

Semiparametric ARCH Models

Robert F. Engle and Gloria González-Rivera
Department of Economics D-008, University of California-San Diego, La Jolla, CA 92093

This article introduces a semiparametric autoregressive conditional heteroscedasticity (ARCH) model that has conditional first and second moments given by autoregressive moving average and ARCH parametric formulations but a conditional density that is assumed only to be sufficiently smooth to be approximated by a nonparametric density estimator. For several particular conditional densities, the relative efficiency of the quasi-maximum likelihood estimator is compared with maximum likelihood under correct specification. These potential efficiency gains for a fully adaptive procedure are compared in a Monte Carlo experiment with the observed gains from using the proposed semiparametric procedure, and it is found that the estimator captures a substantial proportion of the potential. The estimator is applied to daily stock returns from small firms that are found to exhibit conditional skewness and kurtosis and to the British pound to dollar exchange rate.

KEY WORDS: Linear spline; Nonparametric density; Quasi-maximum likelihood estimation.

The most common technique for estimating dynamic models with time-varying variance is maximum likelihood. Since the pioneering work of Engle (1982), the assumption of conditional normality $\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t)$ has been widely used in theoretical as well as in empirical research, even though assuming any other probability distribution function will not violate the spirit of the analysis.

Generalized autoregressive conditional heteroscedasticity (GARCH) models as introduced by Bollerslev (1986) have proven to be particularly suited for modeling the behavior of financial time series. It is precisely in this field of research that enough evidence has been found to make it possible to reject the assumption of normality. It is a well-known fact that the unconditional distribution of returns to financial assets exhibits fatter tails than a normal distribution. GARCH models are able to generate this characteristic under the assumption of conditional normality. Leptokurtosis, however, is found in the conditional distribution as well. For instance, Bollerslev (1987) concluded that the monthly returns to the Standard and Poor's 500 (SP500) Composite Index were better fitted with a GARCH model under the assumption of Student-*t* distributed errors. Hong (1988) rejected conditional normality claiming abnormally high kurtosis in the daily New York Stock Exchange stock returns. Gallant, Hsieh, and Tauchen (1989) gave a theoretical explanation of why we should find leptokurtosis in the conditional distribution of financial returns and confirmed their findings in the daily exchange rate of the British pound to dollar (BP/$). Baillie and Bollerslev (1987) used a conditional Student's *t* to model exchange rates.

The fourth moment is not the only concern of researchers. French, Schwert, and Stambaugh (1987) found conditional skewness significantly different from 0 in the standardized residuals when a GARCH-in-mean was fitted to daily SP500 returns. Looking at the unconditional distribution, Peruga (1988) showed empirical evidence of high skewness and kurtosis in exchange rates. Singleton and Wingender (1986) reported high skewness in individual stocks. Badrinath and Chatterjee (1988) showed that the distribution of the daily value-weighted stock returns was explained as a skewed, elongated distribution. Hence there is some empirical evidence for a violation of the distribution assumptions underlying the most common estimate of GARCH models.

For ARCH and GARCH models, there are already quasi-maximum likelihood results in the work of Weiss (1986) and Bollerslev and Wooldridge (1988), who showed that under a correct specification of the first and second moment consistent estimates of the parameters of the model can be obtained by maximizing a likelihood function constructed under the assumption of conditional normality, even though the true density could be some other. The asymptotic standard errors can be estimated consistently, as was done by White (1982) and Gourieroux, Monfort, and Trognon (1984), although they will not attain the Cramer–Rao bound, reflecting the penalty imposed for not knowing the true conditional density.

The purpose of this article is twofold. First, we quantify the loss of efficiency of the quasi-maximum likelihood estimator (QMLE), which falsely assumes normality. Second, we propose a more efficient estimator, based on a nonparametric estimated density.

The rest of the article is organized as follows. In Section 1, quasi-maximum likelihood results are stated and quasi-maximum likelihood relative efficiency is investigated for two types of densities, one leptokurtic and the other positive skewed. In Section 2, we present a semiparametric GARCH estimator and some Monte Carlo simulations. Section 3 shows some empirical applications to individual stock returns and to the BP/$ exchange rate, including an asymmetric or leverage term in the variance equation. A summary is offered in Section 4.

1. QUASI-MAXIMUM LIKELIHOOD ESTIMATION OF GARCH MODELS

1.1 Conditions for the QMLE Results to Hold

Bollerslev and Wooldridge (1988) studied the quasi-maximum likelihood estimation of multivariate GARCH models, and their results can be viewed as a generalization of those of Weiss (1986), who studied QMLE of univariate ARCH models.

We can state the consistency and asymptotic normality of the QMLE as follows:

Theorem (Bollerslev and Wooldridge 1988). If (a) regularity conditions from Appendix A hold and (b) for some $\theta \in \text{int}\, \Theta$, $E(y_t \mid \mathfrak{I}_{t-1}) = \mu_t(\theta_0)$ and $V(y_t \mid \mathfrak{I}_{t-1}) = \Omega_t(\theta_0)$, then $(A_T^{0-1} B_T^0 A_T^{0-1})^{-1/2} \sqrt{T}(\hat{\theta}_T - \theta_0) \overset{d}{\sim} N(0, I)$, where

$$B_T^0 = T^{-1} \sum_{t=1}^{T} E(s_t(\theta_0) s_t(\theta_0)'), \qquad s_t(\theta_0) = \nabla_\theta l_t,$$

$$l_t = -\tfrac{1}{2} \log|\Omega_t(\theta)| - \tfrac{1}{2}(y_t - \mu_t(\theta))' \Omega_t(\theta)^{-1}(y_t - \mu_t(\theta_t)),$$

and

$$A_T^0 = T^{-1} \sum_{t=1}^{T} E(a_t(\theta_0)), \qquad a_t(\theta) = -\nabla_\theta s_t(\theta)'.$$

In addition, $\hat{A}_T - A_T^0 \overset{p}{\to} 0$ and $\hat{B}_T - B_T^0 \overset{p}{\to} 0$, where

$$\hat{A}_T \equiv T^{-1} \sum_{t=1}^{T} a_t(\hat{\theta}_T)$$

and

$$\hat{B}_T \equiv T^{-1} \sum_{t=1}^{T} s_t(\hat{\theta}_T) s_t(\hat{\theta}_T)'.$$

This theorem was stated and proven by both Weiss (1986) and Bollerslev and Wooldridge (1988), with the second authors explicitly considering the multivariate context. Although the regularity conditions were essentially the same, Weiss assumed finite fourth unconditional moments to insure that several of the series satisfy uniform weak laws of large numbers. In fact, this condition is sufficient but not necessary to satisfy the conditions previously listed in (a). This is fortunate because many of the empirical applications of GARCH models find parameter values that are inconsistent with finite unconditional fourth moments. For example, in the ARCH(1) model, Engle (1982) showed that the fourth moment will be finite only if $\alpha_1^2 < \frac{1}{3}$. The widespread finding of integrated (IGARCH) models (Engle and Bollerslev 1986) means that even second moments do not exist.

Using a different line of proof, Lumsdaine (1989) showed in the univariate GARCH(1, 1) and IGARCH(1, 1) that weaker conditions will still support consistency and asymptotic normality of the QMLE. Her sufficient condition is that

$$E\left(\frac{\partial h_t}{\partial \theta} \frac{\partial h_t}{\partial \theta'} h_t^{-2}\right) < \infty,$$

where $h_t = \Omega_t$ in a univariate model. This condition is typically satisfied because the derivative of h_t with respect to the parameters will generally contain squared residuals. When these are divided by conditional variances, the ratio has a well-defined conditional distribution with typically many finite moments. In fact, examining the preceding quasi-likelihood, it is clear that every residual is automatically divided by its standard deviation and thus divergent variances do not necessarily lead to estimates with nonstandard distributions as long as the proper weights are used.

1.2 Relative Efficiency of QMLE

Definition. Relative efficiency (RE) of the QMLE estimate of a parameter θ is the ratio of its asymptotic variance when the true density f is known to its asymptotic variance when normality has been assumed:

$$\text{RE}_\theta = \frac{\text{var}(\hat{\theta}_{\text{MLE}})}{\text{var}(\hat{\theta}_{\text{QMLE}})}.$$

When the conditional probability density function is correctly specified and some regularity conditions are satisfied, we have the well-known result $A_T^{0\,1/2} \sqrt{T}(\hat{\theta}_T - \theta_0) \overset{A}{\sim} N(0, I)$, where A_T^0, as previously defined, is the information matrix; that is,

$$A_T^0 = -T^{-1} \sum_{t=1}^{T} E\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta \partial\theta'}\right)$$

and l_t is the correctly specified log-likelihood function of observation t.

Furthermore, the information equality holds; that is,

$$B_T^0 \equiv T^{-1} \sum_{t=1}^{T} E\left(\frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'}\right) = A_T^0$$

and the numerator of RE_θ is consistently estimated by the inverse of

$$\tilde{B}_T = T^{-1} \sum_{t=1}^{T} E\left(\frac{\partial l_t(\hat{\theta}_T)}{\partial \theta} \frac{\partial l_t(\hat{\theta}_T)}{\partial \theta'} \middle| \psi_{t-1}\right).$$

When the distributional assumptions are not correct, the previously discussed QMLE results apply and the denominator of RE_θ is consistently estimated by $\tilde{A}_T^{-1} \tilde{B}_T \tilde{A}_T^{-1}$, where

$$\tilde{A}_T = -T^{-1} \sum_{t=1}^{T} E\left(\frac{\partial^2 l_t \hat{\theta}_T)}{\partial \theta \partial \theta'} \middle| \psi_{t-1}\right),$$

where l_t is the incorrectly specified likelihood function under the assumption of normality.

The RE ratio is bounded: $0 < \text{RE} \leq 1$. RE = 1 if the true density function is truly normal.

Given that a GARCH(1, 1) is a good representation of the dynamics of the asset returns, in this section we will deal with this type of models: $y_t = \varepsilon_t$, $\text{var}(\varepsilon_t \mid \psi_{t-1}) = h_t = (1 - \alpha - \beta) + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.

We consider two cases:

1. The conditional density of ε_t follows a gamma distribution with shape parameter c:

$$f(\varepsilon_t \mid \psi_{t-1}) = \frac{\sqrt{c}}{h_t^{1/2}\Gamma(c)} \left(\frac{\sqrt{c}\varepsilon_t}{h_t^{1/2}} + c\right)^{c-1} \times \exp\left(-\frac{\sqrt{c}\varepsilon_t}{h_t^{1/2}} - c\right).$$

This type of probability function exhibits positive skewness. The coefficient of skewness is $\sqrt{\beta_1} = 2/\sqrt{c}$ and the coefficient of kurtosis $\beta_2 = 3 + 6/c$. When c tends to infinity, it converges to the normal distribution (Johnson and Kotz 1970).

2. The conditional density of ε_t follows a Student's t with ν df

$$f(\varepsilon_t \mid \psi_{t-1}) = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{\nu + 1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)^{-1} \times ((\nu - 2)h_t)^{-1/2} \left(1 + \frac{\varepsilon_t^2}{h_t(\nu - 2)}\right)^{-(\nu+1)/2}.$$

This distribution is symmetrical around 0 and exhibits leptokurtosis. The coefficient of kurtosis is $\beta_2 = 3 + 6/(\nu - 4)$. When ν tends to infinity, it tends to the unit normal distribution (Johnson and Kotz 1970).

In Appendix B, it is shown that if the true conditional density is a gamma, then

$$\text{RE}_\alpha = \left(\frac{1}{4D_{B_f}} \sum_t \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \beta}\right)^2 \left(\frac{c - u_t^2}{c + u_t}\right)^2\right) \times \left(\frac{(\beta_2 - 1)}{D_{A_N}} \sum_t \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \beta}\right)^2\right)^{-1}, \quad (1)$$

where β_2 is the coefficient of conditional kurtosis $\beta_2 = E(\varepsilon_t^4 \mid \psi_{t-1})/h_t^2$, $u_t = \sqrt{c}\varepsilon_t/h_t^{1/2}$, and D_{B_f}, D_{A_N}, are the determinants of the matrices B_f and A_N, respectively, where f stands for the true density and N for normality. To calculate RE_β, we use (1) substituting $\partial h_t/\partial\beta$ for $\partial h_t/\partial\alpha$.

If the true conditional density is a Student's t,

$$\text{RE}_\alpha = \left(\frac{1}{4D_{B_f}} \sum_t \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \beta}\right)^2 \left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2\right) \times \left(\frac{(\beta_2 - 1)}{D_{A_N}} \sum_t \frac{1}{h_t^2} \left(\frac{\partial h_t}{\partial \beta}\right)^2\right)^{-1}. \quad (2)$$

As before, to calculate RE_β we use (2) substituting $\partial h_t/\partial\beta$ for $\partial h_t/\partial\alpha$. The calculations of these ratios are given in Appendix B.

Table 1. Efficiency of QMLE

Shape	$\alpha = .1$, $\beta = .8$		$\alpha = .5$, $\beta = .4$		$\alpha = .8$, $\beta = .1$	
	True conditional density: Gamma					
$c = 1$	.25	.25*	.25	.25	.25	.25
2	.16	.17	.17	.17	.21	.21
6	.53	.53	.55	.55	.60	.60
12	.73	.73	.73	.74	.81	.82
30	.90	.91	.92	.92	1.00	1.00
	True conditional density: Student's t					
$\nu = 5$	.41	.41*	.41	.41	.42	.42
8	.82	.82	.83	.83	.87	.87
12	.95	.95	.95	.95	.98	.98

NOTE: Model: $y_t = \varepsilon_t$, $\text{var}(\varepsilon_t \mid \psi_{t-1}) = h_t = (1 - \alpha - \beta) + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$. Ratio of Variances = $\text{var}(\theta_{\text{MLE}})/\text{var}(\theta_{\text{QMLE}})$.

* The first number is the relative efficiency of QMLE for α, the second is for β.

In Table 1, we present the RE results for different values of α, β, c, and ν. The efficiency increases as c and ν increase, reflecting the convergence to a normal probability distribution. For instance, for $c = 12$, the coefficient of conditional skewness is .57 and the conditional kurtosis 3.5. For $\nu = 8$, the conditional skewness is 0 and kurtosis 4.5 (the respective values for a normal are 0 and 3). For $c > 30$ and $\nu > 12$, the respective distribution functions are indistinguishable from the normal. Different values of α and β do not seem to affect much the ratio RE, even though there is a small loss of efficiency when the process has a large β.

The efficiency is particularly low for a gamma ($c = 2$). The asymptotic variance of the QMLE estimator is around six times larger than the minimum variance. For this kind of distribution, the conditional skewness is 1.41 and the conditional kurtosis is 6.

QMLE provides a consistent estimator of the asymptotic covariance matrix, which ensures the right size in the traditional hypothesis tests. Although we will be able to make the correct inference, we have shown that the precision of the estimation of a GARCH(1, 1) model could be very low and therefore the power of hypothesis tests will be considerably reduced. Moreover, the forecasting ability of these models will be affected. The precision of the forecast will be diminished; the forecast intervals will be wider. It is worthwhile searching for estimators that can improve on QMLE.

2. SEMIPARAMETRIC GARCH

In this section, we present a generalization of a GARCH model in which the assumption of a known conditional density is relaxed.

Let us consider the following model:

$$y_t = E(y_t \mid \psi_{t-1}) + \varepsilon_t = y_{t|t-1} + \varepsilon_t, \quad (3)$$

with $h_t^{-1/2}\varepsilon_t \sim \text{iid } g(0, 1)$ or $\varepsilon_t \mid \psi_{t-1} \sim f(0, h_t)$; f is the unknown density function of ε_t conditional on the set of past information ψ_{t-1}. Notice that the mean function

$y_{t|t-1}$ can be an autoregressive integrated moving average model or any type of regression model.

The first four conditional moments of ε_t are

$$\frac{E(\varepsilon_t \mid \psi_{t-1})}{h_t^{1/2}} = 0, \quad (4)$$

$$\frac{E(\varepsilon_t^2 \mid \psi_{t-1})}{h_t} = 1, \quad (5)$$

$$\frac{E(\varepsilon_t^3 \mid \psi_{t-1})}{h_t^{3/2}} = s^c, \quad (6)$$

and

$$\frac{E(\varepsilon_t^4 \mid \psi_{t-1})}{h_t^2} = k^c, \quad (7)$$

where h_t is a random variable parameterized as $h_t = \omega_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2$ to obtain an ARCH(p) model (Engle 1982) or as $h_t = \omega_0 + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}$ to obtain a GARCH(p, q) (Bollerslev 1986), s^c is the coefficient of conditional skewness, and k^c the coefficient of conditional kurtosis, both depending on the shape of the conditional density function f. The shape characteristics of the unconditional probability distribution function of ε_t will depend on the features of the conditional distribution.

The coefficient of skewness of the unconditional distribution (s^u) is defined as $s^u = (E\varepsilon_t^3)\sigma^{-3}$, where σ^2 is the unconditional variance of ε_t. For any distribution such as $E(\varepsilon_t^3) < \infty$ and considering that the expectation of a random variable is the expectation of the conditional expectation, we can write, applying (6),

$$s^u = \frac{E\varepsilon_t^3}{\sigma^3} = \frac{E\varepsilon_t^3}{(Eh_t)^{3/2}}$$

$$= \frac{E(E\varepsilon_t^3 \mid \psi_{t-1})}{(Eh_t)^{3/2}} = \frac{E(s^c h_t^{3/2})}{(Eh_t)^{3/2}}.$$

Given that $\varepsilon_t h_t^{-1/2}$ is iid, the conditional skewness is a constant, and we can write

$$s^u = s^c \frac{Eh_t^{3/2}}{(Eh_t)^{3/2}}. \quad (8)$$

Furthermore, because of Jensen's inequality, $Eh_t^{3/2} \geq (Eh_t)^{3/2}$, given that $h_t > 0$, $|s^u| \geq |s^c|$. Hence the unconditional skewness has the same sign as the conditional and is not smaller in absolute value.

Analogously, using (7) and assuming $E(\varepsilon_t^4) < \infty$, we can write

$$k^u = \frac{E\varepsilon_t^4}{\sigma^4} = \frac{E(E\varepsilon_t^4 \mid \psi_{t-1})}{(Eh_t)^2} = \frac{E(k^c h_t^2)}{(Eh_t)^2}$$

$$= k^c \frac{Eh_t^2}{(Eh_t)^2}. \quad (9)$$

The unconditional kurtosis is not smaller than the conditional, $k^u \geq k^c$.

2.1 Estimation of the Model

By the prediction error decomposition, the log-likelihood function for a sample $\varepsilon_1 \ldots \varepsilon_T$ is, apart from some initial conditions, given by

$$l_T(\theta) = \log L_T(\theta) = \sum_{t=1}^{T} \log f(\varepsilon_t \mid \psi_{t-1}).$$

This function will be maximized to estimate the unknown parameters θ as well as the unknown f. The set of parameters θ includes the ones in the mean equation and the ones in the variance equation. To facilitate the estimation, it is convenient to work with the standardized residuals $\varepsilon_t h_t^{-1/2}$. Taking into account the transformation of the random variable, the log-likelihood function looks like

$$l_T(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \log h_t + \sum_{t=1}^{T} \log g\left(\frac{\varepsilon_t}{h_t^{1/2}}\right). \quad (10)$$

Note that we have restricted the set of density functions to those with mean 0 and variance 1. We will obtain the shape of g independently from location and scale.

To maximize (10), we propose the following procedure:

1. Choose some initial consistent estimates of the set of parameters $\tilde{\theta}$. These estimates may come from applying ordinary least squares to (3) (Engle 1982), or from applying quasi-maximum likelihood estimation to (3) (Bollerslev and Wooldridge 1988; Weiss 1986).
2. Save the residuals ε_t and the conditional variances h_t from step 1 and construct the variable $\varepsilon_t h_t^{-1/2}$. Check if the new variable has mean 0 and variance 1; if not, standardize it.
3. Use any of the nonparametric methods for density estimation to estimate $g(\varepsilon_t h_t^{-1/2})$; call it $\hat{g}$. Write the log-likelihood function as in (10), replacing g by $\hat{g}$.
4. Perform the maximization of the log-likelihood function from step 3, keeping $\hat{g}$ fixed and iterating to convergence.

In step 3, we have chosen the discrete maximum penalized likelihood estimation (DMPLE) technique (Tapia and Thompson 1978) to estimate the nonparametric density. This technique consists of maximizing the actual likelihood of a sample in which the arguments of the function are the heights ($p_1 \ldots p_{m-1}$) of a generalized histogram at some given knots ($n_1 \ldots n_{m-1}$). For a sample $x_1 \ldots x_n$ (in our case $\varepsilon_1/h_1^{1/2} \ldots \varepsilon_n/h_n^{1/2}$) and an interval ($a$, b) divided in m subintervals of length q, the following optimization problem has to be solved:

$$\max L(p_1 \ldots p_{m-1})$$

$$= \sum_{i=1}^{m} \log g(x_i) - \frac{\lambda}{q} \sum_{k=1}^{m-1} (p_{k+1} - 2p_k + p_{k-1})^2,$$

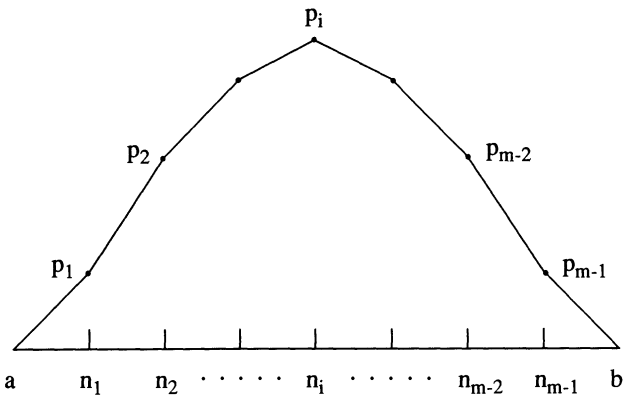

Figure 1. Estimate of a Density Function Using the Discrete Maximum Penalized Likelihood Estimation Technique (Tapia and Thompson 1978).

subject to

$$q \sum_{k=1}^{m-1} p_k = 1, \qquad p_k \geq 0, \quad k = 1 \ldots m - 1,$$

where

$$g(n) = p_k + \frac{p_{k+1} - p_k}{q} (n - n_k),$$

$$n \in [n_k, n_{k+1}); \; g(n) = 0, \; n \notin (n_0, n_m)$$

and λ is the penalty term (chosen by the researcher) to ensure smoothness of the estimate of p. The estimate of the density looks like Figure 1.

Tapia and Thompson (1978) performed Monte Carlo simulations comparing a Gaussian kernel estimate, using the optimal window, to a DMPLE estimate. They showed that the DMPLE estimator is more robust with respect to the choice of the penalty term than the kernel is in respect to the choice of the window width.

In step 4, to perform the maximization we use an iterative algorithm such as that of Berndt, Hall, Hall, and Hausman (BHHH 1974). This requires the calculation of the score function. For a semiparametric GARCH model, the scores are

$$\frac{\partial l_t}{\partial \theta} = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \theta} + \frac{1}{h_t^{1/2}}\left(\frac{\partial \varepsilon}{\partial \theta} - \frac{1}{2}\frac{\partial h_t}{\partial \theta}\frac{\varepsilon_t}{h_t}\right)\frac{\hat{g}_t'}{\hat{g}_t},$$

where θ represents the parameters in the mean and in the variance equation. If θ_2 is the subset of parameters in the variance, $\partial\varepsilon_t/\partial\theta_2 = 0$. DMPLE turns out to be computationally convenient to estimate the ratio $\hat{g}'/\hat{g}$, because $\forall t \in [t_k, t_{k+1})$, $\hat{g}_t' = (p_{k+1} - p_k)/q$.

2.2 Monte Carlo Simulations

To test how the semiparametric procedure performs compared to QMLE in terms of efficiency, we have carried out Monte Carlo simulations for two types of models, GARCH(1, 1) and AR(2)–GARCH(2, 1). All of the following computations have been carried out in the mainframe system Hewlett-Packard 9000/850.

For each model, we have considered that the disturbance term ε_t is conditionally distributed as a Student's t with ν df or as a gamma with shape parameter c. The t_ν distributed random variables were formed as an $N(0, 1)$ distribution divided by the squared root of a χ^2/ν. The normal variate was generated by the International Mathematical and Statistical Libraries (IMSL) subroutine DRNNOR (App. C, footnote 1) and the χ^2 with ν df by DRNCHI (App. C, footnote 2). The gamma-c random variables (g_c) were generated by the IMSL subroutine DRNGAM (App. C, footnote 3).

The artificial data has been generated in the following way. Suppose a generic GARCH model $y_t = y_{t|t-1} + \varepsilon_t$, where $\varepsilon_t = \xi_t h_t^{1/2}$, and the variate ξ_t is iid (0, 1). We generate random numbers from a Student's t with ν df and from a gamma with shape parameter c. These random numbers play the role of ξ_t after some adjustment. Since ξ_t has mean 0 and variance 1, we need to standardize t_ν as well as g_c. The variance of a t_ν is equal to $\nu/(\nu - 2)$. Hence

$$\xi_t = \varepsilon_t h_t^{-1/2} = \frac{t_\nu}{(\nu/(\nu - 2))^{1/2}}.$$

Given some initial values and proceeding recursively, we obtain

$$y_t = y_{t|t-1} + t_\nu h_t^{1/2}\left(\frac{\nu}{\nu - 2}\right)^{-1/2},$$

where the right side depends on past information entirely.

We proceed in a similar way for the gamma. The DRNGAM subroutine produces a variate with mean equal to c and variance equal to c. Hence $\xi_t = \varepsilon_t h_t^{-1/2} = (g_c - c)c^{-1/2}$. From this point on, the generation process continues as before. In Figure 2, we plot the standardized densities of a t with 5 df, a gamma with shape parameter 2, and a normal (0, 1).

All of the estimation results in Tables 2 and 3 are based on 500 replications. In both estimation techniques, QMLE and semiparametric, the convergence criteria used in the BHHH updating regression was $R^2 < .001$. In the semiparametric estimation, the nonparametric density of $\varepsilon_t/h^{1/2}$ has been estimated by the IMSL subroutine DESPL (App. C, footnote 4), which performs the estimation by the penalized likelihood method. The algorithm computes a piecewise linear density function. It requires as inputs the penalty term to ensure smoothness of the estimate, the number of knots, in which the density is estimated, and the support interval (a, b). The number of knots should be chosen as small as possible but large enough to show the structure of the density. The interval (a, b) has been set up as $a = \min(\varepsilon_t/h_t^{1/2}) - .01$ and $b = \max(\varepsilon_t/h_t^{1/2}) + .01$. We have repeated some of the simulations in Tables 2 and 3 with different numbers of knots (from 21 to 200) and different penalty values (from 10 to 100), and the

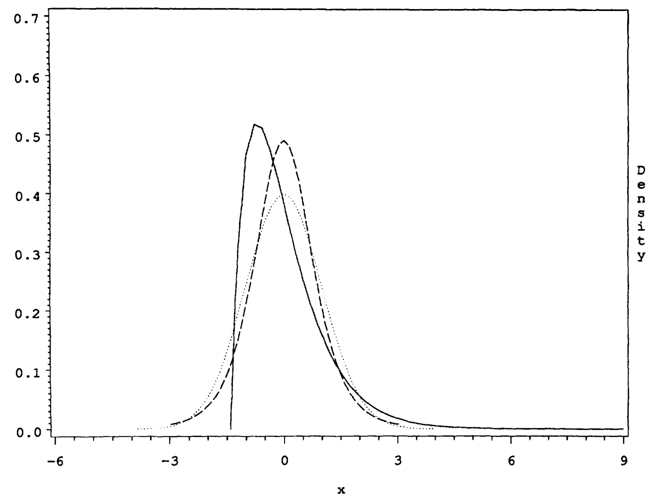

Figure 2. Standardized Density Functions: Normal (0, 1), Student's t With 5 df, and Gamma With Shape Parameter Equal to 2: ---, Standard t_5; ···, N(0, 1), ——, Semiparametric.

results are very robust to the choice of these two parameters. We have used 51 knots and a penalty of 10 for Student's t, and 21 knots and penalty 10 for a gamma.

The parameter values chosen for both models correspond to the empirical findings in the estimation of financial time series. Note that the unconditional fourth moment does not exist in any of the models we consider (App. C, footnote 5). The sum of α and β is close to 1, reflecting the persistence of the volatility, and the parameter β is in the range .6–.8. The parameter values in Table 3 are the same as those of Bollerslev and Wooldridge (1988).

In Table 2, we report the sample means and standard deviations for QMLE and semiparametric estimation of a GARCH(1, 1) for a sample size of $T = 2{,}000$. Different shapes of a gamma distribution have been considered. The larger c is the closer it is to a normal probability distribution. The variability in the QMLE estimates decreases as c increases, reflecting the nearness to normality. The semiparametric estimation produces gains in efficiency for values of c between 2 and 6. The gains are decreasing for larger values of c. For $c = 2$, the Monte Carlo semiparametric ratio of variances shows an improvement in efficiency of 50% over QMLE estimation; for $c = 6$, the gain is only 10%.

When the conditional distribution is a Student's t, we cannot find any gain. We suspect that this poor performance comes from the poor nonparametric estimation of the tails of the density. The difference between a normal and a Student's t probability distribution lies in the tail thickness. For a t distribution with 5 df, the coefficient of kurtosis is 9; for a normal distribution, it is 3. The information in the tails is very dispersed, and its estimation becomes difficult. In general, the behavior of the estimate of the tails of the density is very unstable. We have run some experiments (the results are not reported here) with conditional normal distributed errors to check the proper functioning of the semiparametric estimation. We examined very closely the differences between QMLE and semiparametric estimation, concluding that these come from the poor estimation of the ratio $\hat{g}'/\hat{g}$ in the tails.

In Table 3, we report the means and standard deviations of the QMLE and semiparametric estimates of an AR(2)–GARCH(2, 1) model with a gamma ($c = 2$) and t_5 distributed errors for different sample sizes, $T =$ 2,000, 1,000, and 500. As expected, the variability of the estimates increases with smaller sample sizes. Once more, the gains in efficiency do not seem to be very different from the previous case. With a gamma probability distribution, the ratio of variances shows an im-

Table 2. Effect of the Shape Parameter

	QMLE		Semiparametric		
$T = 2{,}000$	Mean (1)	Standard deviation (2)	Mean (3)	Standard deviation (4)	Ratio $= \frac{(4)}{(2)}$
		Conditional density: $f(\varepsilon_t \mid \psi_{t-1}) \sim$ gamma			
$c = 2$					
α	.204	.030	.196	.021	.70
β	.693	.045	.698	.030	.66
$c = 6$					
α	.202	.022	.202	.021	.95
β	.692	.041	.692	.038	.92
$c = 12$					
α	.201	.020	.201	.021	1.05
β	.697	.034	.695	.036	1.05
		Conditional density $f(\varepsilon_t \mid \psi_{t-1}) \sim$ Student's t			
$\nu = 5$					
α	.206	.029	.204	.029	1.00
β	.690	.049	.692	.048	.97

NOTE: Monte Carlo Results Based on 500 Replications of the Model $y_t = \varepsilon_t$, $\text{var}(\varepsilon_t \mid \psi_{t-1}) = h_t = (1 - \alpha - \beta) + \alpha\varepsilon^2_{t-1} + \beta h_{t-1}$. Parameter Values: $\alpha = .2$, $\beta = .7$

Table 3. Comparison of QMLE to Semiparametric Estimation Procedure

	QMLE		Semiparametric		
	Mean (1)	Standard deviation (2)	Mean (3)	Standard deviation (4)	Ratio = $\frac{(4)}{(2)}$
Part A: Conditional density $f(\varepsilon_t \mid \psi_{t-1}) \sim$ gamma with shape parameter $c = 2$					
$T = 2{,}000$					
ϕ_1	.500	.028	.501	.016	.57
ϕ_2	.150	.027	.149	.017	.62
ω	.102	.019	.101	.015	.78
α_1	.093	.034	.097	.025	.73
α_2	.205	.049	.202	.037	.75
β	.592	.052	.596	.037	.71
$T = 1{,}000$					
ϕ_1	.500	.036	.501	.025	.69
ϕ_2	.149	.037	.152	.025	.67
ω	.106	.033	.106	.022	.66
α_1	.092	.051	.099	.039	.76
α_2	.208	.077	.207	.057	.74
β	.583	.082	.588	.057	.69
$T = 500$					
ϕ_1	.497	.050	.501	.036	.72
ϕ_2	.146	.050	.150	.036	.72
ω	.117	.056	.114	.039	.69
α_1	.094	.078	.096	.061	.78
α_2	.216	.120	.213	.088	.73
β	.560	.131	.576	.090	.68
Part B: Conditional density $f(\varepsilon_t \mid \psi_{t-1}) \sim$ Student t with 5 df ($\nu = 5$)					
$T = 2{,}000$					
ϕ_1	.500	.027	.501	.024	.88
ϕ_2	.146	.027	.147	.029	1.07
ω	.102	.020	.104	.023	1.15
α_1	.093	.033	.096	.035	1.06
α_2	.200	.049	.204	.056	1.14
β	.593	.051	.595	.055	1.07
$T = 1{,}000$					
ϕ_1	.502	.036	.501	.035	.97
ϕ_2	.145	.038	.146	.036	.94
ω	.107	.035	.111	.037	1.05
α_1	.092	.049	.097	.051	1.04
α_2	.205	.079	.214	.086	1.08
β	.581	.090	.579	.092	1.02

NOTE: Monte Carlo results based on 500 replications of the model $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, $\mathrm{var}(\varepsilon_t \mid \psi_{t-1}) = h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta h_{t-1}$. Parameter values: $\phi_1 = .5$, $\phi_2 = .15$, $\omega = .1$, $\alpha_1 = .1$, $\alpha_2 = .2$, and $\beta = .6$.

provement in efficiency between 40% and 60% over QMLE. With a t_5, it is very difficult to improve on QMLE.

If the true density were known to be $\hat{g}$, then the estimated standard errors would be consistent estimates of the relevant elements of the information matrix under the general conditions for maximum likelihood estimation. Instead, this density was itself estimated from

Table 4. GARCH Model: $y_t = b + \varepsilon_t$; $h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma\varepsilon_{t-1}$

	MO			FMC			PDS			AF		
	N(1)	N(2)	NP	N(1)	N(2)	NP	N(1)	N(2)	NP	N(1)	N(2)	NP
b	.0005	.0004	.0004	.0006	.0005	.0001	.0012	.0011	.0014	.0006	.0008	.0003
	(2.01)	(1.72)		(1.11)	(.97)		(2.31)	(2.10)		(1.49)	(2.06)	
ω	.3	.4	.3	.4	.1	.1	.3	.3	.4	1.0	.02	.02
($\times 10^4$)	(3.51)	(3.62)		(1.71)	(1.90)		(2.25)	(2.15)		(4.93)	(1.49)	
α	.0802	.0898	.0784	.0287	.0224	.0207	.0708	.0700	.0738	.2243	.0269	.0292
	(5.15)	(5.32)		(2.67)	(2.88)		(3.51)	(3.37)		(5.34)	(5.13)	
β	.8520	.8318	.8546	.9233	.9634	.9656	.8544	.8602	.8492	.5649	.9690	.9682
	(27.97)	(25.04)		(27.05)	(76.58)		(18.30)	(18.96)		(8.76)	(161.7)	
γ	—	−.0005	—	—	−.0011	−.0013	—	−.0003	—	—	−.0009	−.0009
		(−.92)			(−2.90)			(−.44)			(−2.93)	
log lf	14,224.8	14,227.2	14,616.0	5,706.3	5,718.4	5,850.2	3,494.5	3,494.8	3,577.1	8,482.3	8,526.5	8,805.0

NOTE: N is conditional normal; NP is nonparametric density (for MO, FMC, and AF knots = 91, penalty = 20; for PDS, knots = 41, penalty = 20). Consistent t statistics are in parentheses. Number of observations: MO, 5,903; FMC, 2,769; PDS, 1,453; AF, 3,785.

the data that involve estimating a large number of additional parameters. Only in the special case of adaptation will the standard errors now be correct. In this case, the information matrix is essentially diagonal between the parameters of interest and the nuisance parameters (Bickel 1982; Manski 1984). These conditions are not satisfied for this problem, as the Monte Carlo results illustrate. The estimator does not achieve the full gains. The estimated standard errors are therefore not consistent and are not reported.

3. EMPIRICAL ILLUSTRATION

In this section, we apply the semiparametric GARCH estimation to a set of individual stock returns and to the exchange rate of the British pound in terms of the U.S. dollar (BP/$).

For a set of 115 stocks from Center for Research in Security Prices tapes, we have calculated the Spearman correlation coefficient between size of the firm (ordered from the largest to the smallest according to their capital account) and the first four moments of the unconditional probability distribution of the monthly stock returns for the period 1962–1985. The results are the following:

	Mean	Standard deviation	Skewness	Kurtosis
size	.332	.632	.339	.176
	(.000)	(.000)	(.000)	(.059)

The numbers in parentheses are the p values under H_0: $\rho = 0$. These values suggest that there is a significant relation between the first three unconditional moments and the size of the firm. The small firms have higher mean, higher standard deviation, and higher skewness. The first two facts are well known; small firms are riskier and consequently they offer a larger return. On the contrary, it is not so well reported that small firms tend to have larger skewness. Using daily data for the period 1962–1985, we have estimated a GARCH model for four securities: Mohasco Corporation (MO, furniture), First Mississippi Corporation (FMC, chemicals), Perry Drug Stores (PDS, drugstores), and Airborne Freight (AF, air express transportation).

We have specified two types of GARCH models, the symmetric in variance GARCH(1, 1) and the asymmetric GARCH(1, 1) as in the work of Engle (1990). The latter responds to the need to differentiate between the effects of good news and bad news. Bad news is associated with a large decline in the stock price of the firm; consequently, the value of the firm falls and the debt–equity ratio rises, making the firm riskier (Black 1976). According to this argument, declines in prices will tend to be followed by increases in volatility. To capture the so-called leverage effect, we have introduced the term $\gamma\varepsilon_{t-1}$ in the variance equation. If it is true that bad news increases the volatility of the stock returns, we will expect γ to be negative. Similar features were found important by Nelson (1989) and Schwert (1990).

We have estimated these GARCH models using QMLE, under the assumption of conditional normal errors, and using our semiparametric approach.

In Tables 4–9 we report the estimation results. When we compare the models estimated under QMLE, it seems that the asymmetric GARCH fits FMC and AF stock returns better. The value of the log-likelihood function of the asymmetric GARCH model for FMC and AF is much higher than the one in the symmetric model. Formally, we cannot perform a likelihood ratio test because its distribution is unknown when the assumption of normality fails (White 1982). As we would expect, the asymmetric term in variance is negative but only significant for FMC and AF. The introduction of the asym-

Table 5. Unconditional Distribution of ε_t

Company	Mean (× 10^2)	Variance (× 10^2)	Skewness	Kurtosis
MO	.06	.04	.50	6.77
FMC	.05	.09	.62	4.97
PDS	.15	.04	.54	5.19
AF	.06	.07	.96	10.21

Table 6. Conditional Distribution of $\varepsilon_t h_t^{-1/2}$

Company	Model	Mean	Variance	Skewness	Kurtosis
MO	*N*(1)	.002	1.00	.51	7.41
	N(2)	.004	1.00	.53	7.47
	NP	.003	.99	.51	7.40
FMC	*N*(1)	−.002	1.00	.66	5.18
	N(2)	−.004	.99	.63	4.98
	NP	.009	.99	.68	5.38
PDS	*N*(1)	.004	1.01	.51	5.50
	N(2)	.008	1.01	.53	5.52
	NP	−.0002	.98	.51	5.48
AF	*N*(1)	−.001	1.00	.70	8.04
	N(2)	−.007	1.02	.67	6.83
	NP	.007	.99	.65	7.44

NOTE: *N*(1) = conditional normal; *N*(2) = conditional normal with asymmetric effect in variance equation; *NP* = nonparametric density.

Table 7. Box–Pierce Q Statistic of $\varepsilon_t h_t^{-1/2}$

Company	Model	Q(12)	Q(24)	Q(36)
MO	*N*(1)	11.5	22.0	41.4
	N(2)	11.4	21.9	41.1
	NP	11.5	22.0	41.3
FMC	*N*(1)	21.7	38.4	55.0
	N(2)	23.1	39.6	57.2
	NP	22.5	38.3	54.4
PDS	*N*(1)	10.6	19.2	28.6
	N(2)	10.9	19.3	28.7
	NP	11.1	19.5	29.1
AF	*N*(1)	12.1	25.8	36.0
	N(2)	14.8	22.6	31.2
	NP	15.8	23.6	33.4

NOTE: *N*(1) = conditional normal; *N*(2) = conditional normal with asymmetric effect in variance equation; *NP* = nonparametric density.

Table 8. Autocorrelogram: Residuals From Semiparametric GARCH

Security	1	2	3	4	5	6	7	8	9	10	11	12
						Residuals ε_t						
MO	.008	−.003	−.016	−.014	−.005	−.007	.010	.001	−.03*	.006	−.014	.023
FMC	−.009	−.025	−.012	−.021	.022	−.006	−.003	.000	−.032	.000	−.046*	−.061
PDS	.032	−.014	−.034	−.019	−.018	.039	.022	.021	−.063*	−.038	−.016	.042
AF	.024	−.003	−.006	−.007	.008	−.017	.001	−.010	−.002	−.016	.006	.010
						Squared residuals ε_t^2						
MO	.104*	.064*	.056*	.035*	.036*	.037*	.023	.031*	.027	.026	.035*	.028
FMC	.104*	.009	.029	.028	−.001	.012	.037	.047*	.056*	.009	.004	.045*
PDS	.137*	.111*	.074*	.068*	.053*	.011	.064*	.017	.027	.049	.005	.090*
AF	.089*	.065*	.043*	.042*	.041*	.032	.041*	.018	.030	.004	.029	.027

* Significant at 1% level.

metric term in MO, FMC, and PDS does not very much affect the structure of the model obtained under a symmetric GARCH. This is not the case for AF, where the leverage effect changes the estimates of the model substantially.

If we compare the QMLE model with the semiparametric model, we can reject the null hypothesis of conditional normality at any significance level. As an informal comparison we report the values of the likelihood ratio test, which are 782.4, 263.6, 165.2, and 557 for MO, FMC, PDS, and AF, respectively. The persistence in volatility, measured by the sum of $\alpha + \beta$, ranges from .92 for MO and PDS to .98 for FMC and AF. The conditional distribution of the standardized residuals $\varepsilon_t h_t^{-1/2}$ is positively skewed and leptokurtic. In Figures 3, 4, 5, and 6, there is a comparison between the nonparametric conditional density and a standardized Student's t with 5 df and a normal (0, 1). MO and AF follow quite closely the shape of a standardized t with 5 df, even though those have longer tails. In PDS and FMC, the skewed shape is apparent. At 1% level, we cannot reject the null hypothesis of white-noise standardized residuals as the Box–Pierce Q statistic shows. The assumption of iid standardized residuals is also tested with the BDS statistic. Although the actual size of the BDS statistic, when the procedure is semiparametric ARCH, is not known, we compute nine tests for each set of residuals, and only for two versions of the stock AF can we reject at a 1% level.

In summary, the asymmetric semiparametric GARCH seems to fit the data. A size effect appears to be related to the shape of the unconditional distribution as well as the conditional. Skewness should be taken into account in the estimation of the stock returns.

The second set of examples refers to daily closing prices of the BP/$ from January 2, 1974, until December 30, 1983. This series has been fitted by Gallant et al. (1989). They applied the seminonparametric tech-

Table 9. BDS Statistic[a] for $\varepsilon_t h_t^{-1/2}$ From the Semiparametric GARCH Model

Exponent	.64	.512	.4096
	Company: MO[a]		
m = 2	.05	−.09	−.17
3	−.14	−.12	−.24
4	−.29	−.14	−.30
	Company: FMC		
m = 2	−.23	.08	1.11
3	−.32	−.05	−.85
4	−.30	−.12	−.64
	Company: PDS		
m = 2	−.90	−.33	−.51
3	−.94	−.43	−.56
4	−1.05	−.52	−.68
	Company: AF		
m = 2	−2.00	−.05	.12
3	−2.68[b]	−.15	−.01
4	−3.21[b]	−.23	−.14

[a] PC program provided by David Dechert. For the company MO, the test was run with the first 4,000 observations due to program limitations. The test was also run with the last 4,000 observations, and the results are similar to those reported in this table.
[b] Significant at 1% level.

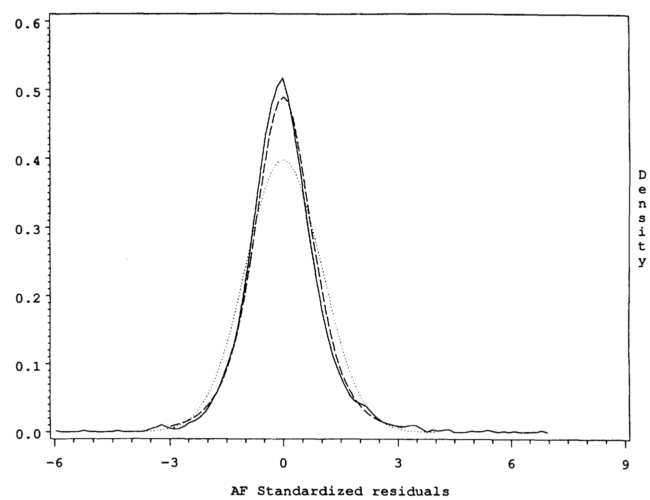

Figure 3. Comparison of the Semiparametric Conditional Density of AF Standardized Residuals to a Normal Density (0, 1) and to a Standardized Student's t With 5 df: ---, Standard t_5; ···, N(0, 1); ——, Semiparametric.

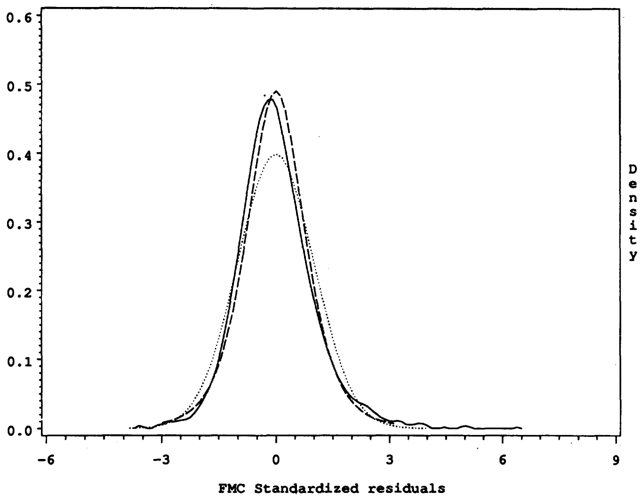

Figure 4. Comparison of the Semiparametric Conditional Density of FMC Standardized Residuals to a Normal Density (0, 1) and to a Standardized Student's t With 5 df: ---, Standard t_5; ···, N(0, 1); ——, Semiparametric.

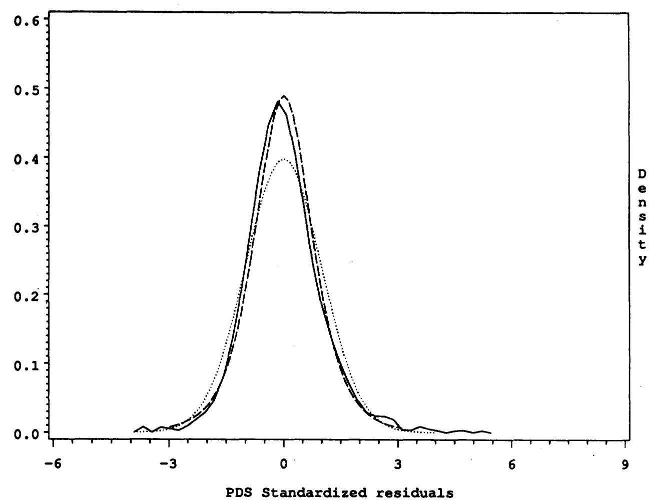

Figure 6. Comparison of the Semiparametric Conditional Density of PDS Standardized Residuals to a Normal Density (0, 1) and to a Standardized Student's t With 5 df: ---, Standard t_5; ···, N(0, 1); ——, Semiparametric.

nique, designed by Gallant and Nychka (1987), which consists of approximating the conditional distribution by a truncated Hermite polynomial expansion using an ARCH-type model as the leading term.

We work with returns to the BP/$ exchange rate calculated as the logarithm of price changes. We adjust the data in the same way that Gallant et al. (1989) did, removing the day-of-the-week effect from the mean as well as from the variance.

The unconditional distribution of these returns is very peaked around 0 and has fat tails (Gallant et al. 1989, fig. 1). The first four unconditional moments are shown in Table 10. The correlogram of the adjusted series is shown in Table 11.

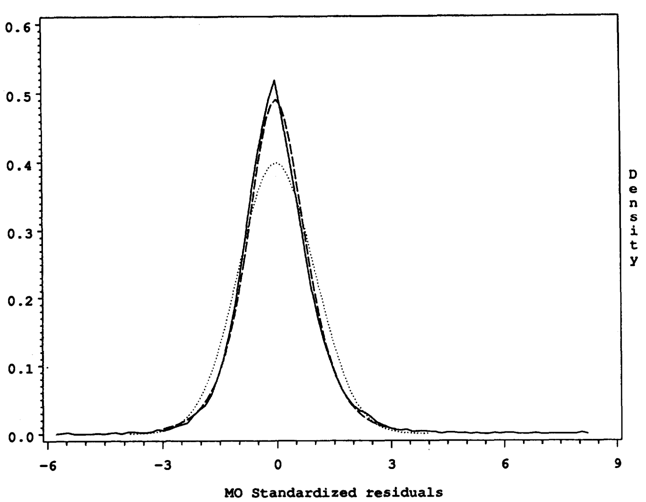

Figure 5. Comparison of the Semiparametric Conditional Density of MO Standardized Residuals to a Normal Density (0, 1) and to a Standardized Student's t With 5 df: ---, Standard t_5; ···, N(0, 1); ——, Semiparametric.

It reveals no linear structure in the mean but a rich structure in the second moments. Given this heterogeneity in the variance, the estimation of a GARCH model seems to be appropriate.

Tables 12, 13, and 14 show the estimation results. We have run a GARCH model and an integrated GARCH model with different distributional assumptions for the disturbance term.

There is strong evidence for integrability in variance. Under a conditional normal density, $\alpha + \beta = .998$ and under a Student's t, $\alpha + \beta = 1.050$. These integrated processes are characterized by persistence in variance; that is, shocks to the conditional variance are permanent. Furthermore, the unconditional fourth moment does not exist; consequently, the unconditional coefficient of kurtosis, previously calculated, is meaningless.

There is an overwhelming rejection of a conditional normal density function for $\varepsilon_t h_t^{-1/2}$. If the alternative is a Student's t, the likelihood ratio test is 598.2. If the alternative is a nonparametric density, an informal comparison is given by the likelihood ratio, which is equal to 504. We reject the null hypothesis of conditional normality at any significance level.

The conditional Student's t is also rejected based on the sample moments of $\varepsilon_t h_t^{-1/2}$. The variance of $\varepsilon_t h_t^{-1/2}$ should be 1 and the coefficient of skewness equal to 0,

Table 10. The First Four Unconditional Moments of the Distribution of Returns to the BP/$ Exchange Rate

	Raw data	Adjusted data
Mean	−.018	.0001
Variance	.350	.998
Skewness	−.413	−.429
Kurtosis	8.902	8.396

Table 11. Correlogram of the Adjusted Series

	Lag									
	1	2	3	4	5	6	7	8	9	10
y_t	−.016	.000	−.006	−.008	.014	−.003	−.006	−.011	.051*	−.008
y_t^2	.152*	.148*	.067*	.064*	.107*	.078*	.032	.039	.025	.049*

* Statistically significant at 1% level.

but the sample variance is 1.41 and the skewness is 1.47.

Rejection of normality and Student's t leads us to estimate an IGARCH with a conditional nonparametric density. We observe that the conditional distribution of $\varepsilon_t h_t^{-1/2}$ is slightly skewed and has fat tails. A plot of the conditional density is given in Figure 7, page 356, and it is compared to a standardized t with 3 df and to a normal(0, 1).

Gallant et al. (1989) found modest evidence for heterogeneity in the conditional density. We have checked the possibility of time-varying standardized third and fourth moments of the integrated semiparametric GARCH. Their correlograms are shown in Table 15, page 357. We also ran the regressions in Table 16, page 357, in the same spirit as time-varying variance models.

Even though there is not a formal test for time-varying skewness and kurtosis, these preliminary results do not seem to indicate any violation of the iid hypothesis. Furthermore, the BDS statistic for the standardized residuals does not reject the assumption of independence.

Summarizing, a semiparametric GARCH can solve the problems found in the estimation of the BP/$ returns and could be an easier alternative to the one presented by Gallant et al. (1989). The series is characterized by an IGARCH with a homogenous nonparametric conditional density that exhibits leptokurtosis and side lobes in the tails.

4. SUMMARY AND CONCLUSIONS

We have introduced a further generalization of GARCH models—the semiparametric GARCH. There is enough empirical evidence to reject the assumption of conditional normality in financial series, a field in which GARCH models have been proven to be extremely successful. It seems sensible to ask how restrictive the assumed normality could be. In this sense, we have investigated the efficiency of the QMLE estimator. We have shown that the loss of efficiency, due to misspecification of the density, could go up to 84% [var($\hat{\theta}_{\mathrm{QMLE}}$) is 6.25 times larger than var($\hat{\theta}_{\mathrm{MLE}}$)]. Consequently, there exists a need to search for more efficient estimators.

If we assume that the mean and variance equations are well specified but we do not know to which probability function they belong, then "the closest" approximation to the true generating mechanism we can think of should come from the data itself. A nonparametric density responds to this concern. In this article, we have proposed a new estimator. The semiparametric GARCH is able to accommodate any particular conditional probability density of the disturbance term.

Monte Carlo results suggest that this semiparametric method can improve the efficiency of the parameter estimates up to 50% over QMLE, but it does not seem to capture the total potential gain in efficiency. In this sense we say that the estimator is not adaptive in the class of densities with mean 0 and variance 1; that is, the estimator is not fully efficient, and it does not achieve the Cramer–Rao lower bound. The information matrix is not block diagonal between the parameters of interest (the ones in the mean and in the variance equation) and the nuisance parameters (the knots of the density).

Table 12. Adjusted Daily Returns BP/$: Model: $y_t = b + \varepsilon_t$; (1) $h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$; (2) $h_t = \omega + \alpha\varepsilon_{t-1}^2 + (1-\alpha)h_{t-1}$

	Normal		Student's t		Semiparametric	
	(1)	(2)	(1)	(2)	(1)	(2)
b	.0202	.0184	.0450	.0437	.0145	.0168
	(.014)	(.014)	(.007)	(.008)		
ω	.0150	.0152	.0004	.0004	.0109	.0151
	(.006)	(.006)	(.000)	(.000)		
α	.1327	.1291	.1979	.1008	.1654	.1270
	(.026)	(.023)	(.026)	(.007)		
β	.8692	—	.8634	—	.8665	—
	(.023)	—	(.011)	—		—
$1/\nu$	—	—	.3031	.2367	—	—
	—	—	(.000)	(.002)	—	—
log lf	−3289.1	−3289.0	−2976.7	−2989.9	−3027.1	−3037.0

NOTE: ν is degrees of freedom. Standard errors are in parentheses: consistent standard errors for normal, outer product of the score for Student's t. In the nonparametric density, knots = 51, and penalty = 20.

Table 13. Adjusted Daily Returns BP/$: Conditional Distribution of $\varepsilon_t h_t^{-1/2}$

Model	Mean	Variance	Skewness	Kurtosis
Normal (1)	−.01	1.00	.05	11.81
Normal (2)	−.01	1.01	.04	11.67
Student-t (1)	−.03	1.07	1.85	44.39
Student-t (2)	−.04	1.41	1.47	36.28
Semiparametric (1)	−.00	.91	.11	13.50
Semiparametric (2)	−.00	1.00	.03	11.60

If we choose the parametric form of the model with a conditional parametric density defined by a shape parameter, this one being part of the parameters to estimate, we can show easily that the expectation of the cross-partial derivatives of the log-likelihood function respects the parameters of interest and the shape parameter is different from 0. In other words, the estimation of the shape parameter affects the efficiency of the estimates of the parameters of interest. This topic will be extended in future research.

To obtain meaningful standard errors for the estimates of the semiparametric model, an alternative estimation procedure can be implemented. Instead of estimating the model in two steps, first the density and second the parameters in the mean and in the variance, we can choose a parametric representation of the density and form a single parameter list composed by all the parameters that define the model and the density function. The maximization of the log-likelihood function will be done with respect to the whole set of parameters at once. We can construct the information matrix so that now it will incorporate the cross-partial derivatives between the parameters that define the density and the parameters of the model. The inverse of the information matrix will retain its interpretation as the variance–covariance matrix.

We have illustrated the use of the semiparametric estimator with two sets of examples. We have chosen a subset of small-firm stock returns, and we have shown that the fitted GARCH process reveals a conditional density that follows closely a Student's t with 5 df, even though there is some positive skewness to account for. The second set of examples is a comparison between the seminonparametric approach reported by Gallant et al. (1989) and our semiparametric method, using the returns to the BP/$ exchange rate. It is shown that the integrated semiparametric GARCH is a simple approach to modeling this "recalcitrant" series.

Table 14. Adjusted Daily Returns BP/$: BDS Statistic for $\varepsilon_t h_t^{-1/2}$ of the Semiparametric (2) Model

Exponent	.64	.512	.4096
$m = 2$	−.02	−.06	−.12
3	−.02	−.07	−.12
4	−.13	−.08	−.15

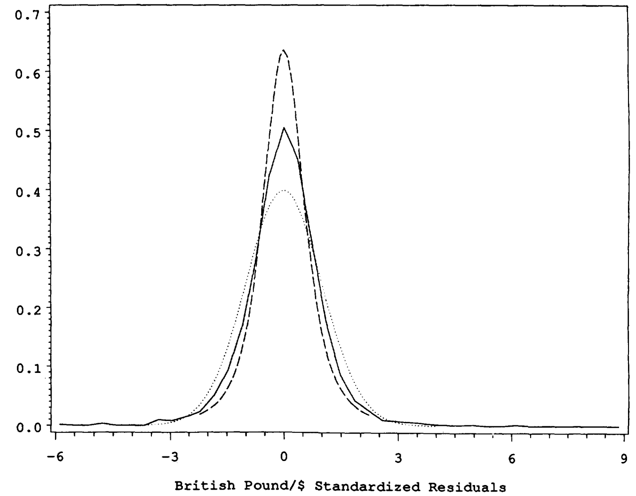

Figure 7. Comparison of the Semiparametric Conditional Density of the BP/$ Exchange Rate Standardized Residuals to a Normal Density (0, 1) and to a Standardized Student's t With 3 df: ---, Standard t_3; ···, N(0, 1); ———, Semiparametric.

ACKNOWLEDGMENTS

We thank Richard Carson, Ian Abramson, Doug Steigerwald, Margarita Genius, the University of California, San Diego, Econometrics Group, anonymous referees, and an associate editor, but we retain responsibility for remaining errors. Financial support by National Science Foundation Grant SES-8910273 is gratefully acknowledged. The data will be archived at the University of Michigan Depository.

APPENDIX A: REGULARITY CONDITIONS OF QMLE

1. θ is compact and has a nonempty interior.
2. The conditional mean and the conditional variance are measurable for all θ and twice continuously differentiable on int θ.
3. (a) $\{l_t(\theta), t = 1, 2, \ldots\}$ satisfies the uniform weak law of large numbers, where $l_t(\theta)$ is the log-likelihood function of observation t. (b) θ_0 is the identifiably unique minimizer of $E(\sum_{t=1}^{T} l_t(\theta))$.
4. (a) $\partial l_t^2/\partial\theta'\partial\theta$ and $E(\partial l_t^2/\partial\theta\partial\theta)$ satisfy the uniform weak law of large numbers. (b) $A_T = -T^{-1}\sum_t E(\partial l_t^2/\partial\theta'\partial\theta)$ is uniformly positive definite.
5. (a) $B_T = T^{-1}\sum_t E(\partial l_t/\partial\theta)'(\partial l_t/\partial\theta)$ is uniformly positive definite. (b) $T^{-1/2}B_T^{-1/2}\sum_t(\partial l_t/\partial\theta)' \to N(0, I_p)$.
6. $(\partial l_t/\partial\theta)'(\partial l_t/\partial\theta)$ satisfies the uniform weak law of large numbers.

APPENDIX B: RELATIVE EFFICIENCY OF QMLE

Consistent estimator of the asymptotic covariance matrix $(A_T^{0-1}B_T^0A_T^{0-1})$ for the model $y_t = \varepsilon_t$, $h_t = (1 - \alpha - \beta) + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$. The scores are

$$\frac{\partial l_t}{\partial\alpha} = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial\alpha}\left(1 - \frac{\varepsilon_t^2}{h_t}\right)$$

Table 15. Adjusted Daily Returns BP/$: Correlogram of the Standardized Third and Fourth Moments of the Integrated Semiparametric GARCH

$\varepsilon_t^3 h_t^{-3/2}$	.006	.000	−.002	−.002	−.014	.000	.000	−.005	.000	.000
$\varepsilon_t^4 h_t^{-2}$	.000	−.002	.000	−.002	.003	−.002	−.002	.000	−.002	−.002

and

$$\frac{\partial l_t}{\partial \beta} = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \beta}\left(1 - \frac{\varepsilon_t^2}{h_t}\right),$$

where

$$\frac{\partial h_t}{\partial \alpha} = \varepsilon_{t-1}^2 - 1 + \beta\frac{\partial h_{t-1}}{\partial \alpha}$$

and

$$\frac{\partial h_t}{\partial \beta} = h_{t-1} - 1 + \beta\frac{\partial h_{t-1}}{\partial \beta}.$$

For $\theta = (\alpha, \beta)$, the expectation of the outer product of the score, matrix B_T^0, is:

$$B_T^0 = \frac{1}{T}\sum_t E\left(\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\right)$$

$$= \frac{1}{T}\sum_t E\left(E\frac{\partial l_t(\theta_0)}{\partial \theta}\frac{\partial l_t(\theta_0)}{\partial \theta'}\bigg| \psi_{t-1}\right).$$

Each element of the matrix B is consistently estimated by

$$B_{11} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2 (\beta_2 - 1),$$

$$B_{22} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2 (\beta_2 - 1),$$

and

$$B_{12} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)\left(\frac{\partial h_t}{\partial \alpha}\right) (\beta_2 - 1),$$

where β_2 is the coefficient of conditional kurtosis $\beta_2 = E(\varepsilon_t^4 \mid \psi_{t-1})/h_t^2$.

The matrix A_T^0 is minus the expectation of the Hessian:

$$A_T^0 = -\frac{1}{T}\sum_t E\left(\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\right)$$

$$= -\frac{1}{T}\sum_t E\left(E\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'}\bigg| \psi_{t-1}\right).$$

Each element of the matrix A is consistently estimated by

$$\hat{A}_{11} = \frac{1}{T}\sum_t \frac{1}{2}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2,$$

$$\hat{A}_{22} = \frac{1}{T}\sum_t \frac{1}{2}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2,$$

and

$$\hat{A}_{12} = \frac{1}{T}\sum_t \frac{1}{2}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)\left(\frac{\partial h_t}{\partial \beta}\right).$$

Obviously, when the conditional density of ε_t is truly normal ($\beta_2 = 3$), the matrices A and B are identical.

The consistent estimator $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ is

$$\widehat{\text{var}}(\alpha) = (\beta_2 - 1)\sum_t \frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2 \frac{1}{D_A} \quad \text{(B.1)}$$

and

$$\widehat{\text{var}}(\beta) = (\beta_2 - 1)\sum_t \frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2 \frac{1}{D_A}, \quad \text{(B.2)}$$

where D_A is the determinant of the matrix $\hat{A}$.

B.2: MLE

If the conditional density is a Student's t with ν df,

$$f(\varepsilon_t \mid \psi_{t-1}) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}$$

$$\times ((\nu - 2)h_t)^{-1/2}\left(1 + \frac{\varepsilon_t^2}{h_t(\nu - 2)}\right)^{-(\nu+1)/2}$$

The scores are

$$\frac{\partial l_t}{\partial \alpha} = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \alpha}\left(1 - \frac{(\nu+1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)$$

and

$$\frac{\partial l_t}{\partial \beta} = -\frac{1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \beta}\left(1 - \frac{(\nu+1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)$$

Table 16. Adjusted Daily Returns BP/$: Regressions of the Standardized Third and Fourth Moments on Their Own Past

Dep. var.	Const.	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	R^2
$\varepsilon_t^3 h_t^{-3/2}$	.023	.006	.000	−.002	−.002	−.014	.000	.000	.000
	(.5)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	
$\varepsilon_t^4 h_t^{-2}$	11.860	−.000	−.002	−.000	−.002	.003	−.002	−.002	.000
	(4.69)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	

NOTE: Standard errors are in parentheses.

and matrix $\hat{B}$:

$$\hat{B}_{11} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2\left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2,$$

$$\hat{B}_{22} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2\left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2,$$

and

$$\hat{B}_{12} = \frac{1}{T}\sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)\left(\frac{\partial h_t}{\partial \beta}\right)\left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2.$$

The estimator of the covariance matrix is $\hat{B}^{-1}$:

$$\widehat{\text{var}}(\alpha) = \sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2 \times \left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2 \frac{1}{D_B} \qquad \text{(B.3)}$$

and

$$\widehat{\text{var}}(\beta) = \sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2 \times \left(1 - \frac{(\nu + 1)\varepsilon_t^2}{\varepsilon_t^2 + h_t(\nu - 2)}\right)^2 \frac{1}{D_B}, \qquad \text{(B.4)}$$

where D_B is the determinant of the matrix $\hat{B}$.

The efficiency of the QMLE procedure is calculated dividing (B.3) by (B.1) and (B.4) by (B.2), respectively.

B.3: MLE

If the conditional density is a gamma with c as a shape parameter:

$$f(\varepsilon_t \mid \psi_{t-1}) = \frac{\sqrt{c}}{h_t^{1/2}\Gamma(c)}\left(\frac{\sqrt{c}\varepsilon_t}{h_t^{1/2}} + c\right)^{c-1} \times \exp\left(-\frac{\sqrt{c}\varepsilon_t}{h_t^{1/2}} - c\right).$$

If we call $u_t = \sqrt{c}\varepsilon_t/h_t^{1/2}$, the scores are:

$$\frac{\partial l_t}{\partial \alpha} = \frac{-1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \alpha}\left(\frac{c - u_t^2}{c + u_t}\right)$$

and

$$\frac{\partial l_t}{\partial \beta} = \frac{-1}{2}\frac{1}{h_t}\frac{\partial h_t}{\partial \beta}\left(\frac{c - u_t^2}{c + u_t}\right).$$

Proceeding in the same way as in B.2, we have the estimator of the covariance matrix $\hat{B}^{-1}$:

$$\widehat{\text{var}}(\alpha) = \sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \beta}\right)^2\left(\frac{c - u_t^2}{c + u_t}\right)^2 \frac{1}{D_B} \qquad \text{(B.5)}$$

and

$$\widehat{\text{var}}(\beta) = \sum_t \frac{1}{4}\frac{1}{h_t^2}\left(\frac{\partial h_t}{\partial \alpha}\right)^2\left(\frac{c - u_t^2}{c + u_t}\right)^2 \frac{1}{D_B}, \qquad \text{(B.6)}$$

where D_B is the determinant of the matrix $\hat{B}$ calculated under a gamma density.

The relative efficiency of QMLE is obtained dividing (B.5) by (B.1) and (B.6) by (B.2), respectively.

APPENDIX C: FOOTNOTES

1. DRNNOR produces a uniform (0, 1) random variable and then uses the inverse of the normal distribution function to generate the normal variate.

2. DRNCHI produces a chi-squared random variable with df degrees of freedom as $\Xi^2 = -2\ln(\prod_{i=1}^n u_i)$, where $n = \text{df}/2$ and u_i are independent random variables from a uniform (0, 1) distribution.

3. DRNGAM produces a gamma random variable with shape parameter c and unit scale parameter. The algorithm uses a 10-region rejection procedure developed by Schmeiser and Lal (1980). For a more complete description of the IMSL subroutines, see IMSL (1987).

4. For a more detailed description, see IMSL (1987).

5. For a GARCH(1, 1) model (Table 2), the necessary and sufficient condition for the existence of the unconditional fourth moment is $\beta^2 + 2\alpha\beta + k^c\alpha^2 < 1$, where k^c is the conditional coefficient of kurtosis. For a GARCH(2, 1) model (Table 3), the condition is $\alpha_2 + \beta^2 + \beta^2\alpha_2 + 2\alpha_1\beta + k^c\alpha_1^2 + k^c\alpha_2^2 + k^c\alpha_1^2\alpha_2 - k^c\alpha_2^3 + 2k^c\alpha_1\alpha_2\beta < 1$. We obtain these conditions considering that $E(\varepsilon_t^4) = E(E(\varepsilon_t^4 \mid \psi_{t-1})) = k^cE(h_t^2)$ and using the law of iterated expectations successively. If the conditional distribution is Normal, $k^c = 3$ and the previous conditions are the same as the ones stated by Bollerslev (1986).

[Received June 1989. Revised March 1991.]

REFERENCES

Badrinath, S. G., and Chatterjee, S. (1988), "On Measuring Skewness and Elongation in Common Stock Return Distributions: The Case of the Market Index," *Journal of Business,* 61, 451–472.

Baillie, R. T., and Bollerslev, T. (1987), "The Message in Daily Exchange Rates: A Conditional Variance Tale," unpublished manuscript.

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

Bickel, P. J. (1982), "On Adaptive Estimation," *The Annals of Statistics,* 10, 647–671.

Black, F. (1976), "Studies of Stock Price Volatility Changes," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association,* pp. 177–181.

Bollerslev, T. P. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics,* 31, 309–328.

——— (1987), "A Conditionally Heteroskedastic Time Series Model for Security Prices and Rates of Return Data," *Review of Economics and Statistics,* 69, 542–547.

Bollerslev, T. P., and Wooldridge, J. M. (1988), "Quasi-Maximum Likelihood Estimation of Dynamic Models With Time Varying Covariances," discussion Paper 505, Massachusetts Institute of Technology, Dept. of Economics.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica,* 50, 987–1008.

——— (1990), "Discussion on Schwert (1990)," *The Review of Financial Studies,* 3, 103–106.

Engle, R. F., and Bollerslev, T. P. (1986), "Modelling and Persistence of Conditional Variances," *Econometric Review,* 5, 1–50.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics,* 19, 3–29.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1989), "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974–83," unpublished manuscript.

Gourieroux, C., Monfort, A., and Trognon, A. (1984), "Pseudo Maximum Likelihood Methods: Theory," *Econometrica,* 52, 681–700.

Hong, C. (1988), "Options, Volatilities and the Hedge Strategy," unpublished Ph.D. dissertation, University of California, San Diego, Dept. of Economics.

International Mathematical and Statistical Libraries (IMSL) (1987), *STAT/LIBRARY: FORTRAN Subroutines for Statistical Analysis,* Houston: Author.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions,* Boston: Houghton-Mifflin.

Lumsdaine, R. L. (1989), "Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1, 1) and IGARCH(1, 1) Models," unpublished manuscript. Harvard University, Department of Economics.

Manski, C. F. (1984), "Adaptive Estimation of Non-Linear Regression Models," *Econometric Reviews,* 3, 145–194.

Nelson, D. B. (1989), "Conditional Heteroskedasticity in Asset Returns: A New Approach," mimeo, University of Chicago, Graduate School of Business.

Peruga, R. (1988), "The Distributional Properties of Exchange Rate Changes Under a Peso Problem," unpublished Ph.D. dissertation, University of California, San Diego, Dept. of Economics.

Schmeiser, B. W., and Lal, R. (1980), "Squeeze Methods for Generating Gamma Variates," *Journal of the American Statistical Association,* 75, 679–682.

Schwert, G. W. (1990), "Stock Volatility and the Crash of '87," *The Review of Financial Studies,* 3, 77–102.

Singleton, J. C., and Wingender, J. (1986), "Skewness Persistence in Common Stock Returns," *Journal of Financial and Quantitative Analysis,* 21, 335–341.

Tapia, R. A., and Thompson, J. R. (1978), *Nonparametric Probability Density Estimation,* Baltimore: Johns Hopkins University Press.

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

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