AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICY UNDER ERROR-TERM NON-NORMALITY

by

Octavio A. Ramírez[1,2,3]

1 Associate Professor, Department of Applied Economics, Texas Tech University, Box 42132, Lubbock TX 79409-2132. Phone: (806) 742-2821, Fax (806) 742-1099, Email: Octavio.Ramirez@ttu.edu

3 Paper presented at the annual meeting of the American Agricultural Economics Association, Chicago, Illinois, August 5-8, 2001

AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICY UNDER ERROR-TERM NON-NORMALITY

ABSTRACT

This paper explores the impact of error-term non-normality on the performance of the normal-error Generalized Autoregressive Conditional Heteroskedastic (GARCH) model under small and moderate sample sizes. A non-normal-, asymmetric-error GARCH model is proposed, and its finite-sample performance is evaluated in comparison to the normal-error GARCH under various underlying error-term distributions. The results suggest that one must be skeptical of using the normal-error GARCH when there is evidence of conditional error-term non-normality. The conditional distribution of the error-term in a previous mainstream application of the normal GARCH is found to be non-normal and asymmetric. The same application is used to illustrate the advantages of the proposed non-normal-error GARCH model.

Keywords: Error- term non-normality, skewness, autoregressive conditional heteroskedasticity.

1. Introduction

The Generalized Autoregressive Conditional Heteroskedastic process (GARCH) (Bollerslev, 1986) and its predecessor, the Autoregressive Conditional Heteroskedastic process (ARCH) (Engle, 1982) have proven useful for modeling a variety of time series phenomena. Many time series variables follow complex autocorrelation structures and are conditionally heteroskedastic. Some, however, are also non-normally distributed.

Bollerslev (1986) indicates that the maximum likelihood (ML) estimator for his GARCH model, which assumes error-term normality, is strongly consistent and asymptotically normal under any true conditional error-term distribution. The asymptotic covariance matrix for the estimator, however, is contingent upon the true error-term distribution. The finite-sample performance of the normal-GARCH model under non-normal true conditional error-term distributions has not been explored. This is important since most time-series applications involve small or moderate sample sizes. In this paper we use standard Monte Carlo simulation procedures to explore the impact of error-term non-normality on the performance of the normal-error GARCH model of Bollerslev (1986) under small and moderate sample sizes.

Partially adaptive estimators parametrically model error-term non-normality to improve efficiency in the estimation of the slope parameters of regression models in finite-sample applications (McDonald and White, 1993). Bollerslev (1987) and Yang and Brorsen (1992) proposed and applied a non-normal-error GARCH model based on the Student-t distribution, which is symmetric but leptokurtotic. We advance a more flexible non-normal-, asymmetric-error GARCH model based on Ramirez and Shonkwiler (2000) partially adaptive inverse hyperbolic sine (IHS) estimator and evaluate its finite-sample performance in comparison to the normal-error GARCH under a variety of true underlying error-term distributions, and through a mainstream empirical example.

2. The Non-Normal IHS-GARCH(p,q) Process

A non-normal-error GARCH(p,q) process analogous to Bollerslev (1986) normal-error GARCH(p,q) process is:

(1) $y_t = x'_t b + \varepsilon_t, \quad \varepsilon_t \sim NN(0,h_t),$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2_{t-i} + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where $NN(0,h_t)$ represents a family of non-normal distributions with mean zero and variance h_t. This process is fully defined by assuming a specific family of non-normal distributions for ε_t. One possibility is Ramirez and Shonkwiler's (2000) expansion of Johnson's (1949) Su family of distributions, which is obtained by letting:

(2) $\varepsilon_t = [\{h_t/G(\Theta,\mu)\}^{1/2}\{\sinh(\Theta v_t)–F(\Theta,\mu)\}]/\Theta, \quad v_t \sim N(\mu,1),$

$F(\Theta,\mu) = E[\sinh(\Theta v_t)] = \exp(\Theta^2/2)\sinh(\Theta\mu)$, and

$G(\Theta,\mu) = \{\exp(\Theta^2)–1\}\{\exp(\Theta^2)\cosh(-2\Theta\mu)+1\}/2$

where $\Theta>0$, $-\infty<\mu<\infty$, and $\sigma>0$ are distributional parameters. Using the results of Johnson, Kotz and Balakrishnan (1994), it can be shown that in this model:

(3) $E[\varepsilon_t] = 0, \quad Var[\varepsilon_t] = h_t,$

$Skew[\varepsilon_t] = E[\varepsilon_t^3] = S(\Theta,\mu) = -1/4w^{½}(w-1)^2[w\{w+2\}\sinh(3\Omega)+3\sinh(\Omega)]/G(\Theta,\mu)^{3/2}$

$Kurt[\varepsilon_t] = E[\varepsilon_t^4] = K(\Theta,\mu) = \{1/8\{w-1\}^2[w^2\{w^4+2w^3+3w^2-3\}\cosh(4\Omega)+4w^2\{w+2\}$
$\cosh(2\Omega)+3\{2w+1\}]/G(\Theta,\mu)^2\}-3$

where $w = \exp(\Theta^2)$ and $\Omega = -\Theta\mu$. The results in (3) imply that $E[y_t] = x'_t b$ regardless of the values of h_t, Θ, and μ, and that the variance of ε_t is the same as in Bollerslev's normal-error GARCH process. The conditional error-term skewness and kurtosis are determined by the parameters Θ and μ. If $\Theta>0$ and μ approaches 0 the error-term distribution becomes symmetric, but it remains kurtotic. Higher values of Θ cause increased kurtosis. If $\Theta>0$ and $\mu>0$, ε_t has a kurtotic and right-skewed distribution, while $\mu<0$ results in a kurtotic and left skewed distribution. Higher values of μ increase both skewness and kurtosis, but kurtosis can be scaled back by reducing Θ (Ramirez and Shonkwiler, 2000).

An advantage of the non-normal-IHS model specification is that the degree of skewness and kurtosis of the conditional error-term distribution can be assumed variable across observations without interfering with the estimation of the linear regression and GARCH process parameters. This is achieved by making Θ and/or μ a function of time or any other potentially relevant factor. Also notice that when $\mu = 0$ the IHS-GARCH model defined above is reduced to the following nested specification:

(4) $y_t = x'_t b + \varepsilon_t,$

$\varepsilon_t = [\{h_t/G(\Theta,0)\}^{1/2}\{\sinh(\Theta v_t)\}]/\Theta,\ \ v_t \sim N(0,1).$

(5) $E[\varepsilon_t] = 0,\ \ Var[\varepsilon_t] = h_t,$

$Skew[\varepsilon_t] = E[\varepsilon_t{}^3] = S(\Theta,0) = 0,$

$Kurt[\varepsilon_t] = E[\varepsilon_t{}^4] = K(\Theta,0) = J(\Theta),$

which implies a symmetric but leptokurtotic error-term model. As Θ goes to zero, ε_t approaches $h_t^{1/2}v_t$ and $J(\Theta)$ becomes zero, indicating that Bollerslev's normal-error

GARCH(p,q) model is nested to the restricted IHS-error GARCH specification in equations (4) and (5) and to the full IHS-error GARCH specification in equations (1), (2) and (3). In practice, under error-term normality, both μ and Θ would approach zero and the proposed IHS-error GARCH estimator would approach Bollerslev's normal-error GARCH estimator. Thus, under the full IHS-error GARCH model specified in equations (1), (2) and (3), the null hypothesis of normality vs. the alternative of non-normality is Ho: $\Theta=\mu=0$ vs. Ha: $\Theta>0$. The null hypothesis of symmetric non-normality versus the alternative of asymmetric non-normality is Ho: $\Theta>0$, $\mu=0$ vs. Ha: $\Theta>0$, $\mu\neq0$.

Given equations (1) and (2), the concentrated log-likelihood function that would have to be maximized when estimating the IHS-GARCH model is obtained using the transformation technique (Mood, Graybill, and Boes 1974):

(6) $$LL = \sum_{i=1}^{n} \ln(G_i) - 0.5\times\sum_{i=1}^{n} H_i^2 \text{ ; where:}$$

$$G_i = \{h_t/G(\Theta,\mu)(1+R_i^2)\}^{-1/2},$$

$$H_i = \{\sinh^{-1}(R_i)/\Theta\}-\mu,$$

$$R_i = [\Theta(y_t\text{-}x'_t b)/\{h_t/G(\Theta,\mu)\}^{1/2}]+F(\Theta,\mu).$$

i=1,…,n refers to the observations, $\sinh^{-1}(x) = \ln\{x+(1+x^2)^{1/2}\}$ is the inverse hyperbolic sine function, and h_t, $F(\Theta,\mu)$, and $G(\Theta,\mu)$ are as given in equations (1) and (2).

3. Properties of the IHS-Error GARCH Estimator

If the distribution of the true conditional error-term (e.g. the error-term underlying the data-generating process) belongs to the expanded form of Johnson's S_U family defined in equation (2), then $E[\sinh(\Theta v_t)] = F(\Theta,\mu)$. This implies that $E[\varepsilon_t] = 0$ and $E[y_t] = x_t b$, regardless of the values of h_t, Θ, and μ. Otherwise:

(7) $E[y_t] = x_t b + \{h_t/G(\Theta,\mu)\}^{1/2}\{E[\sinh(\Theta v_t)]-F(\Theta,\mu)\}/\Theta = x_t b + C$,

Since C is constant with respect to x_t, if the regression equation includes an intercept (bo), the estimator for the intercept will be biased by that constant amount $-C$. If the regressors are fixed in relation to the error-term, the estimators for the slope parameters will remain unbiased. Also, as McDonald and Newey (1988) point out, as long as the error-term is independent of the regressors, any ML-estimator of the location measure of the distribution of y_t conditional on x_t would be a consistent estimator for the regression slopes. Thus, there is no need to assume that ε_t is a member of the expanded S_U family to guarantee unbiased or at least consistent slope parameter estimators.

As any partially adaptive estimator, the proposed IHS-error GARCH estimator would be asymptotically efficient if and only if the true distribution of the conditional error-term is a member of the expanded S_U family and its autocorrelation structure has been properly specified. Under these conditions, standard likelihood theory also guarantees that the maximum likelihood estimators would be asymptotically normal, and that the standard error estimators obtained from the information matrix of the likelihood function would be consistent. When working with finite samples, however, the asymptotic properties are not applicable. In small to moderate sample size applications, the key is to use an estimator based on a flexible family of densities that can accommodate a wide variety of distributional shapes (Ramirez and Shonkwiler, 2000).

Johnson, Kotz and Balakrishnan (1994) indicate that both the log-normal and the normal (Gaussian) family of densities are limiting cases of the S_U family, which also provides for a close approximation for the Pearson family of distributions. They present the Abac for the S_U family and demonstrate that there is an appropriate S_U distribution for any shape factor (e.g. skewness-kurtosis) combination below the log-normal line. Since these shape factor results apply to the proposed expanded form of the S_U family, it is clear that the expanded S_U family allows for any mean and variance, as well as any combination of right/left skewness-leptokurtosis values below the log-normal line. Under zero skewness, it allows for any possible mean-variance-leptokurtosis combination, i.e. it can precisely fit the first four central moments of any symmetric "thick"-tailed distribution.

3. The Monte Carlo Simulation

Monte Carlo simulation is the only alternative to evaluate the finite sample performance of relatively complex models such as the normal and IHS-error GARCH estimators. The basic sample design of Hsieh and Manski (1987), Newey (1988), and McDonald and White (1993) is adopted for the Monte Carlo simulation. A GARCH(1,1) process is assumed for simplicity:

(8) $y_t = b_0 + b_1 x_t + \varepsilon_t = -1 + x_t + h_t^{1/2}\varepsilon_t, \qquad h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \beta_1 h_{t-1}, \qquad h_1 = 1.$

where $x_t = 1$ with a probability of 0.5 and $x_t = 0$ with a probability of 0.5, and x_t is simulated independently of ε_t. This regression model can be interpreted as estimating a shift parameter that separates two distributions that are identical except for a location parameter. The GARCH(1,1) process is simulated under two sets of true parameter values: $\alpha_0 = 1.00$, $\alpha_1 = 0.50$, and $\beta_1 = 0.25$; and $\alpha_0 = 1.00$, $\alpha_1 = 0.25$, and $\beta_1 = 0.50$.

Three specifications for the conditional error-term distribution are taken from McDonald and White (1993): Normal {N(0,1)}; Mixture of normals or variance-contaminated normal {0.9*N(0,1/9)+0.1*N(0,9)}; and Lognormal. Another three non-normal error-term distributions are considered to broaden the spectrum of third-fourth central moment combinations evaluated: A Student-t distribution with three degrees of freedom (as in Phillips, 1998) and two standard normal polynomials {N(0,1)-abs[N(0,1)]^3 and N(0,1)-2[N(0,1)]^2}. An IHS-distributed error-term is also considered, to be used as a benchmark in the evaluation of this model's performance under the alternative error-term distributions discussed above. All error-term distributions are re-scaled and shifted, when necessary, to be drawn from a parent population with zero mean and unitary variance.

Both the mixture of normals and the Student-t are unimodal, thick-tailed, symmetric distributions, with kurtosis coefficients of about 20 and 75, respectively. The log-normal is both thick-tailed and right-skewed, with kurtosis and skewness coefficients of about +5 and +75, respectively. The standard normal polynomials are also thick-tailed and asymmetric, but they are left-skewed instead, exhibiting skewness coefficients of –5 and –2.5, and kurtosis coefficients of about +50 and +10, respectively. The IHS distribution assumed in this case exhibits kurtosis and skewness coefficients of about –4 and +60, respectively.

Different Monte Carlo simulation experiments were conducted with 1000, 200 and 100 samples of sizes of 200, 1000 and 2500, respectively, generated using the same x_t values for each sample. GAUSS 386i programs were used to simulate the data, and the Newton-Raphson algorithm (under a cubic step-length calculation method) preprogrammed within GAUSS 386i constrained maximum likelihood (CML) application module was used for estimating all models. A convergence tolerance level of 10^{-5} was established for the gradients. With few exceptions the CML programs converged properly and produced both parameter and standard error estimates based on the Hessian matrix. The programs utilized are available from the authors upon request.

4. Results

Table 1 presents the Monte Carlo results for the normal and IHS-GARCH models under the smallest sample size of n=200. Both models produce unbiased estimators for the slope parameter regardless of the true error-term distribution being assumed. In the case of the normal GARCH, however, the recommended information-matrix estimator for the standard error of the distribution of the slope-parameter estimator (Bollerslev, 1986) is biased, underestimating the RMSE of the 1000 estimates by an average of 46.6% (Table 4). The analogous information-matrix standard error estimator from the IHS-GARCH model also underestimates the RMSE, but only by an average of 3.2%.

On average, the RMSE of the slope-parameter estimator under the normal GARCH is 314% larger than the RMSE under the IHS-GARCH model. The RMSE differences range from 61% to 702%, depending on the underlying error-term distribution assumed. Under error-term non-normality, both the normal and the IHS-GARCH models are biased estimators for the GARCH(1,1) parameters α_1 and β_1, even when the true error is IHS, which means that the IHS-GARCH is the true maximum likelihood estimator (MLE). In this case, knowledge of the true MLE does not reduce the amount of bias on the estimation of the GARCH(1,1) parameters at n=200.

The IHS-GARCH estimator for α_1, however, appears to be less biased than the normal GARCH estimator for this parameter. The IHS-GARCH yields averages of the α_1 parameter estimates that are closer to the true α_1 value under each of the six underlying non-normal error-terms evaluated, producing an overall average of α_1 estimates of 0.5493 versus 0.7751 when α_1=0.50, and of 0.3027 versus 0.5195 when α_1=0.25 (Table 1). The average bias in the estimation of β_1 appears to be similar under both models. When the simulated error-term is normally distributed, the IHS-GARCH converges to a normal GARCH, producing an unbiased estimator for α_1. At n=200, however, the estimator for β_1 still biased (Table 1).

The RMSE's for both of the normal and the IHS-GARCH estimators for the GARCH(1,1) process parameters α_1 and β_1, which were calculated with respect to the true parameter values, are substantially larger in the case of the normal GARCH (Table 1). This causes a relatively high proportion of normal GARCH models with α_1-β_1 parameter estimate combinations adding up to one or more than one (27.7% versus 8.2% in the case of the IHS-GARCH), which renders the estimated models non-stationary (Bollerslev, 1986). The large RMSE's also produce a relatively larger share of GARCH rejections due to zero-valued parameter estimates for B_1 (an average of 18.6% versus 7.6% in the IHS-GARCH model).

In short, across the 12 non-normal-error GARCH(1,1) combinations evaluated under a sample size of 200, less than 50% of the models estimated using the normal GARCH are stationary with non-zero α_1 and B_1 parameter estimates. In contrast, 83.4% of the models estimated with the proposed IHS-GARCH fulfill these two conditions. If the underlying error-term is normally distributed, an average of 91.6% of the estimated models are stationary with non-zero α_1 and B_1 parameter estimates (Tables 1 and 4).

Another concerning result when using the normal GARCH model under non-normally distributed errors is that the recommended information matrix standard error estimators grossly underestimate the RMSE's of the estimators for α_1 and B_1. When α_1 = 0.5 and β_1 = 0.25, for instance, the averages of the RMSE's across the six underlying non-normal errors evaluated are 0.9331 and 0.2493, while the averages of the standard error estimates are 0.2260 and 0.1376, respectively. In contrast, the IHS-GARCH model produces average RMSE's of 0.2795 and 0.1464 versus average standard error estimates of 0.2529 and 0.1148, respectively (Table 1).

The performance of the normal and the IHS-GARCH models under a considerably larger sample size of n=1000 can be assessed from the statistics in Table 2. As expected, the RMSE's of the slope parameter estimators are substantially smaller than at n=200. The IHS-GARCH slope-parameter estimator again has a substantially lower RMSE than the normal GARCH slope-parameter estimator. In the 12 cases evaluated (six non-normal error distributions by two α_1-β_1 value combinations) the RMSE's are 78% to 1302% larger (457% larger on average) under the normal GARCH (Table 4).

The normal GARCH information-matrix estimator for the standard error of the distribution of the slope-parameter estimator shows a slightly higher bias than at n=200, underestimating the RMSE of the 200 estimates by an average of 54.7% (Table 4). The analogous IHS-GARCH information-matrix standard error estimator only underestimates the RMSE by an average of 1.2%. The amount of bias in the estimators for the

GARCH(1,1) process parameters is still substantial at n=1000. On average, the IHS-GARCH again produces less biased estimates for α_1 (0.4908 vs. 0.6074 when α_1=0.5, and 0.2508 vs. 0.3317 when α_1=0.25).

When β_1=0.25, the average bias in the estimation of this second GARCH process parameter is again similar under the normal and the IHS-GARCH models. When β_1=0.50, the IHS-GARCH produces less biased estimates for this parameter in all cases. If the underlying error-term is normally distributed, the amount of bias in the estimator for β_1 is reduced but not totally eliminated at n=1000 (Table2). The RMSE's of the estimators for the GARCH(1,1) process parameters are again substantially larger in the case of the normal GARCH (Table 2). As a result, even at this larger sample size, the normal GARCH yields a high proportion estimated models that are non-stationary (16.4% vs. 0.25% in the case of the IHS-GARCH). Also, under the normal GARCH, Bollerslev's lagged conditional variance component is rejected an average of 4.8% of the times due to zero-valued parameter estimates for β_1.

In short, across the 12 non-normal-error GARCH(1,1) combinations evaluated at n=1000, only 78.8% of the models estimated using the normal GARCH are stationary with non-zero α_1 and β_1 parameter estimates, while 99.8% of the models estimated with the proposed IHS-GARCH fulfill these two conditions (Table 2). At this larger sample size, under non-normally distributed errors, the normal GARCH underestimates the RMSE's of the estimators for α_1 and B_1 by larger % margins than at n=200. The average of the RMSE's of the α_1 estimators under the six error-term distributions evaluated is 397% larger than the average of the six 200-model averages of the corresponding standard error estimates. The average of the RMSE's of the B_1 estimators is 234% larger. In contrast, the IHS-GARCH RMSE averages are only 10% and 14% larger than the average of the six standard error estimate averages (Table 4).

The previously discussed patterns continue at the largest sample size (n=2500). The standard error estimator from the normal GARCH underestimates the RMSE of the slope parameter estimator by the largest average (65.4%) across the three sample sizes evaluated. The average % RMSE underestimation by the IHS-GARCH remains low (2.8%) and stable across sample sizes (Table 4). The efficiency gains in the estimation of the slope parameter by the IHS vs. the normal GARCH range from 88% to 1381% and average 499%, i.e. they appear to increase slightly with sample size.

Both the normal and the IHS-GARCH estimators for α_1 and B_1 show a lower amount of bias at this largest sample size, with the IHS-GARCH again being less biased in general and on the average. As in the smaller sample sizes, knowledge of the true MLE (i.e. using the IHS-GARCH under an IHS error-term) does not show a particular advantage in this regard (Table 3). Due to the lower RMSE's, the % of estimated normal and IHS-GARCH models that are stationary with non-zero estimates for α_1 and B_1 increases to 89% and 100%, respectively. In the case of the normal GARCH, however, these RMSE's are now underestimated by a larger 474% (α_1) and 270% (B_1), respectively (Table 4).

5. Empirical Example

Finding an application that unambiguously illustrates all of the results from the Monte Carlo simulation discussed above would be a challenging task. Instead, the example in Engle and Kraft (1983), also used by Bollerslev (1986) to illustrate his (GARCH) expansion of Engle's (1982) ARCH process is adopted. In their models, the rate of growth in the implicit U.S. GNP deflator is explained in terms of its own past:

(9) $\pi_t = b_0 + b_1\pi_{t-1} + b_2\pi_{t-2} + b_3\pi_{t-3} + b_4\pi_{t-4} + \varepsilon_t$

where $\pi_t = 100x(GD_t/GD_{t-1})$, GD_t is the implicit price deflator for the GNP (U.S. Department of Commerce, June 2000). The model in equation (9) is estimated using the original time span in Bollerslev (1986) (1948.2 to 1983.4) and an expanded data set (1948.2 to 2000.1), assuming Bollerslev's (1986) normal-error GARCH(1,1) and the proposed IHS-GARCH(1,1) model.

Specifically, an IHS-GARCH model where both the kurtosis and the skewness parameters (θ and μ) are linear functions of time ($\theta = \theta_0 + \theta_1 t$ and $\mu = \mu_0 + \mu_1 t$) is initially assumed. In the case of θ_0, θ_1, μ_0 and μ_1, single-parameter likelihood ratio ($\chi^2_{(1)}$) tests are conducted to verify the asymptotic t-tests results reported in Table 5. The μ_1 parameter is not statistically significant under either the 1948.2 to 1983.4 or the 1948.2 to 2000.1 data. Therefore, this parameter is set equal to zero in the final IHS-GARCH models. Under either data set, both θ_0 and μ_0 are statistically different from zero at the 1% level, indicating that the conditional error-term distribution is leptokurtotic and right-skewed, i.e. that upward inflation spikes are more likely than downward spikes. Since θ_1 is statistically significant as well, the conditional error-term distribution exhibits different levels of kurtosis and skewness through time.

Since both the kurtosis and the skewness coefficients {equation (3)} are monotonically increasing functions of $|\theta|$ ($|\theta_0 + \theta_1 t|$ in this case), the parameter estimates for $\theta_0 = 0.5152$ and $\theta_1 = -0.0060$ (1948.2 to 2000.1 data) indicate that the conditional error-term distribution is more kurtotic and right-skewed at the beginning and at the end of the time period under analysis, and that it is nearly normal at t = 0.5152/0.0061 ≈ 84, which corresponds to 1969.2 (t = 0.4332/ 0.0046 ≈ 94, which corresponds to 1971.4, in the case of the 1948.2 to 1983.4 data). The formulas in equation (3) can be used to calculate the conditional variance, skewness and kurtosis coefficients of ε_t at any time period. The late 1940s, for instance, is a period characterized by relatively large conditional variances (ranging from 4 to 8), skewness (1.6 to 1.8) and kurtosis (5 to 6) coefficients; while the early 1990s exhibit relatively low conditional variances (ranging from 0.9 to 1.1) but similarly large skewness and kurtosis coefficients.

Under both the original and the expanded data sets, the final IHS-GARCH model is statistically superior to the normal GARCH model, according to standard likelihood ratio tests ($\chi^2_{(3)}$ = 140.7356–128.9413 = 11.7951, and $\chi^2_{(3)}$ = 307.2192–282.7756 = 24.4436, respectively). In addition, the standardized residuals ($\hat{\varepsilon}_t/\hat{h}_t$) from the final IHS-GARCH models fail the powerful D'Agostino-Pearson (D'Agostino et al., 1990) normality test ($\chi^2_{(2)}$ = 82.3106, and $\chi^2_{(2)}$ = 28.9622, respectively), while the IHS-transformed standardized residuals {i.e. the $\hat{v}t$'s from equation (2)} do not fail this test ($\chi^2_{(2)}$ = 0.9462, and $\chi^2_{(2)}$ = 3.6569, respectively) (Table 5).

The estimates for the GARCH process and for the intercept and slope parameters of the regression equation are not radically different in this application. However, as expected from the Monte Carlo Simulation results, the corresponding standard error estimates are all substantially lower under the IHS-GARCH (Table 5). The IHS-GARCH advantage in this regard is furthered by the previously discussed simulation evidence about the tendency of the normal GARCH standard error estimators to substantially underestimate the true standard errors when applied under non-normal error-term distribution conditions, as in this case. Inferences based on normal GARCH parameter and standard error estimates would undoubtedly be less reliable.

In regards to forecasting, the two modeling procedures yield different predictions and conditional variance estimates. The average of the inflation rate predictions under the IHS-GARCH (3.4767) is closer to the average of the 205 inflation rates observed during the 1949.1 to 2000.1 period (3.5091) than the average of the predictions under the normal GARCH (3.4472). The root mean square error of the inflation rate predictions is 2.9912 under the normal GARCH vs. 2.9999 under the IHS-GARCH, i.e. it is practically the same under both models. The average of the 205 conditional variance estimates is higher under the IHS-GARCH (1.463 vs 1.418). The difference in the average conditional variance estimates is highest (1.019 vs. 0.906) during the last two decades, when the estimated conditional error term distribution is markedly non-normal.

However, given the simulation evidence discussed above, the IHS-GARCH predictions and conditional variance estimates should be considered more reliable. These factors, in addition to the more realistic assumption about the shape of the conditional error-term distribution, should be reflected on improved confidence intervals for the predictions. True confidence intervals that take in to account the uncertainty due to the estimation of the regression equation and GARCH process parameters as well as the uncertainty arising from the inherent stochastic nature of the true data-generating process can be obtained through standard Monte Carlo Simulation procedures. Specifically, 50,000 sets of parameter values are simulated using the maximum likelihood parameter and covariance matrix estimates. Each set of simulated parameter values is used to generate a vector of 205 inflation rate and conditional variance "predictions" for the last 205 time periods in the analysis.

In the case of the normal GARCH, each of the 50,000 vectors of conditional variance "predictions" is used in conjunction with 205 independent draws from a standard normal distribution to simulate 50,000 vectors containing 205 draws from the conditional error-term distributions corresponding to the last 205 time periods in the analysis. The 50,000 vectors of inflation rate predictions are then added to the corresponding 50,000 vectors of conditional error-term distribution draws to obtain m=50,000 simulated inflation rate values for each of the time periods in the analysis. Then, the boundaries of a true (1-α)% confidence interval for the inflation rate realizations through time are obtained by finding the $(\alpha/2)$ *x* m^{th} and the $[(1-\alpha)+\alpha/2]$ *x* m^{th} largest of these m simulated values for each of the last 205 time periods in the sample. The same process is followed for the IHS-GARCH, except that the conditional (non-normal) error-term distributions are simulated on the basis of equation (2).

The boundaries of the 80% confidence intervals for the inflation rate realizations implied by the normal and IHS-GARCH models are compared with the data in Figures 1 and 2. The difference between these two confidence intervals is best perceived in the relatively inflation stable 1984-2000 period. In the normal GARCH, the inflation rate observations tend to be closer to the middle of the interval, only four observations trespass the lower bound, while nine observations surpass the upper bound of the 80% confidence interval. Given this pattern of observations, the symmetry of the assumed conditional error-term distribution requires a lower bound that is unnecessarily low in order to avoid more of the observed inflation peaks surpassing the upper bound. In the IHS-GARCH, the flexible asymmetry (right-skewness in this case) in the assumed conditional error-term distribution allows for a noticeably higher lower bound, which is very close to the bulk of the observations, coupled with an upper bound that is still high enough to avoid a theoretically excessive number of observations surpassing it.

A similar pattern is observed during the 1949-1960 period. Only during the 1965-1975 period, when the conditional error-term distribution estimated under the IHS-GARCH is nearly normal, are the boundaries of the confidence intervals from the two models almost identical. In addition to these visual patterns, the numerical evidence is clear: Under the normal GARCH, only 14 out of 205 observations (6.8%) are below the lower bound, while 30 (14.6%) exceed the upper bound of the 80% confidence interval. The average width of the confidence interval is 3.68. Under the normal GARCH, 19 observations (9.3%) are below the lower bound and 22 (10.7%) above the upper bound, while the average width of the confidence interval is 3.54.

Similar patterns arise in the case of the 81% through the 95% confidence intervals (Table 6), although the average width of the 95% confidence interval becomes larger under the IHS-GARCH, presumably due to the pronounced right tail of the estimated IHS conditional error-term distribution. Cumulatively for the 80% to 95% confidence intervals, the normal GARCH results in 388 observations beyond the boundaries of these 16 intervals, which is 5.4% lower than the number that would be theoretically expected

($205x\sum_{i=5}^{20} i/100$ = 410). In addition, only 117 (57.1%) out of the theoretically expected 410/2 = 205 (100%) observations are below the lower bounds, and 271 (132.2%) exceed the upper bounds. Under the IHS-GARCH, a total of 410 observations exceed the boundaries of the 16 intervals, 219 (106.8%) being below the lower bounds and 191 (93.2%) surpassing the upper bounds (Table 6). The cumulative average width of the 16 confidence intervals is smaller (70.91 vs. 72.32) in the case of the IHS-GARCH model.

In short, under the same application originally used to illustrate the normal ARCH and GARCH models, the IHS-GARCH confidence intervals are shown to be more consistent with theoretical expectations than the confidence intervals implied by a normal GARCH model. This should be expected given the Monte Carlo Simulation results presented in the previous section.

5. Conclusions and Recommendations

A main conclusion from this research is that one must be skeptical of using the standard normal-error GARCH model when there is evidence of conditional error-term non-normality. The Monte Carlo simulations suggest that the IHS-GARCH model proposed in this study could perform better than the normal-error GARCH model under a variety of non-normal underlying error-term distributions. The RMSE's of the IHS-GARCH estimators for the slope and for the GARCH process parameters are substantially smaller than the RMSE's of the corresponding normal GARCH estimators, for all underlying non-normal error-term distributions and sample sizes evaluated. Under error-term non-normality the IHS-GARCH is a more efficient estimator for these three parameters. Efficient slope-parameter estimators are obviously desirable in applied modeling/forecasting work.

The inefficiency of the normal GARCH estimators for α_1 and B_1 translates into a relatively large number of unwarranted GARCH rejections due to apparent non-stationarity and zero-valued estimates for α_1 and B_1. These rejections would prevent modelers from identifying the correct error-term autocorrelation structure. Further, with the normal GARCH, the usual standard error estimates substantially underestimate the RMSE's (i.e. the true standard errors) of the estimators for the slope and for the GARCH process parameters, providing a false sense of security about the precision with which these parameters have been estimated and invalidating any statistical test based on these standard error estimates. Such a problem, which has clear implications for applied modeling work, does not diminish with sample size. The proposed IHS-GARCH model nearly solves this problem regardless of the sample size.

Both the normal and the IHS-GARCH are biased estimators for the GARCH process parameters, even when the underlying error is IHS, in which case the IHS-GARCH is the true MLE. The magnitude of the bias is noticeably less with the IHS-GARCH, especially at small sample sizes. The magnitude of the bias decreases with sample size and, when B_1 is estimated using the IHS-GARCH, it becomes very small at n=2500. However, at this largest sample size, the 100-sample averages of the α_1 estimates from both the normal and the IHS-GARCH still depart from the true parameter values, even when the underlying error-term distribution is IHS.

The mainstream empirical example demonstrates that some conditional error-term distributions encountered in applied research are not only non-normal but also asymmetric, and therefore the need for the proposed modeling technique. Some of the conclusions from the Monte Carlo simulations, such as increased parameter estimation efficiency, are clearly reflected in the empirical example. The example also illustrates other practical advantages of modeling conditional error-term distribution non-normality, when present, such as more theoretically consistent confidence intervals for the GARCH predictions.

Finally, when comparing the performance of the proposed IHS-GARCH model under IHS errors versus its performance under the other non-normal error-term distributions considered in the Monte Carlo simulation, we conclude that the main advantage of using the true MLE is increased efficiency in the estimation of the slope and of the GARCH process parameters. Since, knowledge of the true MLE is impossible, in practice, the next best alternative is to develop other non-normal-error GARCH models based on flexible non-normal distributions and use testing procedures for non-nested hypotheses (Quang, 1989) to identify the GARCH model that best approximates the true data generating process. Given that the expanded IHS distribution used as a basis for the IHS-GARCH can accommodate any mean and variance together with any skewness-kurtosis combination below the log-normal line, emphasis should be placed on alternative distributions that can do the same above the log-normal line.

6. References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(1986):307-327.

Bollerslev, T. (1986). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69(1987):542-546.

D'Agostino, R.B., A. Belanger, and R.B. D'Agostino Jr. (1990). A Suggestion for using powerful and informative tests of normality. *The American Statistician* 44(4):316-321.

Engle, R.F. and D. Kraft (1983). Multi-period forecast error variances of inflation estimated from ARCH models, in: A. Zellner, ed., Applied time series analysis of economic data (Bureau of the Census, Washington D.C.) 293-302.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50:987-1008.

Hsieh, D.A., and C.F. Manski (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Annals of Statistics*, 15:541-551.

Johnson, N.L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*, New York: Wiley & Sons.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149-176.

McDonald, J.B., and S.B. White (1993). A comparison of some robust, adaptive, and partially adaptive estimators of regression models. *Econometric Reviews*, 12(1):103-124.

McDonald, J.B., and W.K. Newey (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory*, 4:428-457.

Mood, A.M., F.A. Graybill, and D.C. Boes (1974). *Introduction to the Theory of Statistics*, New York: McGraw-Hill.

Newey, W.K. (1988). Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics*, 38:301-339.

Phillips, R.F. (1998). Partially adaptive estimation via a normal mixture: Some further Monte Carlo evidence. *Communications in Statistics*, 27(1):107-114.

Quang, H.V. (1989). Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* 57(2):307-333.

Ramirez O.A and J.S. Shonkwiler (2000). A partially adaptive estimator for regression models with non-spherical errors. *Econometric Reviews* (second submission).

U.S. Department of Commerce, Bureau of Economic Analysis (http://www.stls.frb.org /fred/data/gdp/gnpdef as of June 3, 2000).

Yang, S.R. and B.W. Brorsen (1992). Non-linear dynamics of daily cash prices. *American Journal of Agricultural Economics* 74(1992):706-715.

Figure 1: 80% Confidence Intervals for the Inflation Rate Predictions vs. Data under the Normal GARCH Model

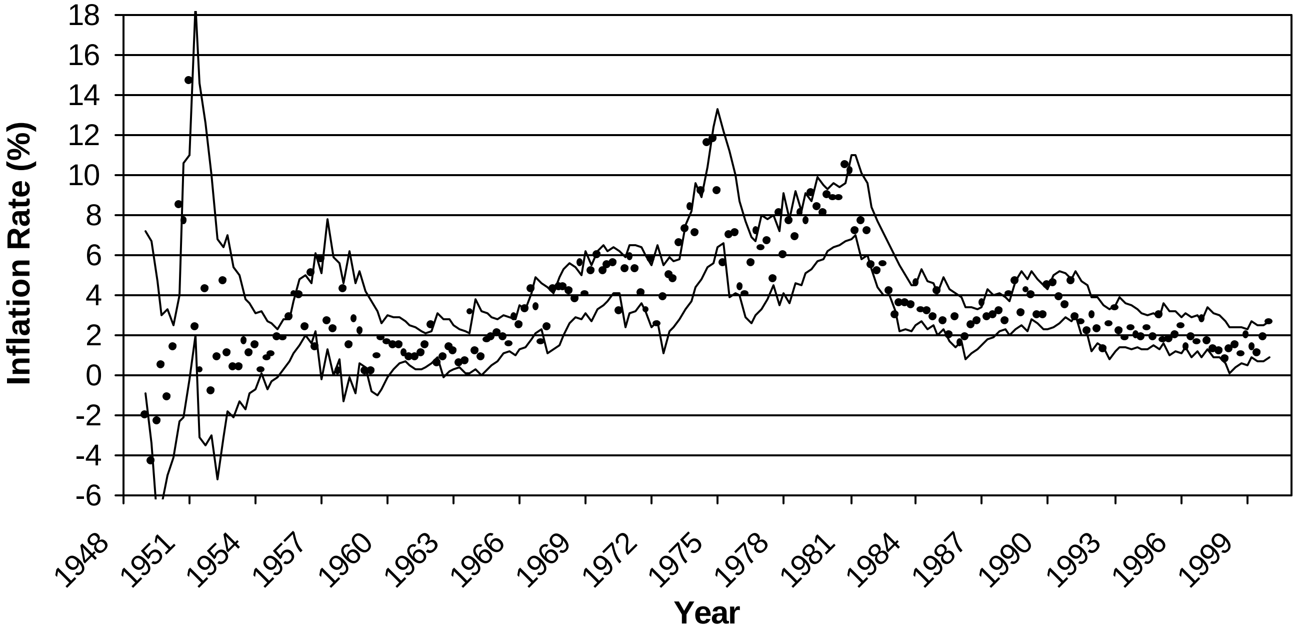

Figure 2: 80% Confidence Intervals for the Inflation Rate Predictions vs. Data under the IHS-GARCH Model