show that $0 < \int \rho(u)k(u)du - \int \rho(u + \alpha)k(u)du = \int [\rho(u) - \rho(u + \alpha)]k(u)du$ for any positive α . Note that

$$\begin{aligned}
 \text{(A.1)} \quad & \int_{-\infty}^0 [\rho(u) - \rho(u + \alpha)]k(u)du \\
 &= \int_0^{\infty} [\rho(u) - \rho(u - \alpha)]k(u)du \\
 &= \int_0^{\infty} [\rho(u + \alpha) - \rho(u)]k(u + \alpha)du + \int_0^{\alpha} [\rho(u) - \rho(u - \alpha)]k(u)du.
 \end{aligned}$$

Also

$$\begin{aligned}
 \int_{\alpha/2}^{\alpha} [\rho(u) - \rho(u - \alpha)]k(u)du &= \int_{-\alpha/2}^0 [\rho(u + \alpha) - \rho(u)]k(u + \alpha)du \\
 &= \int_0^{\alpha/2} [\rho(-u + \alpha) - \rho(-u)]k(-u + \alpha)du \\
 &= \int_0^{\alpha/2} [\rho(u - \alpha) - \rho(u)]k(u - \alpha)du.
 \end{aligned}$$

Note that $0 \leq u \leq \alpha/2$ implies $\alpha - u \geq u$, so that $\rho(u - \alpha) = \rho(\alpha - u) \leq \rho(u)$, and similarly, $k(u - \alpha) \leq k(u)$, implying $[\rho(u) - \rho(u - \alpha)][k(u) - k(u - \alpha)] \geq 0$. Therefore,

$$\begin{aligned} & \int_0^\alpha [\rho(u) - \rho(u - \alpha)]k(u)du \\ &= \int_0^{\alpha/2} [\rho(u) - \rho(u - \alpha)]k(u)du + \int_{\alpha/2}^\alpha [\rho(u) - \rho(u - \alpha)]k(u)du \\ &= \int_0^{\alpha/2} [\rho(u) - \rho(u - \alpha)][k(u) - k(u - \alpha)]du \geq 0. \end{aligned}$$

Then by (A.1) it follows that

$$\begin{aligned} & \int [\rho(u) - \rho(u + \alpha)]k(u)du \\ &= \int_0^\infty [\rho(u) - \rho(u + \alpha)]k(u)du + \int_{-\infty}^0 [\rho(u) - \rho(u + \alpha)]k(u)du \\ &\geq \int_0^\infty [\rho(u) - \rho(u + \alpha)][k(u) - k(u + \alpha)]du \geq 0, \end{aligned}$$

where the last inequality follows by $\rho(u) > \rho(u + \alpha)$ for all $u \geq 0$ and by $k(u) > k(u + \alpha)$ on a subset of $[0, \infty)$ with positive Lebesgue measure. Q.E.D.

APPENDIX B

PROOF OF THEOREM 1: Consider first the case in which Assumption 2.3 (symmetry) holds. Let $\bar{\theta} = (\gamma'_0, \bar{\sigma}, \bar{\eta}')$ for $\bar{\sigma}$ and $\bar{\eta}$ from Assumption 2.4. By assumption 2.3, $Q(\sigma, \eta)$ has a unique maximum at $\bar{\sigma}, \bar{\eta}$. It also follows by Lemma A, which is proven in Appendix A, that for any η , positive scalar s , and any $\alpha \neq 0$, $E[\ln g(su, \eta)] < E[\ln g(su, \eta)]$. Therefore, it follows that

$$\begin{aligned} E_t[l_t(\theta)] &\leq -\ln \sigma - \ln h_t(\gamma) + E[\ln g(\sigma_0 h_t(\gamma_0) u_t / \sigma h_t(\gamma), \eta)] \\ &= Q(\sigma h_t(\gamma) / h_t(\gamma_0), \eta) - \ln h_t(\gamma_0) \leq Q(\bar{\sigma}, \bar{\eta}) - \ln h_t(\gamma_0) = E_t[l_t(\bar{\theta})], \end{aligned}$$

where $E_t[\cdot]$ is the conditional expectation given $\{z_{t-s}^y\}_{s=1}^t$, the first inequality is strict if $f_t(\gamma) \neq f_t(\gamma_0)$, and the second inequality is strict if $\sigma h_t(\gamma) / h_t(\gamma_0) \neq \bar{\sigma}$ or $\eta \neq \bar{\eta}$. Therefore, $E_t[l_t(\theta)] \leq E_t[l_t(\bar{\theta})]$, with strict inequality if $f_t(\gamma) \neq f_t(\gamma_0)$, or $\sigma h_t(\gamma) / h_t(\gamma_0) \neq \bar{\sigma}$, or $\eta \neq \bar{\eta}$. By Assumption 2.2, if $\gamma \neq \gamma_0$ then one of $f_t(\gamma) \neq f_t(\gamma_0)$ and $\sigma h_t(\gamma) / h_t(\gamma_0) \neq \bar{\sigma}$ has positive probability, implying $E_t[l_t(\theta)] < E_t[l_t(\bar{\theta})]$ with positive probability. Therefore, by iterated expectations, $\bar{L}(\theta) = E[E_t[l_t(\theta)]] < E[E_t[l_t(\bar{\theta})]] = \bar{L}(\bar{\theta})$. Furthermore, if $\gamma = \gamma_0$ but $\theta \neq \bar{\theta}$, then it follows by Assumption 2.3 that $E_t[l_t(\theta)] = Q(\sigma, \eta) - \ln h_t(\gamma_0) < Q(\bar{\sigma}, \bar{\eta}) - \ln h_t(\gamma_0) = E_t[l_t(\bar{\theta})]$, so again $\bar{L}(\theta) < \bar{L}(\bar{\theta})$ by iterated expectations.

Consider next the case in which $f_t(\gamma_0) = 0$. Similarly to the previous proof, for $\theta = (\gamma', \sigma, \eta)'$ and $Q(\sigma, \eta) = -\ln \sigma + E[\ln g(\sigma_0 u_t / \sigma, \eta)]$, $E_t[l_t(\theta)] = Q(\sigma h_t(\gamma) / h_t(\gamma_0), \eta) - \ln h_t(\gamma_0) \leq Q(\bar{\sigma}, \bar{\eta})$ with strict inequality if $\sigma h_t(\gamma) / h_t(\gamma_0) \neq \bar{\sigma}$ or $\eta \neq \bar{\eta}$, so the conclusion follows from Assumption 2.4 and the hypothesis that $h_t(\gamma) / h_t(\gamma_0)$ is not constant if $\gamma \neq \gamma_0$. Q.E.D.

PROOF OF THEOREM 2: By Assumption 2.6, $Q(\alpha, \sigma, \eta)$ has a unique maximum at $\bar{\alpha}, \bar{\sigma}, \bar{\eta}$. Let $\bar{\theta} = (\gamma'_0, \bar{\alpha}, \bar{\sigma}, \bar{\eta}')$. It then follows that

$$\begin{aligned} E_t[l_t(\theta)] &= Q(\alpha_0 - h_t(\gamma_0)^{-1} \{f_t(\gamma) - f_t(\gamma_0) + \alpha h_t(\gamma)\}, \sigma h_t(\gamma) / h_t(\gamma_0), \eta) - \ln h_t(\gamma_0) \\ &\leq Q(\bar{\alpha}, \bar{\sigma}, \bar{\eta}) - \ln h_t(\gamma_0) = E_t[l_t(\bar{\theta})], \end{aligned}$$

with strict inequality if $\sigma h_t(\gamma)/h_t(\gamma_0) \neq \bar{\sigma}$, or $\alpha_0 - h_t(\gamma_0)^{-1}\{f_t(\gamma) - f_t(\gamma_0) + \alpha h_t(\gamma)\} \neq \bar{\alpha}$, or $\eta \neq \bar{\eta}$. By Assumption 2.5, if $\gamma \neq \gamma_0$, then either $\sigma h_t(\gamma)/h_t(\gamma_0) \neq \bar{\sigma}$ with positive probability or $\sigma h_t(\gamma)/h_t(\gamma_0) = \bar{\sigma}$ with probability one and $[f_t(\gamma) - f_t(\gamma_0)]/h_t(\gamma_0) \neq \bar{\alpha} - \alpha_0 + \alpha(\bar{\sigma}/\sigma)$ with positive probability, implying that the above inequality holds with positive probability, and hence $\bar{L}(\theta) < \bar{L}(\bar{\theta})$. Also, if $\gamma = \gamma_0$, but $\theta \neq \bar{\theta}$, then $\bar{L}(\theta) < \bar{L}(\bar{\theta})$ also holds by Assumption 2.6. *Q.E.D.*

PROOF OF THEOREM 3: By standard matrix results the asymptotic variance of a QMLE of γ is the inverse second moment matrix of the residuals from the population regression of $\partial l_t(\bar{\theta})/\partial \gamma$ on the other elements of $\partial l_t(\bar{\theta})/\partial \theta$. To calculate the least-squares residual, note that correct specification implies that $\bar{\sigma} = \sigma_0$, $\bar{\eta} = \eta_0$, and $\bar{\alpha} = \alpha_0 = 0$, so that $U_t = u_t$. Also, by symmetry, $s(u_t)$ is an odd function of u and $r(u_t)$ and $\partial l_t(\theta_0)/\partial \eta$ are even functions of u , implying orthogonality of $s(u_t)$ and $(r(u_t), \partial l_t(\theta_0)/\partial \eta)$. Furthermore, u_t is independent of (A_t, B_t) . Then, in the case where α is not estimated, so that $-\sigma_0^{-1}s(u_t)$ is not included among the elements of $\partial l_t(\theta_0)/\partial \theta$, the asymptotic variance of a QMLE of γ is

$$V = (E[S_t S_t'])^{-1} = \{\text{Var}(A_t)E[r(u_t)^2] + E[B_t B_t']E[s(u_t)^2]\}^{-1},$$

where $S_t = (A_t - E[A_t])r(u_t) + B_t s(u_t)$. Here S_t is the population residual from the regression of $\partial l_t(\theta_0)/\partial \gamma$ on the other elements of the score. If α is also estimated, the asymptotic variance is

$$V_\alpha = (E[S_t^\alpha S_t^{\alpha'}])^{-1} = \{\text{Var}(A_t)E[r(u_t)^2] + \text{Var}(B_t)E[s(u_t)^2]\}^{-1},$$

$$S_t^\alpha = (A_t - E[A_t])r(u_t) + (B_t - E[B_t])s(u_t).$$

The form of the asymptotic variance matrices follows from the form of the efficiency bound for semiparametric estimators of γ in (1.3) that is given in Steigerwald (1994). The partitioned inverse formula implies that

$$V_\alpha - V = V\bar{B}\bar{B}'V/(J^{-1} - \bar{B}'V\bar{B}). \tag{Q.E.D.}$$

PROOF OF THEOREM 4: A QMLE constructed from (1.2) converges in probability to the maximizer $\gamma(\alpha)$ of $E_\alpha[l_t(\theta)]$. Thus if α is assumed to equal zero when the true innovation density is asymmetric, $\gamma(\alpha)$ is the limit of a QMLE constructed from (1.2). Then by the theorem of the maximum (or the implicit function theorem applied to the first order conditions for γ) and by the information matrix equality, it follows that

$$\partial \gamma(\alpha) / \partial \alpha |_{\alpha=0} = V^{-1}E[S_t s(u_t) / \sigma_0] = \sigma_0^{-1}V^{-1}E[B_t]E[s(u_t)^2].$$

REFERENCES

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.
 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., AND J. WOOLRIDGE (1992): "Quasi-Maximum Likelihood Estimators and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews*, 11, 143-172.
 ENGLE, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.
 ENGLE, R., AND G. GONZALEZ-RIVERA (1991): "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-360.
 ENGLE, R., D. LILIEN, AND R. ROBINS (1987): "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.
 FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.

- HSIEH, D. (1988): "The Statistical Properties of Daily Foreign Exchange Rates: 1974–1983," *Journal of International Economics*, 24, 129–145.
- (1989): "Modeling Heteroscedasticity in Daily Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 7, 307–317.
- LEE, S. (1991): "Asymptotic Properties of the Maximum Likelihood Estimator of the GARCH-M and IGARCH-M Models," Manuscript, University of Rochester.
- LEE, S., AND B. HANSEN (1991): "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," Manuscript, University of Rochester.
- LUMSDAINE, R. (1991): "Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1, 1) and IGARCH(1, 1) Models," Manuscript, Princeton University.
- NELSON, D. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.
- NEWBY, W. (1986): "Generic Uniqueness of the Population Quasi-Maximum Likelihood Parameters," Manuscript, MIT.
- STEIGERWALD, D. (1994): "Efficient Estimation of Models with Conditional Heteroskedasticity," Manuscript, University of California, Santa Barbara.
- WOOLDRIDGE, J. (1990): "An Encompassing Approach to Conditional Mean Tests with Applications to Testing Nonnested Hypotheses," *Journal of Econometrics*, 45, 331–350.