ECONOMETRIC REVIEWS, 5(1), 1-50 (1986)

MODELLING THE PERSISTENCE OF CONDITIONAL VARIANCES

Robert F. Engle and Tim Bollerslev

Department of Economics, UCSD
La Jolla, CA 92093

and

Department of Economics, Northwestern University
Evanston, IL 60201

Key Words and Phrases: *Autoregressive conditional heteroskedasticity;* ARCH; GARCH; *integration in variance; Student-t distribution; conditional kurtosis; time aggregation; non-linear conditional heteroskedasticity; asset pricing; exchange rate determination.*

ABSTRACT

This paper will discuss the current research in building models of conditional variances using the Autoregressive Conditional Heteroskedastic (ARCH) and Generalized ARCH (GARCH) formulations. The discussion will be motivated by a simple asset pricing theory which is particularly appropriate for examining futures contracts with risk averse agents. A new class of models defined to be integrated in variance is then introduced. This new class of models includes the variance analogue of a unit root in the mean as a special case. The models are argued to be both theoretically important for the asset pricing models and empirically relevant. The conditional density is then generalized from a normal to a Student-t with unknown degrees of freedom. By estimating the degrees of freedom, implications about the conditional kurtosis of these models and time aggregated models can be drawn. A further generalization allows the conditional variance to be a non-linear function of the squared innovations. Throughout, empirical estimates of the logarithm of the exchange rate between the U.S. dollar and the Swiss franc are presented to illustrate the models.

0747-4938/86/0501-0001$3.50/0

TABLE OF CONTENTS

I. INTRODUCTION

The economic theory of behavior under uncertainty forms the basis for much of modern finance and monetary theory. An agent must make a decision based upon the distribution of a random variable some time in the future. In many rational expectations models it is assumed that only the mean of the conditional distribution affects the decision, however for more general utility functions and risk averse agents, a measure of dispersion will also be of primary importance. Conventional econometric methods have not been responsive to the need to develop quantitative measures of risk and uncertainty. This is particularly clear in time series analysis where the very convenient assumption of linear covariance stationary models implies that most measures of uncertainty would remain constant over time.

This paper will discuss the current research in building models of conditional variances using the Autoregressive Conditional Heteroskedastic (ARCH) and Generalized ARCH (GARCH) formulations. The paper will motivate the discussion by introducing a simple asset pricing theory which is particularly appropriate for examining forward and futures contracts with risk averse agents. A new class of models defined to be integrated in variance is then introduced. This new class of models include the variance analogue of a unit root in the mean as a special case. The models are argued to be of both theoretical importance for the asset pricing models as well as empirically relevant. The conditional density is then generalized from a normal to a Student-t with unknown degrees of freedom. By estimating the degrees of freedom implications about the conditional kurtosis of these models and time aggregated models can be drawn. A further generalization allows the conditional heteroskedasticity to be a non-linear function of the squared innovations. Throughout, empirical estimates of the logarithm of the exchange rate between the U.S. dollar and the Swiss franc are presented to illustrate the models. The data are weekly observations on the Wednesday spot closing price and were generously provided by Frank Diebold. The same data set was examined in Diebold and Nerlove (1985).

II. THE SIMPLEST ARCH MODEL

To illustrate the unsatisfactory nature of time series and standard econometric models for modelling risk and uncertainty, consider the first order autoregressive model:

(1) $$y_t = \phi\, y_{t-1} + \epsilon_t\,, \quad \epsilon_t \text{ i.i.d.},$$

where

$$E(\epsilon_t) = 0, \quad V(\epsilon_t) = \sigma^2$$

and for stationarity $|\phi| < 1$. The expected value of y_{t+1} is simply zero but the conditional mean of y_{t+1} given the past information equals:[1]

$$E_t(y_{t+1}) = \phi\, y_t$$

which obviously is a random variable depending on the information set at time t. Similarly, the variance and the conditional variance can be expressed as:

$$V(y_{t+1}) = \sigma^2/(1-\phi^2)$$

$$V_t(y_{t+1}) = E_t[y_{t+1} - E_t(y_{t+1})]^2 = \sigma^2$$

Therefore, even though the conditional and unconditional variances are different, unless $\phi = 0$, they are both constants. By repeated substitution in (1) we get

[1]Of course, here and in what follows the conditional expectations are only defined almost surely. Without the loss of any essential results we shall simply ignore the technicalities arrising from this.

(2) $$y_{t+s} = \phi^s y_t + \sum_{i=1}^{s} \phi^{s-i} \epsilon_{t+i}$$

where the first term equals the conditional expectation of y_{t+s} at time t and the second term gives the corresponding forecast error. Thus, the conditional variance of y_t s steps ahead can be computed by

(3) $$V_t(y_{t+s}) = \sigma^2 \sum_{i=1}^{s} \phi^{2(s-i)}$$

so that although the conditional variance depends upon the forecast horizon it does not depend on the available information set or the forecast origin, t.

In more general models, the conditional variance s steps into the future would depend on the available information set and could therefore change over time. One might claim that the success of time series models is attributable to the use of the conditional mean for forecasting rather than the unconditional mean. Presumably similar gains are available for variances from using more sophisticated models of the conditional variances. The ARCH model introduced in Engle (1982) was designed to generalize the class of models used in time series analysis with exactly this point in mind. The conditional variance as well as the conditional mean were parameterized as functions of the available information set. While this allowed a vast increase in the potential models only a few of the parameterizations have been particularly useful. These are the ARCH(q) and the GARCH(p,q) models to be described below.

The simplest generalization of the AR(1) model given in (1) has an ARCH(1) model for the errors

$$y_t = \phi y_{t-1} + \epsilon_t$$

$$(4) \qquad E_{t-1}(\epsilon_t) = 0$$

$$V_{t-1}(\epsilon_t) = h_t = \omega + \alpha \epsilon_{t-1}^2$$

where $|\phi| < 1$, $\omega > 0$ and $\alpha \geq 0$.

The first order ARCH model described in (4) has several interesting properties. The disturbances are serially uncorrelated but not independent as they are related through second moments. If $\alpha < 1$, the unconditional variance of ϵ_t is given by:

$$V(\epsilon_t) = \sigma^2 = \omega (1-\alpha)^{-1} .$$

Using this notation the conditional variance of ϵ_t can therefore be rewritten as:

$$(5) \qquad h_t - \sigma^2 = \alpha (\epsilon_{t-1}^2 - \sigma^2)$$

so that the conditional variance will be above the unconditional variance whenever the squared surprise last period exceeds its unconditional expectation. The marginal distribution of y_t will be symmetric if the conditional distribution of ϵ_t is symmetric. If ϵ_t is conditionally normal, the fourth unconditional moment of ϵ_t will exceed $3\sigma^4$ so that the marginal distribution of ϵ_t exhibits fatter tails than the normal.

It follows immediately from (4) that the conditional mean and the conditional variance of y_{t+1} evaluated at time t may be expressed as:

$$E_t(y_{t+1}) = \phi y_t$$

$$V_t(y_{t+1}) = h_{t+1} = \omega + \alpha(y_t - \phi y_{t-1})^2$$

so that both the conditional mean and the conditional variance of the one-step ahead forecast depend on the available information set. In particular the conditional variance is increased by either large positive or negative surprises in y_t.

Multi-step forecasts from the AR(1)-ARCH(1) model can be computed using the expression for y_{t+s} given in (2) leading to:

$$V_t(y_{t+s}) = \sum_{i=1}^{s} \phi^{2(s-i)} E_t(h_{t+s}) .$$

Provided $\alpha < 1$ and $i > 2$, it follows from (5) by the law of iterated expectations that:

$$E_t(h_{t+i}) = \sigma^2 + \alpha E_t(h_{t+i-1} - \sigma^2)$$

and the multi-step conditional variance equals:

$$(6) \qquad V_t(y_{t+s}) = \sigma^2 \sum_{i=0}^{s-1} \phi^{2i} + \alpha^{s-1}(h_{t+1}-\sigma^2) \sum_{i=0}^{s-i} \phi^{2i} \alpha^{-i}.$$

Thus, the conditional variance is clearly not independent of the current information set, but for long term forecasts and α less than one the dependence on $(h_{t+1} - \sigma^2)$ becomes negligible and the expression in (6) is well approximated by (3).

Assuming ϵ_t to be conditionally normal distributed, it follows by the prediction error decomposition, that for a sample of T observations the log likelihood function is, apart from some initial conditions, given by

$$(7) \qquad \text{LogL} = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}(\log(h_t) + \epsilon_t^2 h_t^{-1}) .$$

Expressions for the gradient, the Hessian and Fishers information matrix are given in Engle (1982), along with a discussion of methods for maximization of (7).

All of the above arguments and calculations extend in a straight forward manner to the more general ARCH(q) model in which h_t depends linearly on q lagged values of ϵ_t^2,

$$V_{t-1}(\epsilon_t) = h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

where now $\alpha_i \geq 0$, $i = 1,\dots,q$. For a more detailed discussion of the ARCH(q) model the reader is referred to the original paper by Engle (1982).

III. SURVEY OF APPLICATIONS AND EXTENSIONS

In a series of papers, the ARCH model has been analyzed, generalized, extended to the multivariate context, and used to

test for time varying risk premia in the term structure of interest rates and in other financial markets. These papers include Engle (1983) and Engle and Kraft (1983) where a measure of the variance of inflation are given. The ARCH model is extended to a multivariate framework in Kraft and Engle (1982). In Engle, Granger and Kraft (1984) and Granger, Robins and Engle (1984) bivariate ARCH models of inflation with changing covariances as well as variances are constructed. Engle, Brown and Stern (1984) and Granger and Engle (1984) examine the effectiveness of ARCH models for forecasting purposes. The power of ARCH tests and the finite sample properties of various ARCH estimators, are analyzed in Engle, Hendry and Trumble (1985) by means of Monte Carlo methods. In Bollerslev (1985a, 1985b) the Generalized ARCH or GARCH model is developed, and the GARCH model with conditionally Student-t distributed errors is studied in Bollerslev (1985c). Engle, Lilien and Robins (1985) and Bollerslev, Engle and Wooldridge (1985) examine the term structure of interest rates and a three asset Capital Asset Pricing Model (CAPM) to determine whether risk premia are varying over time. These papers introduce the ARCH in mean or ARCH-M model in a univariate and multivariate context respectively. Engle and Watson (1985) recast the GARCH-M model in a full information state space form.

In addition to this work, a variety of papers have begun to appear from different parts of the world. Without attempting an exhaustive list of these, particularly interesting are papers by Milhøj (1985a) who develops far more general moment conditions than those in the original Engle (1982) paper. Linnel-Nemec (1984a, 1984b) establishes stationarity and ergodicity of ARCH models. Pantula (1984) and Weiss (1982) derive the limiting distribution of ARCH estimators in more general contexts. Pagan, Hall and Trivedi (1983), Weiss (1984) and Coulson and Robins (1985) provide empirical time series examples of ARCH and related models of changing variances. Domowitz and Hakkio (1985), Diebold and Pauly (1985), Diebold and Nerlove (1985), Milhøj (1985b), Hsieh (1985) and McCurdy and Morgan (1985), apply the ARCH, the ARCH-M and the GARCH model to the foreign exchange market. Amsler (1984a, 1984b) investigate whether using the risk premia estimated by ARCH models will make long bonds satisfiy the Shiller variance bounds. Poterba and Summers (1984) derive a pricing formula for stock market prices in the spirit of the asset pricing formulas presented in this paper. The price is related to its own variance, which is modelled as a simple AR(1) process. Similar ideas are employed in the paper by French, Schwert and Stambaugh (1985). Blanchard and Watson (1984) and Bodie, Kane and McDonald (1983, 1984) present evidence that macroeconomic and financial time series models can usefully be reformulated as a form of multivariate ARCH processes.

These applications can all be viewed in a common econometric framework with a focus on modelling the behavior of economic agents to risk in a time series context. The general model is a simultaneous equations system where the elements of the ARCH disturbance covariance matrix also appear as regressors. To establish notation, let y_t be a vector of endogenous random variables

and ψ_t the information set available at time t. Also, let μ_t and H_t denote the conditional mean and the conditional variance of y_t respectively:

$$\mu_t = E[y_t|\psi_{t-1}]$$

$$H_t = E[(y_t-\mu_t)(y_t-\mu_t)'|\psi_{t-1}] .$$

Finally, let X_t be a vector of weakly exogenous and lagged dependent variables in the sense of Engle, Hendry and Richard (1983). The conditional mean and variance of the general model can then be parameterized as:

$$\mu_t = X_t\beta + \text{vech}(H_t)\ \delta$$

$$H_t = A + (I\otimes X_t)'C\ (I\otimes X_t) .$$

where vech(H_t) denotes the vector of all the unique elements of H_t obtained by stacking the lower triangle of H_t, A is a positive definite matrix and C is a positive semi-definite matrix. This class of models represents the reduced form of a structural simultaneous equations system where the mean of the observed behavior is affected by the covariance matrix of all observed endogenous variables. Such a system could be described as:

$$B\ y_t + D\ \mu_t + F\ \text{vech}(H_t) + G\ X_t = \epsilon_t$$

which is closely related to Wallis (1980)'s rational expectations version of a simultaneous system except that in this case both conditional means and conditional variances and covariances enter the structural relation.

If in the above reduced form $\delta=0$, this is simply a multivariate heteroskedasticity model. If in addition X_t enters the heteroskedasticity function only through squares and cross products of $\epsilon_t = y_t - \mu_t$, the multivariate ARCH model of Kraft and Engle (1982) is obtained. The multivariate GARCH model of Bollerslev, Engle and Wooldridge (1985) has H_t depending on past squares of ϵ_t's and past values of H_t's. In effect, this corresponds to a different lag structure on past squared ϵ_t's.

In the univariate context with $\delta \neq 0$, the ARCH-M model is obtained. Engle, Lilien and Robins (1985) uses this type of model behavior in modelling the term structure of interest rates where the variance of the excess holding yield on a long bond is a determinant of the expected return on that asset. The ARCH-M model is extended to a multivariate setting in Bollerslev, Engle and Wooldridge (1985), where a Capital Asset Pricing Model (CAPM) with time varying betas is estimated. The expected return on an asset is proportional to the covariance with the market return, that is the CAPM gives the restriction

$$\mu_t = \delta H_t w_t$$

where w_t is the vector of value weights for all the assets, assumed to be in the information set. The results for a model with three risky assets: bills, bonds and stocks, are promising. There is strong evidence for a time varying conditional covariance matrix for the asset returns. The coefficient of aggregate relative risk aversion, δ, is statistically significant. The trivariate model in which the risk premium is associated with the covariance with the market outperforms a set of univariate models where the premia depend only on the own variances. The model, however, does fail some of the more demanding specification tests suggesting further work along these lines might be fruitful.

IV. SIMPLE ASSET PRICING THEORIES

To motivate the new class of models called integrated in variance to be defined below, a very simple illustrative model of asset pricing will be presented. It has a familiar form and can be solved to give previously unreported pricing equations.

Models of asset pricing involve agents maximizing expected utility over uncertain future events. The competitive outcome sets a price for the asset today which equates supply with demand. For example suppose a representative agent must allocate his wealth, W_t, between shares of a risky asset q_t at a price p_t and those of a sure asset x_t with a price equal to 1. The shares of the risky asset will be worth y_{t+1} each at the end of the period, and the sure asset will be worth $r_t x_t$, where r_t denotes one plus the risk free rate. Note, that if there are no dividends $y_{t+1} = p_{t+1}$. Also, suppose the representative agent has a mean variance utility function in end of period wealth, $q_t y_{t+1} + r_t x_t$. With these assumptions the allocation problem becomes:

$$(8) \qquad \max_{q_t} \; 2E_t(q_t y_{t+1} + r_t x_t) - \gamma_t V_t(q_t y_{t+1})$$
$$\text{s.t. } W_t = x_t + p_t q_t$$

which has the solution

$$p_t = \frac{1}{r_t} E_t(y_{t+1}) - \frac{\gamma_t q_t}{r_t} V_t(y_{t+1}) . \tag{9}$$

If the outstanding stock of the risky asset is fixed at q, then (9) describes the asset pricing model. Typically, W_t, r_t and γ_t are also taken as constants in such a formulation, even though this is not required by the model. However, for ease of exposition we shall treat W, q, γ and r as constants in what follows.

It is worth considering exactly what is implied by the relation given in (9). Both the conditional mean and the conditional variance are functions of the conditioning information set which includes p_t. The situation is therefore similar to the Walrasian auctioneer who must declare a price known to all agents that clears the market. The equation therefore has no error term unless there are other sources of uncertainty and these may be problematic from a theoretical point of view. We are indebted to Roman Frydman for this observation.

If now the risky asset is interpreted as a forward contract for delivery in s periods, the price a pure speculator characterized by (8) would be willing to pay is simply

$$p_t = r^{-s}[E_t(y_{t+s}) - \delta V_t(y_{t+s})] \tag{10}$$

where r^{-s} gives the discounted present value at the sure rate r and $\delta = \gamma q$. Note, p_t is the price paid in full at time t for delivery at time t+s. Of course, $p_{t+s} = y_{t+s}$ by definition. In the foreign exchange market this model relates the price of a forward contract for delivery in s periods to the expected future value of the spot exchange rate y_{t+s} and the variance of that rate discounted at the sure domestic interest rate r. Typically it is assumed that δ=0 so that there is no risk premium and the forward rate is an unbiased predictor of the future spot rate corrected for interest rate differentials.

A simple redating of the model shows that the price of the forward contract at time t+1 for s $\geq$ 2 periods remaining to maturity can be expressed as:

$$p_{t+1} = r^{1-s} [E_{t+1}(y_{t+s}) - \delta V_{t+1}(y_{t+s})] .$$

Taking expectations at time t, multiplying by r^{-1} and subtracting from (10) gives:

$$p_t = r^{-1} E_t(p_{t+1}) - \delta r^{-s}[V_t(y_{t+s}) - E_t(V_{t+1}(y_{t+s}))] . \tag{11}$$

Now, suppose y_t can be represented by an infinite moving average process where the increments are uncorrelated but have time varying conditionnal variance h_t,

$$y_t = \sum_{i=1}^{\infty} \theta_i \epsilon_{t-i} = \theta(L)\epsilon_t \tag{12}$$

$$V_t(y_{t+1}) = V_t(\epsilon_{t+1}) = h_{t+1} .$$

It follows that

$$(13)\qquad V_t(y_{t+s}) = E_t\Big(\sum_{i=1}^{s} \theta_{s-i}\,\epsilon_{t+i}\Big)^2 = \sum_{i=1}^{s} \theta_{s-i}^2\, E_t(h_{t+i}) .$$

Consequently the expression in square brackets in (11) is simply $\theta_{s-1}^2 h_{t+1}$ leading to:

$$(14)\qquad p_t = r^{-1} E_t(p_{t+1}) - \delta r^{-s}\theta_{s-1}^2\, h_{t+1}$$

which is the familiar formula for a one period holding yield with the explicit calculation of the effect of the changing variance of y_{t+s} for a risk averse agent. Notice, however, that the expectation of the final spot price no longer enters this equation directly, but only through its effect on $E_t(p_{t+1})$. Also, if θ_{s-1} becomes small for large s, the only relevant information about the risk of the contract is already incorporated in that expectation.

In the simple model above the only source of uncertainty derives from the future spot price to which the contract relates. In many other situations however, there is a flow of uncertain distributions which accrue to the owner of the asset. For example, the price of a share of a stock is presumably determined by the present discounted value of the expected dividend stream. The rate of interest on a long bond is conjectured by the expectations hypothesis to be the present discounted value of the sequence of short term rates expected during the life of the bond. The value of a fruit tree is the present discounted value of the expected

value of the fruit. In each case, the model is based on the assumption of risk neutral investors and exogenously given consumption.

However, risk averse investors would everything else equal pay less for streams of income the more uncertain the streams. The precise fashion in which the variability of future payoffs enters the asset pricing formulation depends among other things upon the utility function of the agents and the intertemporal substitutability of the payouts. The two simplest formulations are the variance of the discounted payoffs and the discount of the variance of the payoffs, i.e. either of the following two forms:

$$(15)\qquad p_t = \sum_{s=1}^{\infty} r^{-s}[E_t(y_{t+s}) - \delta\, V_t(y_{t+s})]$$

or

$$p_t = \sum_{s=1}^{\infty} r^{-s}E_t(y_{t+s}) - \delta\, V_t\Big(\sum_{s=1}^{\infty} r^{-s}y_{t+s}\Big)$$

where y_{t+1}, $y_{t+2}, \ldots$ refer to the future income stream generated by the asset. The latter of these two can also be expressed as

$$(16)\qquad p_t = \sum_{s=1}^{\infty} r^{-s}E_t(y_{t+s}) - \delta \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \omega_{ij}\mathrm{Cov}(y_{t+i}, y_{t+j}).$$

Under the same set of assumptions as in (12) about the process generating y_t (15) can be converted to a holding yield expression:

$$p_t = r^{-1} [E_t(p_{t+1}) + E_t(y_{t+1}) - \delta\lambda\, h_{t+1}] \tag{17}$$

where λ depends upon the polynomial $\theta(L)$ and the risk free rate r. Assuming h_{t+1} is generated by an integrated GARCH(1,1) model, to be defined below, a similar expression holds for (16). The important feature of this result is that $V_t(p_{t+1})$ does not independently enter the asset pricing equation. The variance of next periods payout is sufficient information for risk averse agents. This differs from the Summers and Poterba (1984) formulation where $V_t(p_{t+1})$ enters the pricing equation explicitly.

Examining expressions (11) through (17) it is clear that if $\delta \neq 0$, the conditional variance of y_t in the future will affect the price of the asset today. The extent of the effect cannot be seen directly from (14) or (17) which are stated in terms of future expectations about the price, but can be seen from the discount expressions (10), (15) or (16). If variances can be forecast as in the ARCH example in the preceeding section, then the current information on y_t and the current conditional variance will have an effect on the current price. However, the size of the effect will depend upon the persistence of the variance. That is, it will depend upon how important current information is in predicting future variances. This question is of great importance since it explains how new information influences current asset prices.

V. MULTI-STEP FORECASTING WITH THE GARCH MODEL

Bollerslev (1985a) has recently introduced the Generalized ARCH or GARCH model which incorporates in a simple form many of the properties which have been found useful in ARCH estimation. A random variable ϵ_t is said to follow a GARCH(p,q) process if

$$E_{t-1}(\epsilon_t) = 0$$

$$V_{t-1}(\epsilon_t) = h_t = \omega + \sum_{i=1}^{q} \alpha_i\, \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i\, h_{t-i} \tag{18}$$

$$= \omega + \alpha(L)\epsilon_t^2 + \beta(L)h_t$$

where $\omega > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for all i. This formulation generalizes the ARCH(q) model by allowing non-zero β_i's. Simple substitution reveals that the GARCH model is simply an infinite order ARCH model with exponentially decaying weights for larger lags. Thus, a low order GARCH model may have properties similar to high order ARCH models without the problems of estimating many parameters subject to non-negativity constraints. In fact, the ARCH models estimated in Engle (1983), Engle and Kraft (1983), and Engle, Lilien and Robins (1985), impose linearly declining weights in the α_i's so that the only free parameters are q and the sum of the weights. Selection of q can be addressed by model selection techniques but is somewhat awkward. Thus the GARCH model appears to be a natural and simple generalization of the ARCH model, and empirical evidence suggests that it fits as well or even better than the ARCH model with linearly declining weights with roughly the same mean lag, see Bollerslev (1985a) for more details.

Sastry Pantula has pointed out that the GARCH model can be rewritten in an alternative and from a theoretical time series point of view maybe more interpretable form. Let $\upsilon_t \equiv \epsilon_t^2 - h_t$, then the conditional variance in expression (18) may be rearranged as:

$$(19) \qquad \epsilon_t^2 = \omega + \sum_{i=1}^{m} (\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i \upsilon_{t-i} + \upsilon_t$$

where $m = \max\{p,q\}$, $\alpha_i \equiv 0$ for $i > q$ and $\beta_i \equiv 0$ for $i > p$. This reveals that ϵ_t^2 follows an ARMA(m,p) process with serially uncorrelated innovations υ_t. Note, however, the peculiar feature that the innovations υ_t have different support in each time period and consequently are heteroskedastic with a very complicated distribution. It is clear that the autoregressive part of the process is given by the sum of the α_i's and the β_i's, and the moving average part is characterized by minus the β_i coefficients. Such an analogy has been used by Bollerslev (1985b) to motivate the use of the autocorrelogram and partial autocorrelogram for ϵ_t^2 in model identification.

If ϵ_t^2 is generated by a GARCH(p,q) model, the forecast of the conditional variances is a non-trivial function of the information set today, just as in the ARCH(1) model illustrated in Section II. This allows a solution to the rational expectations models of the preceeding section. The conditional variance s steps ahead can be written:

$$(20) \qquad h_{t+s} = \omega + \sum_{i=1}^{n} [\alpha_i \epsilon_{t+s-i}^2 + \beta_i h_{t+s-i}] + \sum_{i=s}^{m} [\alpha_i \epsilon_{t+s-i}^2 + \beta_i h_{t+s-i}]$$

where $n = \min\{m, s-1\}$ and by definition summation from 1 to 0 and from $s > m$ to m both are equal to zero. Thus,

$$(21) \qquad E_t(h_{t+s}) = \omega + \sum_{i=1}^{n} [(\alpha_i + \beta_i)E_t(h_{t+s-i})] + \sum_{i=s}^{m} [\alpha_i \epsilon_{t+s-i}^2 + \beta_i h_{t+s-i}]$$

which gives a recursive solution for the conditional expectations. In particular for the GARCH(1,1) model and $s \geq 2$ this expression reduces to

$$(22) \qquad E_t(h_{t+s}) = \omega + (\alpha + \beta)\, E_t(h_{t+s-1}).$$

In the stationary case where all the roots of the autoregressive polynomial $\alpha(z)+\beta(z)=0$ lie outside the unit circle and therefore the sum of the α_i's and the β_i's is less than one, the unconditional variance of ϵ_t is given by:

$$E(\epsilon_t^2) = \sigma^2 = \omega[1 - \sum_{i=1}^{m} (\alpha_i + \beta_i)]^{-1} .$$

By simple substitution equation (21) becomes

$$E_t(h_{t+s}) = \sigma^2 + \sum_{i=1}^{n} [(\alpha_i + \beta_i)(E_t(h_{t+s-i}) - \sigma^2)] + \sum_{i=s}^{m} [\alpha_i \epsilon^2_{t+s-i} + \beta_i h_{t+s-i}] \tag{23}$$

and it follows that $E_t(h_{t+s}) \to \sigma^2$ as $s \to \infty$ when the process is stationary.

To illustrate the ideas discussed above consider the time series of weekly data on the exchange rate between the U.S. dollar and the Swiss franc from July 1973 through August 1985 for a total of 632 observations. The data are interbank closing spot bid prices on Wednesdays taken from the International Monetary Markets Yearbook. This particular series is chosen for the analysis because the first differences of the logarithm show no significant serial correlation in the mean, a finding also reported in the paper by Diebold and Nerlove (1985), who kindly provided the data. Thus, attention can simply be focused upon the variance models.

A plot of the first differences of the logarithm of the series is given in Figure 1. It follows, that even though the series seems to be uncorrelated over time, the observations are clearly not independent. There is a tendency for large changes to be followed by large changes but of unpredictable sign. The modified Box-Pierce test statistic for up to tenth order serial correlation takes the value Q(10) = 9.127 corresponding to the .480

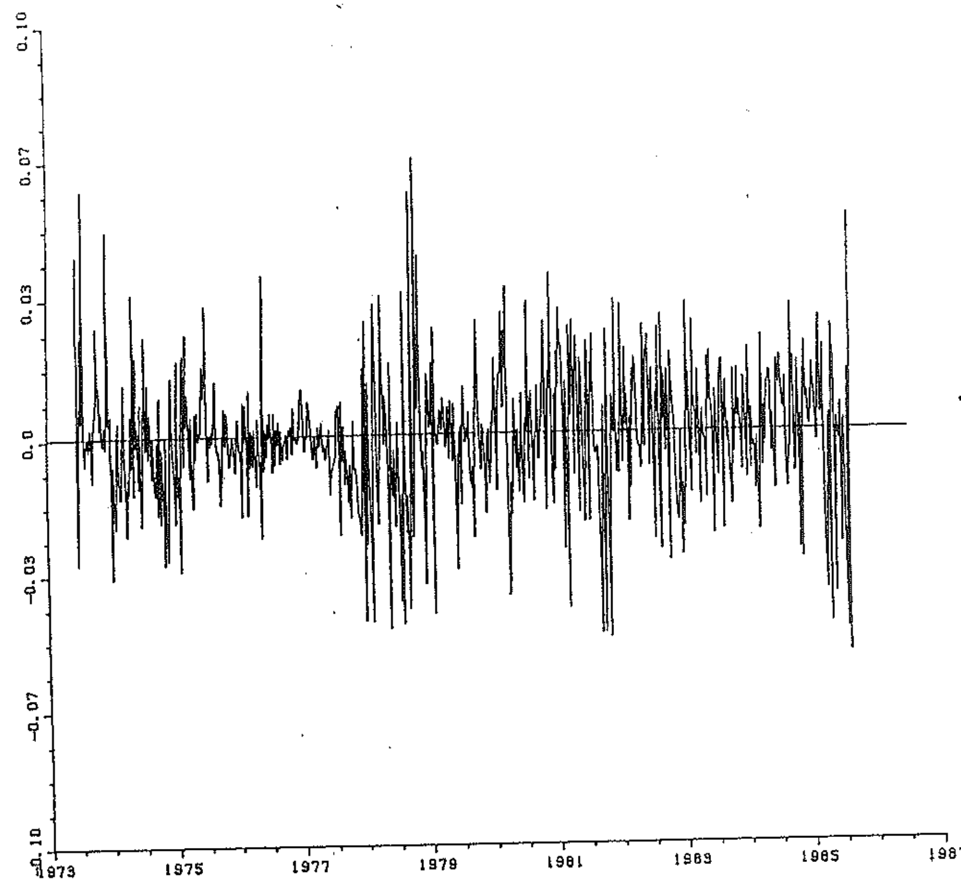

FIGURE 1

First Differences of Logarithm of U.S. Dollar/Swiss Franc Exchange Rate

fractile in a $\chi^2(10)$ distribution, whereas the test statistic for the first ten squared changes equals $Q^2(10) = 99.435$ which is highly significant at any level. For a precise definition of the Box-Pierce test statistic and a discussion of its applicability in testing for absence of serial correlation in squared residuals see McLeod and Li (1983). The first ten autocorrelations, ρ_k, and partial autocorrelations, ϕ_{kk}, for the squared differences are given in Table I.

TABLE I

Correlation Structure for Squared Log Differences

k	1	2	3	4	5	6	7	8	9	10
ρ_k	.114	.117	.169	.212	.083	.110	.162	.012	.102	-.003
ϕ_{kk}	.114	.106	.148	.178	.024	.049	.094	-.070	.053	-.077

Judged by their asymptotic standard error, .040, both the autocorrelations and the partial autocorrelations die out fairly slowly. Thus, taking into account the small sample variability of the correlation estimates, cf. Bollerslev (1985b), a GARCH(1,1) model with $\alpha + \beta$ close to one seems a reasonable first model candidate.

Assuming conditional normality maximum likelihood estimation of the GARCH(1,1) model yields:[2]

$$\log(y_{t+1}/y_t) = \epsilon_{t+1}$$

$$\epsilon_{t+1}|\psi_t \sim N(0,h_{t+1})$$

[2]To start up the recursions defined by (7) and (18) we need presample values for ϵ_0^2 and h_0. A natural choice is given by the sample analogue $\frac{1}{T}\sum \epsilon_t^2$.

$$h_{t+1} = \underset{(1.36\cdot 10^{-5})}{3.27\cdot 10^{-5}} + \underset{(.011)}{.996}\epsilon_t^2 \underset{(.011)}{-.871}(\epsilon_t^2 - h_t) \quad (24)$$

$$\log L = 1750.061$$

where the estimates in parentheses are asympototic standard errors obtained from the last auxilary regression in the Berndt, Hall, Hall and Hausman (1974) algorithm used to maximize the likelihood function.[3] Note, the estimated model in (24) does not have a finite unconditional fourth moment which may undermined the validity of the correlation structure for ϵ_t^2 as an identification tool. It is clear from the parameterization in (24) that the sum of α and β is very close to one, and that a conventional Wald type test would accept the hypothesis $\alpha + \beta = 1$. Multi-step forecast from (24) would approach the unconditional variance quite slowly as powers of .996 approach zero. The estimated mean lag of this variance expression, $1/(1-\beta)$, equals 7.752 or about eight weeks.

A series of Lagrange Multiplier tests for higher order ARCH, or equivalently GARCH, and moving average terms in ϵ_t were applied to determine the adequacy of the model specification. None of these tests detects model failure, which is quite surprising with over 600 observations. For instance the test for up to fourth order ARCH in addition to the GARCH(1,1) model equals .568, or .096 in a $\chi^2(3)$ distribution. The LM test for a constant in the

[3]Due to the complexity of the analytical derivatives for the GARCH model, numerical derivatives were used here and in what follows.

mean, i.e. a random walk with drift, equals .196, and the test for up to fourth order serial correlation takes the value 3.628, none of which are significant at traditional levels:

VI. INTEGRATED VARIANCE MODELS

The estimate for the GARCH(1,1) model as well as the theoretical expression in (23) suggests consideration of a particular class of GARCH models which have the property that the multi-step forecasts of the variance do not approach the unconditional variance. A GARCH model with this property is here defined to be integrated in variance.

DEFINITION

The GARCH(p,q) process characterized by the first two conditional moments:

$$E_{t-1}(\epsilon_t) = 0$$

$$E_{t-1}(\epsilon_t^2) = h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where $\omega \geq 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$ for all i and the polynomial

$$1 - \sum_{i=1}^{q} \alpha_i z^i - \sum_{i=1}^{p} \beta_i z^i = 0$$

has $d > 0$ unit root(s) and max{p,q}−d root(s) outside the unit circle is said to be:

(i) Integrated in variance of order d if $\omega = 0$.

(ii) Integrated in variance of order d with trend if $\omega > 0$.

It is obvious, that for the GARCH(p,q) process to be integrated in variance a necessary condition is that the α_i's and the β_i's sum to one. In such models the multistep variance conditional on ψ_t will always depend upon h_{t+1}, but for shorter forecast horizons additional information will also be important in forming optimal forecasts of the conditional variances. The integrated GARCH(p,q) models, both with or without trend, are therefore part of a wider class of models with a property called "persistent variance" in which the current information remains important for the forecasts of the conditional variances for all horizons.

To illustrate consider the integrated GARCH(1,1) model

$$h_{t+1} = \alpha\epsilon_t^2 + (1 - \alpha)h_t$$

where $0 < \alpha < 1$. From (22) it follows that for this particular model

$$E_t(h_{t+s}) = h_{t+1} . \qquad (25)$$

The conditional variance s steps in the future is the same as the conditional variance one step ahead for all horizons s. This model is obviously very closely related to the traditional random walk with a unit root in the conditional mean. In a random walk without a drift the prediction of the mean s steps in the future is just equal to the level today. Thus, the information today is important in forecasting forever. Shocks to to the system are permanent. In the same way shocks to the conditional variance in the integrated GARCH(1,1) model will not be forgotten.

In the integrated GARCH(1,1) model with trend

$$h_{t+1} = \omega + \alpha\epsilon_t^2 + (1 - \alpha)h_t$$

and therefore

$$E_t(h_{t+s}) = s\omega + h_{t+1} .$$

Again this model resembles very closely the traditional random walk with a drift in the mean. Even though the relative importance of h_{t+1} in the expression for the conditional variance s steps ahead decreases with the horizon, the effect is persistent.

The empirical plausibility of integrated GARCH models has already been established by the findings in Engle, Lilien and Robins (1985) and Bollerslev, Engle and Wooldridge (1985) that ARCH and GARCH models for interest rates typically exhibit parameters which are not in the stationary region. The estimates presented above for the U.S. dollar/Swiss franc exchange rate also suggests an integrated variance model.

Reestimating the GARCH(1,1) model imposing the restriction for integration in variance, yields:

$$(26) \qquad h_{t+1} = \underset{(.008)}{.090}\epsilon_t^2 + (1-.090)\, h_t$$

$$\log L = 1745.225.$$

The log likelihood is reduced below that of the unrestricted GARCH(1,1) by 4.836 which implies a likelihood ratio test statistic of 9.672 with two degrees of freedom. However, this reduction is almost entirely due to the restriction $\omega = 0$ since the t-statistic on this parameter equals 2.408. Indeed if the model is reestimated allowing a trend in the variance, the log likelihood is essentially unchanged relative to (23) and the parameter α is estimated to .118(.013). Here we leave out the intercept on the theoretical ground that a trend in the variance seems unreasonable for this particular model.

It follows from (19) that the integrated GARCH(1,1) model may be rewritten as

$$\Delta\epsilon_t^2 = -(1-\alpha)(\epsilon_{t-1}^2 - h_{t-1}) + (\epsilon_t^2 - h_t)$$
$$= -(1-\alpha)\,\upsilon_{t-1} + \upsilon_t$$

Therefore, even though the population autocorrelation and partial autocorrelations for $\Delta\epsilon_t^2$ don't exist if (26) is the true model, it is interesting to note that the sample correlation structure for $\Delta\epsilon_t^2$ resembles that of an MA(1) process with moving average parameter close to minus one fairly well, see Table II.

The appropriate procedure for testing for integration in variance is not yet clear. It is possible that all of the well known difficulties when testing for unit roots in the mean as discussed by Dickey and Fuller (1979) applies in this context as

TABLE II

Correlation Structure for Differences of Squared Log Differences

k	1	2	3	4	5	6	7	8	9	10
ρ_k	-.514	.002	-.018	.113	-.100	.007	.110	-.133	.114	-.142
ϕ_{kk}	-.514	-.357	-.312	-.105	-.123	.129	.061	-.049	.090	-.079

well, see Engle and Granger (1985) for some recent references on this subject. However in a few simple Monte Carlo runs, the standard test statistics appear to be relatively well behaved. The knife edge condition for the unit root in variance does not seem as severe as in the mean. Estimation by maximum likelihood essentially does a GLS correction and this may be sufficient to give standard asymptotic normality. At the moment, this is a conjecture.

Another complication when testing in (26) and in GARCH models in general involve the issues of one-sided tests, as the null hypothesis is on the boundary of the parameter space. Therefore, under the above conjecture, it may be simpler to construct Lagrange Multiplier tests for the adequacy of model (26). The troublesome trend term, ω, then appears to be in the wrong tail as the likelihood function increases in the negative direction from zero. Thus, the local LM test will accept the null for any one sided test. Similarly, the LM test against the unrestricted GARCH(1,1) with $\omega = 0$ equals 3.565 which is close to the 95 percent fractile in a $\chi^2(1)$ distribution. However, the test is in the

explosive direction and therefore if one chooses to consider only stationary alternatives it would also accept the null at any level. The tests for other forms of ARCH and serial correlation in the mean are again insignificant. In particular the test for up to fourth order ARCH equals 4.026 corresponding to the .597 fractile in a $\chi^2(4)$ distribution, and the test for up to fourth order serial correlation takes the value 2.164 corresponding to the .294 fractile in a $\chi^2(4)$ distribution.

VII. DISTRIBUTION OF MULTISTEP FORECAST ERRORS AND TIME AGGREGATION

Section V examined the second conditional moment of multistep forecasts. In this section, the properties of higher order moments will be developed. Engle (1982) and Bollerslev (1985a) established that the unconditional distribution of ARCH and GARCH processes with conditionally normal errors in general have fatter tails than the normal, i.e. they are leptokurtotic. They also showed that the conditions for existence of finite unconditional fourth moments restrict the maximum temporal dependence by restricting the admissable parameter space, and therefore in particular that all integrated GARCH models will have infinite unconditional fourth moments in the population. In this section we discuss the distribution of the fourth conditional moments.

Although most of the applications of the ARCH and GARCH models have assumed that the conditional distribution is normal, there is no necessity for this assumption. Recently, Bollerslev (1985c) has postulated that the conditional distribution is a standardized Student-t with unknown degrees of freedom υ. By estimating this parameter along with the other GARCH parameters it is possible not only to have a fat tailed unconditional distribution but also a fat tailed conditional distribution. If υ is twenty-five or larger the conditional distribution is essentially indistinguishable from the normal, while smaller values of υ give larger fourth conditional moments.

Consider now the GARCH(1,1) process characterized by the first four conditional moments:

$$
\begin{aligned}
&E_t(\epsilon_{t+1}) = 0\\
&E_t(\epsilon^2_{t+1}) = h_{t+1} = \omega + \alpha\epsilon^2_t + \beta h_t\\
&E_t(\epsilon^3_{t+1}) = 0\\
&E_t(\epsilon^4_{t+1}) = \kappa h^2_{t+1}.
\end{aligned}
\tag{27}
$$

Of course, for the GARCH(1,1) with conditionally normal errors $\kappa = 3$ and for the standardized Student t with $\upsilon > 4$ degree of freedom we have $\kappa = 3(\upsilon-2)(\upsilon-4)^{-1} > 3$ so the conditional distribution is also leptokurtic in that situation.

It follows that the fourth conditional moment of a forecast error s periods in the future equals

$$E_t(\epsilon^4_{t+s}) = \kappa E_t[E_{t+s-1}(h^2_{t+s})] = \kappa\, E_t(h^2_{t+s}),$$

and for $s \geq 2$

$$
\begin{aligned}
E_t(h^2_{t+s}) &= E_t[E_{t+s-2}(\omega + \alpha\epsilon^2_{t+s-1} + \beta h_{t+s-1})^2]\\
&= E_t[\alpha^2(\kappa-1)h^2_{t+s-1} + (\omega + (\alpha+\beta)h_{t+s-1})^2]\ .
\end{aligned}
\tag{28}
$$

In the stationary case, the series defined by this expression will converge whenever the unconditional fourth moment exists, i.e. if $\kappa\alpha^2 + 2\alpha\beta + \beta^2 < 1$ see Bollerslev (1985a).

For the integrated GARCH(1,1) model where $\omega = 0$ and $\beta = 1-\alpha$, (28) reduces to

$$E_t(h^2_{t+s}) = [1 + (\kappa-1)\alpha^2]^{s-1}\, h^2_{t+1}\ . \tag{29}$$

The further into the future this process is forecast, the larger the conditional kurtosis of the forecast errors. The rate of increase depends positively upon the kurtosis of the original process and upon the parameter α.

Of special interest for questions of time aggregation is the conditional kurtosis of the sum of future ϵ_t's. Define

$$\kappa^{(2)}_t = E_t(\epsilon_{t+1} + \epsilon_{t+2})^4 \,/\, [E_t(\epsilon_{t+1} + \epsilon_{t+2})^2]^2 \tag{30}$$

which will be the conditional kurtosis of a random walk process which is sampled every two periods. Therefore, if a conditional kurtosis of κ is found using weekly data, a kurtosis of $\kappa^{(2)}_t$ would be expected for biweekly data. Note, that in general the biweekly kurtosis will have a time subscript if the weekly model in (27) is true. Simple calculations yield:

$$E_t(\epsilon_{t+1}+\epsilon_{t+2})^4 = E_t(\epsilon^4_{t+1}) + E_t(\epsilon^4_{t+2}) + 6\, E_t(\epsilon^2_{t+1}\epsilon^2_{t+2})$$

$$= \kappa\, h^2_{t+1} + \kappa\, E_t(h^2_{t+2}) + 6\, E_t(\epsilon^2_{t+1}h_{t+2}) .$$

Using (28) and substituting back into (30) we get:

$$(31) \qquad \kappa_t^{(2)} = \kappa + \kappa \frac{h^2_{t+1}[(3-\kappa)(\frac{2\beta}{\kappa} - \alpha^2) + 2\alpha(\alpha+2)] + 2\omega h_{t+1}(\frac{3-\kappa}{\kappa})}{[h_{t+1}(1+\alpha+\beta) + \omega]^2} .$$

Thus, the kurtosis of the conditional distribution for the two-step estimator will be smaller than the one-step kurtosis whenever the numerator in this expression is negative. In particular, when $3 < \kappa < 2\beta\alpha^{-2}$ it is possible to have a decrease in $\kappa_t^{(2)}$ relative to κ. For $\kappa = 3$ and $\alpha > 0$ there is an unambig- uous increase in kurtosis. If α and β are both zero, so the process defined in (27) has constant variance, $\kappa^{(2)}$ becomes time independent,

$$\kappa^{(2)} = \kappa + \frac{3-\kappa}{2}$$

and the familar result from the central limit theorem that $\kappa^{(2)}$ is closer to three than κ applies. The time dependence of $\kappa_t^{(2)}$ also vanishes for the integrated GARCH(1,1) model. In that situation the two step conditional kurtosis becomes:

$$(32) \qquad \kappa^{(2)} = \kappa + \kappa[(3-\kappa)(\frac{2(1-\alpha)}{\kappa} - \alpha^2) + 2\alpha(\alpha+2)] .$$

General conditions for a decrease in kurtosis is therefore easy to find in this case. For example if $\alpha = .1$, $\kappa^{(2)} < \kappa$ provided

$3.940 < \kappa < 137.060$, and for $\alpha = .05$ the condition for a decrease in kurtosis is $3.365 < \kappa < 677.635$.

For the exchange rate data, the integrated GARCH(1,1) model given in (26) was reestimated allowing for conditionally standardized Student-t distributed errors treating $1/\nu$ as an additional parameter to be estimated. The log likelihood function for a sample of T observations becomes in this situation

$$(33) \qquad \text{LogL} = \sum_{t=1}^{T} \Big[\log(\Gamma(\tfrac{\nu+1}{2})) - \log(\Gamma(\tfrac{\nu}{2})) - \tfrac{1}{2}\log(\nu-2)h_t)$$

$$- \tfrac{1}{2}(\nu+1)\, \log(1 + \epsilon_t^2 h_t^{-1}(\nu-2)^{-1}\Big] , \qquad \nu > 2$$

where $\Gamma(\cdot)$ denotes the usual gamma function. It is possible to show, that for $\nu \to \infty$, or equivalently $1/\nu \to 0$, (33) converges to the log likelihood function with conditionally normal errors defined in (7). Maximization by the BHHH algorithm leads to the estimates:

$$\epsilon_{t+1}|\psi_t \sim t_{\nu,h_{t+1}} , \qquad 1/\nu = \underset{(.009)}{.096}$$

$$(34) \qquad h_{t+1} = \underset{(.013)}{.104}\epsilon_t^2 + (1-.104)h_t$$

$$\text{logL} = 1751.735 .$$

The likelihood ratio test statistic for conditional normality, i.e. $1/\nu = 0$, takes the value 13.019 which is highly significant at any reasonable level, see Bollerslev (1985c) where it is shown by means of Monte Carlo methods that the likelihood ratio test statistic is more concentrated towards the origin than a $\chi^2(1)$ distribution in this situation. Also, a conventional t-test on $1/\nu$ overwhelmingly rejects the assumption of conditional normality for this model. It follows from the formula $\kappa = 3(\nu-2)(\nu-4)^{-1}$ that the conditional kurtosis in (34) is estimated to 3.94. This compares quite well with a sample kurtosis of the standardized residuals equal to 3.74. The unrestricted GARCH(1,1) was also estimated with conditionally Student-t distributed errors with the finding that $1/\nu = .115(.016)$. The likelihood ratio test statistic for $1/\nu = 0$ equals here 10.183, which again is highly significant. The estimated conditional kurtosis for this model is 4.28 while the sample analog is 3.93.

To examine the effects of time aggregation the series was split into two subseries of 315 non-overlapping biweekly observations. We confirm the findings by Boothe and Glassman (1985) and others summarized therein that the unconditional kurtosis declines with time aggregation dropping from 4.47 for the whole sample to an average of 3.55 for each of the two subsamples. However, if the true model is indeed the integrated GARCH(1,1) model in (34), the population fourth moment in each case is infinite. The decline in the sample kurtosis might therefore merely be due to the decline in sample size.

The interesting question in this context is the effect of time aggregation on the conditional fourth moment. Recognizing that an integrated GARCH(1,1) weekly model does not aggregate to an integrated GARCH(1,1) biweekly model, we proceed to estimate

this biweekly model. The result for the first subsample is

$$h_{t+1}^{(2)} = \underset{(.023)}{.118}\varepsilon_t^2 + (1-.118)h_t^{(2)}, \qquad 1/\nu = \underset{(.002)}{.074} .$$

The estimated conditional kurtosis for this model is 3.63. This is in close accordance with the implied estimate from the weekly model (34). Substituting the parameter values from the estimated weekly model in (32) we obtain an estimate of $\kappa^{(2)} = 3.96$. Note, this estimate of $\kappa^{(2)}$ is essentially unchanged from the weekly estimate, $\kappa = 3.94$. The parameters are just at the point where the conditional kurtosis does not change with aggregation. The second biweekly sample has a larger estimated value of ν, and therefore a kurtosis closer to 3.

VIII. NON-LINEAR CONDITIONAL HETEROSKEDASTICITY

One explanation for the success of the conditional Student-t distribution in modelling the exchange rate series is that the observations are contaminated by some outliers or extreme values which make the conditional distribution look fat tailed. If indeed these extreme values are interpreted as outliers, then they may not be helpful in predicting future variances, and furthermore the estimates in the variance function may be unduely influenced by a few extreme observations. To examine this issue, a series of alternative specifications for the conditional variance were estimated. Here we shall only consider the two most successful of these alternative specifications. These are:

(35) $h_{t+1} = \omega + \beta h_t + \gamma_1 |\epsilon_t|^{\lambda_1}$

and

(36) $h_{t+1} = \omega + \beta h_t + \gamma_2(2F(\epsilon_t^2/\lambda_2) - 1)$

where F denotes the standard normal cumulative distribution function. If $\lambda_1 = 2$ in (35) the model is simply GARCH(1,1). In (36) if λ_2 is large relative to the standard deviation of ϵ_t, the model is essentially linear.

Models (35) and (36) were both estimated with the same maximum likelihood routine using the BHHH algorithm simply by substituting in the new expressions for the conditional variances. Assuming conditional normality the likelihood function increases from 1750.061 for the unrestricted GARCH(1,1) to 1752.571 and 1753.461 for models (35) and (36) respectively. The likelihood ratio test statistic for (35) versus the unrestricted GARCH(1,1) equals therefore 5.020 corresponding to the .975 fractile in a $\chi^2(1)$ distribution. The maximum likelihood estimate of λ_1 is 1.543(.130). The estimated value of λ_2 in (36) is $.765 \cdot 10^{-3}$ $(.253 \cdot 10^{-3})$. Since the unconditional residual variance equals $\sigma^2 = .268 \cdot 10^{-3}$, it follows that $(\hat{\lambda}_2)^{1/2} = 1.690 \cdot \sigma$, and there is an important attenuation of the largest residuals in (36). Also, (36) exhibits a slightly higher likelihood function than (35) and converged more easily.

The improvements for the non-linear models are somewhat greater when the conditional Student-t distribution is employed. In that case the likelihood function increases from 1756.497 for the unrestricted GARCH(1,1) to 1760.297 and 1761.730 for models (35) and (36) respectively. However, the non-linear models with conditionally t-distributed errors were much harder to iterate to convergence. Both the estimates of $\hat{\lambda}_1 = 1.428(.149)$ and $\hat{\lambda}_2 = .408 \cdot 10^{-3}(.148 \cdot 10^{-3})$ are lower when $1/\nu > 0$. It is interesting to note that the estimates of β are very similar across the six models, ranging from a high of .891(.013) for (36) with conditionally normal errors to a low of .852(.031) for (35) with conditionally t-distributed errors.

This new non-linear class of conditional variance models may provide important improvements for some time series, however in this case it is not clear that the improvements are worth the analytical complexity which arises from the non-linearities. An alternative approach to the non-linear models might be to use non-parametric techniques. For a recent discussion and some references on this subject see Engle, Granger, Rice and Weiss (1986).

IX. SOLVING THE ASSET PRICING EQUATION

Using the class of models described above it is now possible to obtain simple closed form solutions to the asset pricing equations given in (10), (15) and (16). In each case a solution is a stochastic process for the asset price which depends upon the process of the forcing variable. In particular, for the forward contract in (10) the process of the price should depend upon the process of the spot price, y_t,

(10) $p_t = r^{-s}[E_t(y_{t+s}) - \delta V_t(y_{t+s})]$.

There are several special cases to be considered in solving equation (10). To illustrate let y_t be a random walk with innovations which follow an integrated GARCH(1,1) process. In that situation $E_t(y_{t+s}) = y_t$ and

$$V_t(y_{t+s}) = E_t(\sum_{i=1}^{s} \epsilon^2_{t+i}) = E_t(\sum_{i=1}^{s} h_{t+i}) = sh_{t+1}$$

so that

(37) $p_t = r^{-s}[y_t - \delta sh_{t+1}]$.

For a future contract where no money changes hands until the terminal date at t+s, the alternative sure rate of return is equal to zero so that r=1. Thus, in that case (37) simplifies to:

(38) $p_t = y_t - \delta sh_{t+1}$.

Therefore, if $\delta \neq 0$ there will be a time varying risk premium in the future contract. For contracts far in the future new information will have a substantial effect on asset prices as it changes agent's perceptions of the variance of the final payoff as well as all the intermediate variances. The persistence of the variance gives time varying risk premia even for contracts many periods in the future and implies sizeable effects on asset prices. Note, because the innovation in y_t enters both linearly and as a square in (37) the future price series may exhibit some skewness.

Suppose now as an alternative case, that y_t is a simple random walk with serially independent innovations and variance σ^2. Then $V_t(y_{t+s}) = s\sigma^2$ and the solution to (10) proceeds in identical fashion, leading to the future contract pricing equation:

(39) $p_t = y_t - \delta s\sigma^2$.

This is essentially the solution achieved by Hellwig (1982), although in a different context. The variance of the spot price also enters the pricing equation in this situation, but does not give rise to a time varying risk premium. New information casts no light on future uncertainty.

Finally, consider the case where the spot price y_t is a random walk with innovations distributed as a stationary GARCH(1,1). Letting $\sigma^2 = \omega/(1-\alpha-\beta)$, it follows from (23) that the multi-step forecasts of the conditional variance of the innovations, h_{t+s}, can be written as

$$E_t(h_{t+s} - \sigma^2) = (\alpha+\beta)^{s-1}(h_{t+1} - \sigma^2)$$

and

$$V_t(y_{t+s}) = \sum_{i=1}^{s} \sigma^2 + E_t(h_{t+i} - \sigma^2)$$

$$= s\sigma^2 + [h_{t+1} - \sigma^2] \frac{1 - (\alpha+\beta)^s}{1 - \alpha - \beta} .$$

By substitution in (10) the solution for the future constract is:

$$P_t = y_t - \delta s\sigma^2 + \delta[h_{t+1} - \sigma^2] \frac{1-(\alpha+\beta)^s}{1-\alpha-\beta} . \quad (40)$$

When $\alpha+\beta < 1$, the current information continues to be an important part of the time varying risk premium even for large s, but the relative importance decreases with the length of the contract. This is in contrast to the solution for the integrated GARCH(1,1) model given in (38) where the importance of h_{t+1} increases with the length of the contract.

Of course, the set of solutions presented here is far from exhaustive, but merely serves to illustrate the general idea. Namely, that a given solution to the asset pricing equations (10), (15) or (16) depends in a crucial way on the distribution of the forcing variable, y_t, and in particular on the conditional variance of y_t.

X. CONCLUSIONS

The statistical properties as well as the empirical relevance of the ARCH model in modelling risk and uncertainty in economies is by now well established. In this paper, however, several new unanswered questions have been raised.

Both from an empirical and theoretical economic point of view do the integrated GARCH models constitute an interesting development, but the statistical properties for this new class of models are largely unknown. Some preleminary Monte Carlo evidence suggests that the knife edge properties when estimating and testing for a unit root in the mean might not be as severe in the inte-

grated variance models. So far this is a conjecture. Similarly, the empirical regularity and statistical properties of the more complicated non-linear conditonal heteroskedasticity models remain open questions. Another related issue involves the problem of testing for one sided alternatives when the null hypothesis is on the boundary of the admissable parameter space, as is typically the case when testing in GARCH models with non-negativity constraints. It would also be of interest to see how the aggregation formulas derived here hold up against a more detailed and broader empirical analysis. Finally, the validity of the asset pricing formulas presented throughout the paper, especially the pricing equations for futures contracts given in the last section, in explaining observed price behavior is another very interesting question. We leave the answer to all of these questions for future research.

ACKNOWLEDGEMENTS

We are deeply indebted to many for discussions helping to clarify key issues. Particular thanks go to Dick Baillie, Buz Brock, Roman Frydman, Clive Granger, Dale Poirier, Mike Rothschild, Larry Weiss and Mike Wickens. For remaining misunderstandings omissions and errors, we humbly apologize. The research was supported by NSF Grant No. SES 84-20680.

REFERENCES

Amsler, C., (1984a). Term Structure Variance Bounds and Time Varying Liquidity Premia. Economics Letters, 16, 137-144.

_______ (1984b). Including Time Varying Liquidity Premia in Term Structure Variance Bounds. Michigan State University: Econometrics Workshop Paper #8308.

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

Blanchard, O.J., & Watson, M.W., (1984). Are Business Cycles All Alike? NBER Working Paper No. 1392.

Bodie, Z., A. Kane, & McDonald, R., (1984). Why Haven't Nominal Rates Declined. Financial Analysts Journal, March-April 1984.

_______ (1984). Why are Real Interest Rates So High? NBER Working Paper No. 1141.

Bollerslev, T., (1985a). Generalized Autoregressive Conditional Heteroskedasticity. Forthcoming Journal of Econometrics. Also UCSD, Department of Economics Discussion Paper No. 85-16.

_______ (1985b). A Note on the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process. UCSD, Department of Economics, Discussion Paper No. 85-23.

_______ (1985c). A Conditionally Heteroskedastic Time Series Model for Security Prices and Rates of Return Data. UCSD, Department of Economics, Discussion Paper No. 85-32.

Bollerslev, T., Engle, R.F., & Wooldridge, J.M., (1985). A Capital Asset Pricing Model with Time Varying Covariances. UCSD, Department of Economics, Discussion Paper No. 85-28.

Boothe, P., & Glassman, P., (1985). The Statistical Distribution of Exchange Rates: Empirical Evidence and Economic Implications. University of Alberta, Department of Economics, Research Paper No. 85-22.

Coulson, N.E., & Robins, R.P., (1985). Aggregate Economic Activity and the Variance of Inflation: Another Look. Economics Letters, 17, 71-75.

Dickey, D.A., & Fuller, W.A., (1979). Distribution of Estimators for Autoregressive Time Series with a Unit Roof. Journal of the American Statistical Association, 77, 427-431.

Diebold, F.S., & Nerlove, N., (1985). ARCH Models of Exchange Rate Fluctuations. Unpublished manuscript, University of Pennsylvania, Department of Economics.

Diebold, F.S., & Pauly, P., (1985). Endogenous Risk in a Portfolio Balance Rational Expectations Model of the Deutschemark-Dollar Rate. University of Pennsylvania, Department of Economics, Econometric Discussion Paper # 85-17.

Domowitz, I., & Hakkio, C.S., (1985). Conditional Variance and the Risk Premium in the Foreign Exchange Market. Journal of International Economics, 19, 47-66.

Engle, R.F., (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation. Econometrica, 50, 987-1008.

_______ (1983). Estimates of the Variance of U.S. Inflation Based on the ARCH Model. Journal of Money Credit and Banking, 15, 286-301.

Engle, R.F., Brown, S., & Stern, G., (1984). Short Run Forecasts of Electricity Sales: A Comparison of New Methodologies. Unpublished manuscript UCSD, Department of Economics.

Engle R.F., & Granger, C.W.J., (1985). Cointegration and Error Correction: Representation, Estimation and Testing. Forthcoming Econometrica. Also manuscript UCSD, Department of Economics.

Engle, R.F., Granger, C.W.J, & Kraft, D., (1984). Combining Competing Forecasts of Inflation Using a Bivariate ARCH Model. Journal of Economic Dynamics and Control, 8, 151-165.

Engle, R.F., Granger, C.W.J., & Rice, J., Weiss, A., (1986). Non-Parametric Estimates of the Relation Between Weather and Electricity Demand. Forthcoming Journal of the American Statistical Association, June 1986.

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

Engle, R.F., Hendry, D., & Richard, R.F., (1983). Exogeneity. Econometrica, 51, 277-304.

Engle, R.F., & Kraft, D., (1983). Multiperiod Forecast Error Variances of Inflation Estimated from ARCH Models. Applied Time Series Analysis of Economic Data, (A. Zellner, Ed.), Bureau of the Census, 293-302.

Engle, R.F., Lilien, D., & Robins, R., (1985). Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. Forthcoming Econometrica. Also UCSD, Department of Economics, Discussion Paper No. 85-17.

Engle, R.F., & Watson, M., (1985). Applications of Kalman Filtering in Econometrics. Invited Paper to the Fifth World Congress of the Econometric Society, Cambridge, Mass. Forthcoming in Volume of Proceedings. Also UCSD, Department of Economics, Discussion Paper No. 85-31.

French, K.R., Schwert, G.W., & Stambaugh, R.F., (1985). Expected Stock Returns and Volatility. University of Rochester, Graduate School of Management, Working Paper Series No. MERC 85-10.

Granger, C.W.J., Robins, R.P., & Engle, R.F., (1984). Wholesale and Retail Prices: Bivariate Time Series Modelling with Forecastable Error Variances. In Model Reliability, (Belesley and Kuh, Eds.), MIT Press.

Hellwig, M.F., (1982). Rational Expectations Equilibrium With a Conditioning on Past Prices: A Mean Variance Example. Journal of Economic Theory, 26, 279-312.

Hsieh, D.A., (1985). The Statisitcal Properties of Daily Foreign Exchange Rates: 1974-1983. Mimeo, University of Chicago, Graduate School of Business.

Kraft, D., & Engle, R.F., (1982). Autoregressive Conditional Heteroskedasticity in Multiple Time Series Models. UCSD, Department of Economics, Discussion Paper No. 82-23.

McCurdy, T.H., & Morgan, I.G., (1985). Testing the Martingale Hypothesis in the Deutschmark/U.S. Dollar Futures and Spot Markets. Queens University, Department of Economics, Discussion Paper No. 639.

McLeod, A.J., & Li, W.K., (1983). Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. Journal of Time Series Analysis, 4, 269-273.

Milhøj, A., (1985a). The Moment Structure of ARCH Processes. Scandinavian Journal of Statistics, 12, 281-292.

________ (1985b). A Conditional Variance Model for Daily Deviations of an Exchange Rate. Unpublished manuscript, University of Copenhagen, Institute of Statistics.

Nemec, A.F., Linnell, (1984a). Conditionally Heteroscedastic Autoregressions. University of Washington, Department of Statistics, Technical Report #43.

__________ (1984b). Least Squares Estimation of Conditionally Heteroscedastic Autoregressions. University of Washington, Department of Statistics, Technical Report #48.

Pagan, A.R., Hall, A.D., & Trivedi, P.K., (1983). Assessing the Variability of Inflation. Review of Economic Studies, 50, 585-596.

Pantula, S.G., (1984). Autoregressive Conditionally Heteroscedastic Models. Unpublished manuscript, North Carolina State University, Department of Statistics.

Poterba, M.P., & Summers, L.H., (1984). The Persistance of Volatility and Stock Market Fluctuations. Harvard University, Institute of Economic Research, Discussion Paper No. 1092.

Wallis, K.F., (1980). Econometric Implications of the Rational Expectations Hypothesis. Econometrica, 48, 49-73.

Weiss, A.A., (1982). Asymptotic Theory for ARCH Models: Stability, Estimation and Testing. Forthcoming Econometric Theory. Also UCSD, Department of Economics, Discussion Paper No. 82-36.

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

DATA APPENDIX

Weekly Swiss Franc/Dollar Exchange Rate

(data to be read across rows)

2.7470	2.8690	2.7820	2.7850
2.8530	2.8640	3.0390	3.0410
3.0210	3.0180	3.0190	3.0100
3.0170	3.0210	2.9890	3.0000
3.0190	3.1000	3.1320	3.1710
3.1780	3.1970	3.1960	3.1900
3.1990	3.1910	3.3520	3.3640
3.3740	3.3910	3.2760	3.2540
3.1960	3.1260	3.1220	3.1340
3.0950	3.0490	3.0260	3.0650
3.0730	3.0390	2.9670	2.9380
2.8990	2.9150	2.8940	2.9940
2.9520	3.0090	3.0020	3.0100
2.9730	2.9970	2.9330	2.8930
2.9630	2.9850	2.9760	3.0140
2.9990	3.0160	3.0150	2.9850
2.9850	2.9470	2.9050	2.9020
2.8470	2.8740	2.8140	2.8070
2.7700	2.6850	2.6860	2.6250

2.5470	2.5250	2.5610	2.5790
2.5700	2.5440	2.4910	2.5370
2.5010	2.4780	2.3980	2.4360
2.4840	2.4670	2.5280	2.5350
2.5610	2.5700	2.5450	2.5580
2.5200	2.4770	2.4910	2.4780
2.4930	2.4910	2.4940	2.4950
2.5380	2.5620	2.5950	2.6770
2.7060	2.6780	2.6800	2.6750
2.6790	2.6760	2.6870	2.7230
2.7190	2.7080	2.6980	2.6540
2.6260	2.6260	2.6440	2.6390
2.6550	2.6690	2.6500	2.6340
2.6280	2.6270	2.6210	2.5990
2.6000	2.6110	2.6020	2.5930
2.5430	2.5640	2.5610	2.5810
2.5790	2.5300	2.5580	2.5390
2.5430	2.5250	2.5300	2.5190
2.4890	2.5000	2.4920	2.4650
2.4050	2.4990	2.4780	2.4830
2.4710	2.4790	2.4780	2.4920
2.4960	2.4750	2.4890	2.4720
2.4770	2.4720	2.4810	2.4650
2.4690	2.4550	2.4480	2.4460
2.4480	2.4350	2.4390	2.4410
2.4470	2.4370	2.4540	2.4520
2.4520	2.4480	2.4450	2.4550
2.4780	2.5060	2.5250	2.5160
2.5150	2.5140	2.5350	2.5530
2.5650	2.5570	2.5430	2.5440
2.5370	2.5180	2.5240	2.5180
2.5200	2.5280	2.5190	2.5140
2.1020	2.1220	2.1720	2.1650
2.1540	2.1460	2.1390	2.0900
1.9900	1.9620	1.9750	1.8810
1.8530	1.8990	1.8860	1.7940
1.7610	1.7770	1.7810	1.7810
1.8380	1.8200	1.8160	1.7980
1.8140	1.8700	1.8560	1.8550
1.8930	1.8960	1.9130	1.8720
1.8790	1.8540	1.8820	1.9000
1.9340	1.9650	1.9650	1.9580
1.9510	1.9340	1.9320	1.9820
1.9750	2.0140	2.0520	2.1000
2.0830	2.1080	2.1400	2.1280
2.0730	2.0930	2.1030	2.1560
2.1180	2.0580	2.1160	2.1220
2.1510	2.1410	2.1730	2.2020
2.1340	2.1750	2.2040	2.2180
2.2250	2.2020	2.1610	2.1150
2.0850	2.0850	2.0230	1.9980
1.9520	1.9310	1.9910	2.0000

2.0280	2.0130	1.9830	2.0360
2.0490	2.0580	2.0460	2.0760
2.0860	2.0490	2.0540	2.0650
2.0620	2.0580	2.0250	2.0610
2.0760	2.1160	2.1330	2.1410
2.0880	2.1090	2.1230	2.1130
2.0900	2.1180	2.1540	2.1910
2.1390	2.1530	2.1790	2.1750
2.1790	2.1620	2.1320	2.0940
2.1250	2.1140	2.1230	2.1550
2.1620	2.1780	2.1780	2.1570
2.1850	2.2120	2.2140	2.1890
2.2330	2.2530	2.2320	2.2440
2.2350	2.2290	2.1970	2.1940
2.1680	2.1160	2.1280	2.1770
2.1580	2.1710	2.1650	2.1900
2.2160	2.2480	2.2800	2.2680
2.2770	2.2650	2.2330	2.2710
2.3060	2.3320	2.3740	2.4030
2.4210	2.4340	2.4650	2.4450
2.4110	2.4000	2.3910	2.4640
2.4900	2.5340	2.5080	2.4950
2.5370	2.5580	2.4850	2.4940
2.4170	2.4500	2.5020	2.5210
2.5450	2.5550	2.5580	2.5760
2.6170	2.6520	2.6760	2.6830
2.6760	2.7370	2.8120	2.8230
2.8400	2.8980	2.8500	2.7550
2.6500	2.6420	2.6530	2.5310
2.5950	2.6400	2.6860	2.5810
2.5890	2.5900	2.5840	2.6010
2.5300	2.5350	2.5420	2.4510
2.3350	2.3690	2.2440	2.3620
2.5000	2.4930	2.4890	2.4930
2.4580	2.4370	2.4180	2.4020
2.3880	2.4030	2.4180	2.4360
2.3780	2.3980	2.3830	2.3890
2.3610	2.3530	2.3350	2.2970
2.2820	2.2370	2.2190	2.2270
2.2080	2.1900	2.1600	2.1190
2.0760	2.0260	2.0180	2.0510
1.9600	2.0140	1.9610	1.9860
1.9650	1.9120	1.8270	1.8350
1.8930	1.9390	1.9160	1.8750
1.8620	1.8670	1.9300	1.9570
1.9530	1.9720	1.9870	1.9710
1.8800	1.9130	1.8930	1.8650
1.8590	1.8180	1.8070	1.8130
1.7840	1.7130	1.6890	1.6130
1.6690	1.6530	1.6140	1.5910
1.5250	1.4920	1.5810	1.5420
1.5130	1.5060	1.6090	1.6190

1.6490	1.7210	1.7230	1.7090
1.7020	1.6410	1.6200	1.6440
1.6640	1.6670	1.6740	1.7030
1.6300	1.6720	1.6740	1.6710
1.6670	1.6770	1.6870	1.6890
1.7040	1.7150	1.7100	1.7110
1.7230	1.7130	1.7270	1.7320
1.7210	1.7340	1.7310	1.6750
1.6520	1.6510	1.6540	1.6260
1.6380	1.6570	1.6540	1.6560
1.6550	1.6640	1.6480	1.6280
1.6150	1.5750	1.5790	1.6030
1.6470	1.6570	1.6520	1.6380
1.6420	1.6440	1.6340	1.6030
1.6010	1.6030	1.5840	1.5710
1.5750	1.5830	1.6120	1.6250
1.6030	1.6170	1.6410	1.6650
1.7140	1.7250	1.7680	1.7960
1.8600	1.7880	1.7200	1.7000
1.6670	1.6500	1.6640	1.6650
1.6510	1.6530	1.6300	1.6240
1.6400	1.6190	1.5920	1.5980
1.5990	1.6500	1.6320	1.6420
1.6590	1.6570	1.6300	1.6300
1.6310	1.6500	1.6400	1.6380
1.6440	1.6620	1.6930	1.7390
1.7070	1.7070	1.7330	1.7520
1.8200	1.8010	1.7700	1.7870
1.7710	1.8060	1.8260	1.8810
1.9210	1.9620	1.9740	1.9190
1.9430	1.9380	1.8570	1.9050
1.9130	1.9480	2.0010	1.9810
2.0160	2.0630	2.0820	2.0420
2.0790	2.1070	2.0970	2.0520
2.0270	2.0720	2.1120	2.0670

ECONOMETRIC REVIEWS, 5(1), 51-56 (1986)

MODELING THE PERSISTENCE OF CONDITIONAL VARIANCES: A COMMENT

Professors Engle and Bollerslev have delivered an excellent blend of "forest" and "trees"; their important extensions of the basic ARCH(q) model nicely complement the comprehensive survey. In these comments I will focus on some general issues which perhaps received insufficient attention, as well as on specific issues related to the authors' new results. In the former category are the statistical versus economic motivations for GARCH, the effects of GARCH on other standard diagnostics (in particular, tests for linear and nonlinear serial correlation), and multivariate GARCH modeling. In the latter category are specification caveats regarding integrated variance models, and the effects of temporal aggregation on the unconditional distribution of GARCH processes.

The statistical motivation for the ARCH(q) model, as well as all of the extensions introduced by Engle and Bollerslev, is best seen by recalling the fundamental Wold decomposition: every covariance stationary stochastic process may be written as the sum of 1) a linearly deterministic component, and 2) a linearly indeterministic component with one-sided MA representation:

(1) $$\sum_{j=0}^{\infty} b_j \varepsilon_{t-j} \text{ , where}$$

(2) $$\sum_{j=0}^{\infty} b_j^2 < \infty, \text{ and}$$

$$E(\varepsilon_t \varepsilon_\tau) = \sigma^2 < \infty \text{ if } t = \tau, \text{ 0 otherwise .}$$

 0747-4938/86/0501-0051$3.50/0

Notice that, while the innovations are uncorrelated, they need not be independent since neither their conditional nor unconditional density need be Gaussian. Furthermore, it is only the unconditional innovation variance $E(\varepsilon_t^2)$ which must be constant and finite; this is a restriction of constancy and finiteness on the expectational integral of the conditional innovation variance (i.e., on $E_{\Omega}(E_{\varepsilon/\Omega}\ \varepsilon_t^2)$), but it in no way requires constancy of the conditional variance ($E_{\varepsilon/\Omega}(\varepsilon_t^2)$) itself. In summary, then, (1) allows for a time-varying conditional mean while (2) allows for a time-varying conditional variance. Although all stationary ARMA models satisfy (1) and (2), the converse is not true. Similarly, while all GARCH models satisfy (1) and (2), there are many conditional variance structures not captured by the models considered by Engle and Bollerslev. The search for a parsimonious and descriptively accurate subclass of conditional-mean models has largely ended; we now routinely consider only models with rational spectral densities (e.g. ARMA models). The Engle-Bollerslev paper makes great progress in the search for a similar "consensus" conditional variance specification. I believe that the GARCH(p,q) model, with Student's-t conditional density, is the major contribution of the paper due to the improvements in descriptive accuracy and parsimony which it facilitates relative to the conditionally normal ARCH(q) model. The nonlinear conditional variance models will probably prove less useful, unless as-yet-nonexistent evidence is uncovered indicating that the linear specification is an inadequate approximation for economic phenomena.

The statistical considerations sketched above indicate that the presence of GARCH is possible, but is it probable in economic terms? My guess is yes. First, the results of Clark (1973), as extended by Stock's (1984) important work on time deformation, indicate that inappropriate use of a "calendar" time scale may lead to GARCH-type volatility clustering. Consider, for example, an economic variable evolving at the rate of one step per unit of some non-calendar time (e.g., "business-cycle time" or "information-arrival time). Then, relative to calendar time, the process actually evolves more quickly in some periods than in others, with associated movements in prediction error variances. Stock's work also provides a good example of the association of continuous-time phenomena and GARCH effects, as stressed in Professor Sims' comments. The linkage between time deformation and GARCH needs further study, presumably leading to testable restrictions between the two.

Assuming that there exist economic time series not subject to time deformation but nevertheless displaying GARCH, we need a truly economic theory leading to GARCH in an equilibrium model of optimizing agents. Because economic theory usually does not generate testable implications for higher-ordered moments of economic variables, this appears an arduous task, but some progress may not be too far off.

To the extent that GARCH is present in observed economic time series or in model disturbances, it invalidates other important diagnostics, such as tests for serial correlation. Domowitz and Hakkio (1984) develop an LM test for serial correlation which is robust to heteroskedasticity of unknown form, while Diebold (1986a) develops serial correlation tests specifically robust to ARCH. The relative power of the two approaches depends on the accuracy of the ARCH approximation; further work on "robustifying" other model diagnostics is needed as well.

While all covariance-stationary stochastic processes have a Wold representation (1)-(2), it is not necessarily the most efficient form for prediction; a well-known example is the class of bilinear models. GARCH effects in the residuals of linear models may indicate the superiority of a nonlinear conditional mean specification, leading to improved point prediction relative to a linear model with GARCH disturbances. Some initial progress in distinguishing GARCH from bilinearity (and in combining the two models) has been made by Weiss (1986), but further research is needed.

Multivariate GARCH, which is only briefly discussed in the paper, may prove particularly useful in the modeling of time-varing risk premia and general disturbance-related economic phenomena; the main problem so far has been the huge number of parameters which must be estimated. Even the "diagonal" model in which conditional variances depend only on own lags and squared innovation lags, and conditional covariances depend only on own lags and innovation cross products, is tractible only in low-dimensional cases. Furthermore, the diagonality restriction sometimes appears invalid, as in Engle, Granger and Kraft (1984). A factor-analytic model with GARCH effects may prove to be a particularly attractive alternative. Consider the N-dimensional time series $\{y_t\}$, given by:

(3) $$y_{it} = \lambda_i F_t + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where:

$$E F_t = E \varepsilon_{it} = 0, \text{ for all } i \text{ and } t$$

$$E F_t F_\tau = 0, \text{ all } t \neq \tau, \text{ and } E F_t \varepsilon_{i\tau} = 0, \text{ all } i, t, \tau$$

$$F_t / \Omega_{t-1} \sim (0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i F_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

$$E \varepsilon_{it} \varepsilon_{j\tau} = \theta_i \text{ if } i = j \text{ and } t = \tau, \text{ 0 otherwise.}$$

In obvious matrix notation, we write the model as:

(4) $$y_t = \lambda F_t + \varepsilon_t.$$

Immediately, then,

(5) $$\Sigma_t \equiv \text{cov}_{t-1}(y_t) = \sigma_t^2 \lambda \lambda' + \Theta,$$

where $\Theta = \text{cov}(\varepsilon_t)$.

The substantive motivation of such an approach is quite strong. The "common factor," F, represents general influences which tend to affect all variables, albeit with different strengths captured by the λ_i's. The "unique factors," represented by

the ε_i's, are uncorrelated variable-specific shocks. The rich (and testable) conditional variance-covariance structure of the observed variables arises from their joint dependence on the common factor F; this leads to *commonality* in temporal volatility movements across economic variables, which is frequently observed. For example, Diebold and Nerlove (1986) find strong evidence of one common GARCH-factor in a seven-variable system of dollar exchange rates and provide estimates of the corresponding multivariate model.

In closing, I have a few remarks concerning integrated variance models and temporal aggregation. While the integrated variance model may prove very useful, it should be considered "dangerous" at this preliminary stage. First, the authors correctly note that all of the problems which plague conditional mean unit root tests may carry over to the conditional variance. In particular, it may prove very difficult in specific cases to determine whether a trend or a unit root (or both) is operative, in spite of the fact that the models have *very* different properties and implications. Second, the problem of initial conditions becomes very important for the estimation of integrated variance models, which are nonstationary. Third, while one motivation for the use of the GARCH approximation is that appropriate variance "forcing" variables are rarely known in the time-series context, this arguement is less convincing when the variance appears integrated. Integration implies *persistent* movements in variance, in which case we should search harder for *some* economic explanation of the movements. For example, while interest-rate equations may appear to have integrated-variance disturbances, it may be due to a failure to include monetary-regime dummies for the conditional variance intercept, ω. This would correspond to stationary GARCH movements within regimes, with an unconditional "jump" occuring between regimes.

Finally, in work complementary to the authors' results on the effects of temporal aggregation on the *conditional* density, Diebold

(1986b) has shown that temporal aggregation of GARCH processes leads to unconditional normality, in spite of the fact that successive observations are not independent. Thus, as a distributional model for asset returns, GARCH leads to high frequency unconditional leptokurtosis, volatility clustering, and convergence to normality under temporal aggregation, all of which are observed in the data.

Francis X. Diebold
Federal Reserve Board

ADDITIONAL REFERENCES

Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. Econometrica, 41, 135-156.

Diebold, F.X. (1986a). Testing for serial correlation in the presence of ARCH. Forthcoming, Proceedings of the American Statistical Assoc., Business and Economic Statistics Section.

Diebold, F.X. (1986b). Temporal aggregation of ARCH processes and asset return distributions. Univeriesty of Pennsylvania econometrics discussion paper.

Diebold, F.X. and M. Nerlove (1986). Multivariate ARCH models of exchange rate fluctuations. University of Pennsylvania econometrics discussion paper.

Domowitz, I. and C.S. Hakkio (1984). Testing for serial correlation and common factor dynamics in the presence of heteroskedasticity. Northwestern University discussion paper.

Stock, J.H. (1984). Testing for time deformation. Harvard University discussion paper.

Weiss, A.A. (1986). ARCH and bilinear time-series models: comparison and combination. Journal of Business and Economic Statistics, 4, 59-70.

ECONOMETRIC REVIEWS, 5(1), 57-61 (1986)

COMMENT

As economic theory has turned toward behavior under uncertainty econometricians have increasingly worked with models that explicitly involve the dispersion of random variables. Robert Engle's work with conditional heteroskedasticity has been notable for its methodological innovations, the variants it has inspired other econometricians to produce, and the rapidity with which it has become a standard part of the methodological repertoire of applied econometricians. In this survey Engle and Bollerslev have described several of the variants of conditional heteroskedasticity models, and have introduced a new class of models, integrated in variance.

The models considered in this paper allow conditional heteroskedasticity to be persistent. The structure of this persistence is similar to that for conditional means in more traditional linear models. Section II of the paper provides a simple but important example of the way linear models have been adapted to this purpose. Taking the paper's equation (4) with $\phi = 0$,

$$\varepsilon_t | \Psi_{t-1} \sim N(0, h_t); \quad h_t = \omega + \alpha\varepsilon_{t-1}^2 \qquad (1)$$

In this model it is necessary to have $\omega > 0$ and $\alpha \geq 0$ so that conditional variances will be nonnegative; a finite unconditional variance further demands $\alpha < 1$. These conditions are usually met in empirical applications of two-parameter models of conditional heteroskedasticity. As such models begin to involve more parameters, e.g.,

 0747-4938/86/0501-0057$3.50/0

$$\varepsilon_t | \Psi_{t-1} \sim N(0,h_t); \quad h_t = \omega + \Sigma_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2, \tag{2}$$

the conditions for nonnegative conditional variance are typically violated by maximum likelihood estimates; e.g., Diebold and Nerlove (1985). One can enforce these restrictions by Bayesian methods, as I have described elsewhere (Geweke 1986a, 1986b). However, if one were to posit not (1) but

$$\varepsilon_t | \Psi_{t-1} \sim N(0,h_t); \quad \log(h_t) = \omega + \alpha \log(\varepsilon_{t-1}^2) \tag{3}$$

variances would be positive by construction. Since the gradients are simple,

$$\partial \log L/\partial\omega = (1/2)\Sigma_{t=2}^{n} \left[\varepsilon_t^2(\varepsilon_{t-1}^2)^{-\alpha} e^{-\omega} - 1\right],$$

$$\partial \log L/\partial\alpha = (1/2)\Sigma_{t=2}^{n} \log(\varepsilon_{t-1}^2)\left[\varepsilon_t^2(\varepsilon_{t-1}^2)^{-\alpha} e^{-\omega} - 1\right],$$

and the log-likelihood function is globally concave with Hessian

$$-(1/2)e^{-\omega}\Sigma_{t=2}^{n} \varepsilon_t^2(\varepsilon_{t-1}^2)^{-\alpha} \begin{bmatrix} 1 & \log(\varepsilon_{t-1}^2) \\ \log(\varepsilon_{t-1}^2) & [\log(\varepsilon_{t-1}^2)]^2 \end{bmatrix}$$

standard classical methods for inference should be straightforward. It would be interesting to see how this model compares with (1), and its obvious extension with (2), in some of the empirical applications taken to date.

The model (3) is motivated by thinking about the nonnegativity of conditional variances. In more elaborate models of conditional variance these considerations are even more important. An example is provided by the integrated GARCH models introduced in this paper. The leading, simple case is

$$\varepsilon_t | \Psi_{t-1} \sim N(0,h_t); \quad h_t = \alpha\varepsilon_{t-1}^2 + (1-\alpha)h_{t-1}$$

This model is motivated by the traditional random walk with a unit root in the conditional mean, and it shares with the traditional model the property $E_t(h_{t+s}) = h_{t+1}$. However, h_t is positive by construction, and in this regard it is quite different from a random walk. Just how different is suggested by the higher moments of the k-step-ahead conditional distribution of $\{\varepsilon_t\}$. Let $\varepsilon_t = h_t^{1/2} z_t$, $z_t \sim \text{IIDN}(0,1)$. We have

$$h_{t+s} = [\alpha z_{t+s-1}^2 + (1-\alpha)]h_{t+s-1},$$

$$h_{t+s}^2 = [\alpha^2 z_{t+s-1}^4 + (1-\alpha)^2 + 2\alpha(1-\alpha)z_{t+s-1}^2]h_{t+s-1}^2,$$

$$E(h_{t+s}^2 | \Psi_t) = (2\alpha^2 + 1)E(h_{t+s-1}^2 | \Psi_t) = (2\alpha^2 + 1)^{s-1} h_{t+1}^2$$

Since $E(\varepsilon_{t+s}^4 | \Psi_t) = E(h_{t+s}^2 | \Psi_t)E(z_{t+s}^4) = 3(2\alpha^2 + 1)^{s-1} h_{t+1}^2$,

$$\text{var}(\varepsilon_{t+s}^2 | \Psi_t) = [3(2\alpha^2 + 1)^{s-1} - 1]h_{t+1}^2$$

As s increases, the expectation of ε_{t+s}^2 remains the same but its variance increases exponentially. This means that for large s $P[\varepsilon_{t+s}^2 < \varepsilon_t^2] \doteq 1$, but there is a small probability that $\varepsilon_{t+s}^2 >> \varepsilon_t^2$. Just how fast this happens is shown by the examples in Table 1, which provide numerical approximations to the true distribution of ε_{t+s} based on 10,000 Monte Carlo replications for some values of α and s. Clearly, the integrated GARCH process is not typical of anything we see in economic time series. Similar problems arise if one takes

$$\log(h_t) = \alpha\log(\varepsilon_{t-1}^2) + (1-\alpha)\log(h_{t-1});$$

TABLE I[a]

Fractile points for c.d.f. of $\varepsilon_{t+s}|\psi_t$,

$\varepsilon_t|\psi_{t-1} \sim N(0,\ h_t),\quad h_t = \alpha\varepsilon_{t-1}^2 + (1-\alpha)h_{t-1}$

Fractile	s=2	s=4	s=6	s=6	s=12	s=18
	α = 1.0			α = 0.5		
.005	-3.58	-4.23	-3.96	-3.67	-3.75	-3.34
.050	-1.63	-1.06	-.54	-1.43	-1.14	-.82
.250	-.37	-.10	-.03	-.42	-.27	-.14
.500	.00	.00	.00	.00	.00	.00
.750	.36	.11	.03	.41	.25	.15
.950	1.60	1.03	.56	1.43	1.15	.84
.995	3.75	4.59	4.09	3.42	4.10	3.37

[a]Fractile points are based on 10,000 Monte Carlo replications. An indication of the error due to numerical approximation is the asymmetry of the distribution, since the true distribution is symmetric about zero.

TABLE II[a]

Fractile points for c.d.f. of $\varepsilon_{t+s}|\psi_t$,

$\varepsilon_t|\psi_{t-1} \sim N(0,\ h_t),\ \log(h_t) = \alpha\log(\varepsilon_{t-1}^2) + (1-\alpha)\log(h_{t-1})$

Fractile	s=2	s=4	s=6	s=6	s=12	s=18
	α = 1.0			α = 0.5		
.005	-3.51	-4.37	-4.45	-2.35	-1.15	-.38
.050	-1.59	-1.04	-.56	-.76	-.23	-.05
.250	-.38	-.10	-.03	-.127	-.019	-.003
.500	.00	.00	.00	.000	.000	.000
.750	.35	.11	.03	.124	.021	.003
.950	1.60	1.08	.54	.76	.22	.05
.995	3.62	4.58	3.28	2.15	1.03	.41

[a]Fractile points are based on 10,000 Monte Carlo replications. An indication of the error due to numerical approximation is the asymmetry of the distribution, since the true distribution is symmetric about zero.

see the numerical approximations to the true distributions in Table 2. The idea of drift in variance is intuitively appealing, but is difficult to formalize by analogy with models of drift in mean.

In most of the models of the persistence of conditional variance that have been considered to date the degree of persistence and the form of the unconditional distribution are very tightly linked. In (1) or (3), for example, the autocorrelation function of the conditional variances and the moments of the standardized conditional distribution are both functions of α; more autocorrelation implies greater kurtosis in the unconditional distribution (up to the point where the fourth moment fails to exist). It would be desirable to treat persistence and the form of the unconditional distribution as distinct properties within the same model, and it is encouraging to see the authors' use of the student-t with unknown degrees of freedom and the report of the analytical tractability of this model. This is a worthwhile generalization which could be pursued in models of the persistence of conditional means as well.

John Geweke
Duke University

ADDITIONAL REFERENCES

Geweke, J., (1986a). Exact inference in the inequality constrained normal linear regression model. Journal of Applied Econometrics, forthcoming.

Geweke, J., (1986b). Exact inference in models with autoregressive conditional heteroskedasticity. Dynamic Econometric Modeling (E. Berndt, H. White, & W. Barnett, Eds.) Cambridge: Cambridge University Press. Forthcoming.

ECONOMETRIC REVIEWS, 5(1), 63-69 (1986)

AN EXCURSION INTO CONDITIONAL VARIANCELAND*

I. Introduction

It was a pleasure to read such a clearly written and helpful paper on a rapidly developing topic, and Rob Engle and Tim Bollerslev are to be congratulated on making their ideas so accessible. The concept of conditional variances bids fair to be almost as important in practice as that of conditional means in some areas of economics, including efficient markets of the type studied in the paper, and the accurate evaluation of forecast variances. Moreover, ARCH processes have a wonderful role to play in teaching econometrics, providing classical illustrations of many points of theory which hitherto have been left at the abstract level. Consider the simplest data generating process with ARCH characteristics (see Engle, Hendry and Trumble, 1985):

$$(1)\quad \varepsilon_t = \eta_t\sqrt{\omega + \alpha\, \varepsilon^2_{t-1}} \quad \omega,\ \alpha > 0 \text{ with } \eta_t \sim IN(0,\ 1)$$

Clearly:

$$(2)\quad E(\varepsilon^2_t|\varepsilon_{t-1}) = \omega + \alpha\, \varepsilon^2_{t-1}$$

whereas if $|\alpha| < 1$:

---

* The title is a quote from correspondence with Rob Engle. Financial Support from the Economic and Social Research Council Grant B00220012 is gratefully acknowledged.

0747-4938/86/0501-0063$3.50/0

(3) $E(\varepsilon_t^2) = \omega/(1-\alpha) = \sigma^2$ and $E(\varepsilon_t\varepsilon_{t-1}) = 0$

Hence $\{\varepsilon_t\}$ is:

(a) *unconditionally* homoscedastic white-noise and

(b) *conditionally* normally distributed although it is ARCH; but

(c) successive ε_t are *not* independent, and

(d) $(\varepsilon_1 \ldots \varepsilon_T)$ is *not* (jointly) distributed as a multivariate normal (it has fatter tails).

Moreover, in the linear regression model:

(4) $y_t = x_t'\beta + \varepsilon_t$ where x_t is strongly exogenous for β

(see Engle *et al.* (1983)); then:

(e) the Ordinary Least Squares estimator $\hat{\beta}$ of β is *B.L.U.E.* and the estimated variance of $\hat{\beta}$ is an unbiased estimator of the actual sampling variance; but

(f) OLS can be extremely *inefficient* asymptotically relative to the Maximum Likelihood Estimator if α is close to unity; and

(g) the MLE of β is also unbiased (but non-linear).

Further, Engle *et al.* (1985) show that:

(h) a *pretest* estimator first testing $Ho : \alpha = 0$ and using OLS or MLE (according to the not reject/reject outcome) is *unbiased* in this context. Finally:

(i) the Durbin-Watson statistic is an *inconsistent* test of Ho as is easily proved. Thus, this simple ARCH process neatly illustrates a wide range of important theoretical econometrics ideas and distinctions. Three issues arising in this pedagogical role merit further analysis: (1) the persistence of shocks, (2) the fatness of tails to distributions and (3) how the presence of ARCH in data gets explained in a model.

II. Persistence of Variances

If you do live simulations in a classroom using (1) to illustrate ARCH behaviour in error terms, be prepared for some embarrassments: frequently no ARCH effects appear! This is

because (1) actually has remarkably little persistence in the conditional variance despite equation (5) in the paper. Why? As noted above, the unconditional expectation of $\omega + \alpha\,\varepsilon_{t-1}^2$ is σ^2 and $\Pr(|n_t| < 1) = 0.68$. So in (1) above, two-thirds of the time, the next shock ε_t is smaller than ε_{t-1} in absolute value (and 40% of the time it is halved or more). Equation (5) in Engle and Bollerslev arises because in variances, big shocks get large weights and so a large variance is required to characterise the whole distribution. However, the probability of a sequence of "big shocks" is small: e.g. $\Pr(|n_t| > 2) < .05$ and this would generate an unusually large ε_t^2 only if it happened to coincide with a value of ε_{t-1}^2 already above average. This negative outcome happens even if $\alpha = 1$, which only serves to retain *all* (rather than *part*) of the previous shock (but again merely for one period). This description of what appears to happen numerically when generating data by (1) is consistent with the findings of Engle *et al.* (1985) and accounts for their results to terms of many of the samples simply not "revealing" the underlying ARCH effect.

The GARCH process generates manifestly more persistence as of course does ARCH with a lag of longer than one period. The extreme is the new class of IGARCH processes, which the authors admit were formulated by analogy with unit roots in linear autoregressions. The latter certainly seem set for a bright future (see Engle and Granger, 1986 and Hendry, 1986): are the former? Here the need to restrict ω to zero to avoid a trend in variance is worrying and I would conjecture that current information gradually fades in relevance as horizons increase (as in their equation (6)), rather than persisting eternally. Certainly, on weekly data this might be hard to detect as different from a unit root, which therefore would be a good first approximation, although "policy implications" (e.g. asset purchase decisions) could differ radically between roots on or within the unit circle for relevant horizons.

III. Fat Tails

Since the disturbance term on an econometric equation is the composite of everything not explicitly included in the model, GARCH errors could reflect a multitude of sins of omission. Whether or not to "treat" a significant diagnostic test outcome of ARCH or revise the overall model specification must depend on the context, including both the data series under analysis and the theoretical model guiding the research. A large shock treated as heteroscedacticity when estimating parameters is downweighted (as in ARCH MLE), in contrast to its role in say an adaptive Kalman Filter (as discussed e.g. by Struth, 1984). Agents in an ARCH world would presumably try to discern permanent from transitory changes in variances and hence seek to appraise whether or not outliers were informative (or perhaps simply represented measurement error).

While ARCH processes have fatter tails for $\{\varepsilon_t\}$ than the Normal, the evidence in the paper and in Bollerslev (1985) suggests that there remains a need to assume fatter tails than Normal in the conditional distributions. Like Prucha and Kelejian (1984) they consider 't'-distributions and obtain improved fits. This suggests a need to discount some of the larger shocks rather than allow them full (inverse) weight in estimation: alternatively, in a loose sense very large shocks do not cause as much variance persistence as the average shock. Or relating back to Section II, while GARCH generates more outliers than the Normal, it also generates more small values (these persist too) which together are insufficiently leptokurtic for the data, hence necessitating assuming an initially leptokurtic distributional form. (Other discussants noted alternative functional forms for GARCH itself, including log GARCH and these may generate more leptokurtic outcomes.)

IV. The Origins of ARCH in Models

Many data series manifest ARCH (according to eye-ball tests at least) and modelling this by ARCH errors is clearly one potential route. The authors are careful to distinguish

cases where ARCH only influences the variance from where it also (or instead) enters the mean. In view of their findings on ARCH for spot prices it would have been very interesting to see tests of their theory model of how forward prices are determined (which I found somewhat unpersuasive in the absence of evidential support, given its switch of exogeneity between p_t and q_t and the assumed fixed stock of a risky asset like foreign currency).

When data are highly autoregressive or even integrated of order one, denoted I(1) (see Granger, 1981) there must be an important interaction between functional form assumptions and the occurrence of (I)(G)ARCH errors, analogous to that between heteroscedasticity and functional form in static models. Thus, some tests of ARCH-type assumptions seem merited. In terms of the expository model, instead of the usual tests for σ^2 being constant, tests for the constancy of (ω, α) should be employed. Similarly, one might attempt to discriminate between conditional and unconditional heteroscedasticity/ functional form mis-specification (as in White, 1980). For the US M1 data in Baba et al. (1985), the ARCH-like behaviour of the percentage growth in M1 is modelled by corresponding patterns in interest rates and hence is treated as a property of the mean - but models of interest rates may in turn require ARCH errors to explain their rapid changes in variances. Thus, the usual modelling problems of choices between representations recur and so testing against reasonable alternatives remains important.

V. Conclusions

(G)ARCH models are an important contribution to modelling varying variances, with implications as spelt out by Engle and Bollerslev for forecast confidence intervals, valid inferences, and efficient estimation, and potential applications to many economic phenomena. I also consider they have a useful role in illustrating econometrics concepts.

A further role is considered in Engle and Hendry (1985) for testing super-exogeneity. If β in (4) is asserted not to be an invariant, but to alter with changes in some variable z_t this could be tested directly. However, if a legitimate information set I_{t-1} for predicting z_t existed, then the dependence of β on the anticipated components $E(z_t|I_{t-1})$ and $var(z_t|I_{t-1})$ could be evaluated. A more powerful test would also check for dependence on unanticipated components of mean and variance corresponding to the decomposition of z_t into $\{E(z_t|I_{t-1})\}$ and $\{\nu_t = z_t - E(z_t|I_{t-1})\}$ (the latter being a conventional component of tests for weak exogeneity) and of ν_t^2 into $\{E(\nu_t^2|I_{t-1})\}$ and $\{\nu_t^2 - E(\nu_t^2|I_{t-1})\}$. The variance terms could fruitfully be modelled by a (G)ARCH process, and the deviation therefrom. Doubtless, many further roles for (G)ARCH models will appear in due course.

David F. Hendry
Nuffield College, Oxford
August 1986

ADDITIONAL REFERENCES

Baba, Y., D.F. Hendry and R.M. Starr, (1985). U.S. Money Demand, 1960-1984, Discussion Paper, University of California at San Diego.

Engle, R.F. and D.F. Hendry, (1985). Testing Superexogeneity and Invariance, Discussion Paper, Nuffield College, Oxford.

Granger, C.W.J., (1981). Some Properties of Time-Series Data and Their Use in Econometric Model Specification, Journal of Econometrics, 16, 121-130.

Hendry, D.F. (ed.), (1986). Econometric Modelling with Cointegrated Variables, Special Issue, Oxford Bulletin of Economics and Statistics, 48, 3.

Prucha, I.R. and H.H. Kelejian, (1984). The Structure of Simultaneous Equations Estimators: A Generalization towards Nonnormal Disturbances, Econometrica, 52, 721-736.

Struth, F., (1984). Modelling Expectations Formation with Parameter-Adaptive Filters. An Empirical Application to the Livingston Forecasts, Oxford Bulletin of Economics and Statistics, 46, 211-239.

White, H., (1980). A Heteroscedastic-Consistent Covariance Matrix Estimator, and a Direct Test for Heteroscedasticity, Econometrica, 48, 421-448.

ECONOMETRIC REVIEWS, 5(1), 71-74 (1986)

COMMENT

The authors discuss an interesting class of models called the Generalized Autoregressive Conditionally Heteroskedastic (GARCH) models, in which both the conditional mean and the conditional variance of a time series are functions of the past behavior of the time series. The popularity of the ARCH model is clear from the number of applications published since it was introduced in 1982. Motivated by a simple asset pricing theory the authors introduce two new models: (i) integrated in variance and (ii) nonlinear conditional variances. The paper highlights the relevance of conditional heteroskedastic models and raises some important open problems.

The problem of modelling conditional variance is a very important topic and deserves a serious consideration. A unified approach to the problem is warranted. Consider a simple AR(1) model $Y_t = \rho Y_{t-1} + \epsilon_t$ where the innovations ϵ_t are uncorrelated $(0,\sigma^2)$ random variables. The models considered in the paper assume that $E[\epsilon_t|\Psi_{t-1}] = 0$ and $E[\epsilon_t^2|\Psi_{t-1}] = h_t$ where Ψ_{t-1} is the information up to time t-1 and h_t is a function of the past. They considered two formulations for h_t, (i) GARCH (q,p) where $h_t = \omega + \Sigma_{i=1}^{q}\alpha_i\ \epsilon_{t-i}^2 + \Sigma_{j=1}^{p}\beta_j h_{t-j}$ and (ii) nonlinear conditional heteroskedasticity where h_t is a nonlinear function of ϵ_{t-j}^2 and h_{t-j}, $j > 0$. Let us consider a class of models given by $\epsilon_t = W_t h_t^{1/2}$ where W_t is a sequence of iid random variables with zero mean and unit variance. Note then that $E[\epsilon_t|\Psi_{t-1}] = 0$ and $E[\epsilon_t^2|\Psi_{t-1}] = h_t$. In the paper it is assumed that the distribution of W_t is either the standard normal distribution or a student's t distribution.

0747-4938/86/0501-0071$3.50/0

Suppose we assume that $\{\epsilon_t^2\}$ is a regular stationary time series with an absolutely continuous spectral density. Then we can approximate the correlation structure of the time series $\{\epsilon_t^2\}$ by that of a stationary ARMA process. Note that $\epsilon_t^2 = h_t + v_t$ where $v_t = (W_t^2-1)\, h_t$ is a sequence of uncorrelated random variables. If $h_t = \omega + \Sigma_{i=1}^q \theta_i\, \epsilon_{t-i}^2 - \Sigma_{j=1}^p \beta_j v_{t-j}$ then the process ϵ_t^2 has the same correlation structure as that of an ARMA(q,p) process with the AR parameters θ_i and the MA parameters β_j. Therefore, for $q \geq p$, the square of a GARCH(q,p) innovation process has the correlation structure of an ARMA(q,p) process with the AR parameters $\theta_i = \alpha_i + \beta_i$ for $i \leq p$, $\theta_i = \alpha_i$ for $i > p$ and the MA parameters β_j.

A more general model is to allow for a function f of ϵ_t^2 to be a regular stationary process. For example, we may assume that $X_t = \ln \epsilon_t^2$ is a regular stationary process with an absolutely continuous spectral density. Therefore the correlation structure of X_t can be well approximated by that of an ARMA process. Note that $X_t = \ln h_t + a_t$ where $a_t = \ln W_t^2$ is a sequence of iid random variables. If we now let

$$\ln h_t = c + \Sigma_{i=1}^q \gamma_i X_{t-i} - \Sigma_{j=1}^p \delta_j (X_{t-j} - \ln h_{t-j})$$

then the time series $X_t = \ln \epsilon_t^2$ has the same correlation structure as that of an ARMA(q,p) process with the AR parameters γ_i and the MA parameters δ_j. Assume that $q \geq p$, and let $\alpha_i = \gamma_i - \delta_i$ for $i \leq p$ and $\alpha_i = \gamma_i$, $i > p$ and δ_j. Also, assume that $\Sigma_{i=1}^q \alpha_i + \Sigma_{j=1}^p \delta_j < 1$. Note then the ℓog-GARCH(q,p) is given by $\epsilon_t = W_t\, h_t^{1/2}$ where

$$h_t = e^c \left[\pi_{i=1}^q (\epsilon_{t-i}^2)^{\alpha_i}\right] \left[\pi_{j=1}^p (h_{t-j})^{\delta_j}\right] ,$$

and $\{W_t\}$ is a sequence of iid random variables. We can define ℓog-GARCH(q,p) to be <u>integrated in variance</u> if $\Sigma_{i=1}^q \alpha_i + \Sigma_{j=1}^p \delta_j = 1$. It can be shown that if $W_t = Z_t$ or $W_t = \text{sgn}(Z_t - 0.5)\exp[0.5(Z_t - 0.5)]$ where $\{Z_t\}$ is a sequence of standard normal variables then both $X_t = \ln \epsilon_t^2$ and ϵ_t^2 have <u>all of the moments finite</u>. Since the innovations a_t in the model for X_t are iid the traditional time series identification and estimation procedures can be used to model the ℓog-GARCH(q,p) process. For the ℓog-GARCH model the marginal probability density of $\{\epsilon_t^2\}$ can be derived. Consider now the problem of testing for unit roots.

<u>1. Testing for unit roots in the mean model</u>:

If the innovation process is either GARCH(q,p) or ℓog=GARCH(q,p) <u>with</u> $\Sigma_{i=1}^q \alpha_i + \Sigma_{j=1}^p \delta_j < 1$ then the Dickey-Fuller criterion for testing $\rho = 1$ based on the ordinary least squares estimator $\hat{\rho}_{OLS}$ is valid. If the innovation process is integrated in variance then $n^{-1}\, \Sigma_{t=1}^n \epsilon_t^2$ converges to a random variable and hence I do not believe that the Dickey-Fuller criterion will be valid to test $\rho = 1$.

<u>2. Testing for unit roots in the variance model</u>:

Since the innovations a_t in the ℓog-GARCH(q,p) model are iid, Dickey and Fuller (1979) and Said and Dickey (1984) procedures can be used to test the integrated in variance hypothesis. However, there is no reason to believe why such tests should be valid when the innovation model is GARCH(q,p).

Finally, the maximum likelihood procedure suggested in the paper requires the evaluation of numerical derivatives. Also, the ML estimators may be sensitive to the distributional assumptions. This is because higher moments enter the information matrix. A 4-stage least squares procedure similar to the one suggested in Pantula (1985) may be more robust.

I must congratulate the authors for their contribution to the econometric literature. In the next few years, much progress is sure to be made on these and other alternative approaches to modelling conditional heteroskedasticity. In this process of advancement, the ideas provided by Engle and Bollerslev are important and thought provoking.

Sastry G. Pantula
North Carolina State University

ADDITIONAL REFERENCES

Pantula, S. G. (1985). Estimation of autoregressive models with ARCH errors. Technical Report, North Carolina State University, Raleigh, NC.

Said, S. E. & Dickey, D. A. (1985). Hypothesis testing in ARIMA (p,1,q) models. *Jour. of the Amer. Stat. Assoc.*, 80, 369-374.

ECONOMETRIC REVIEWS, 5(1), 75-80 (1986)

COMMENT

One would think that providing a summary of a literature that has had only a few years to develop would be a relatively simple task. For the literature on ARCH models, this is clearly not the case. Upon my first reading of this paper by Professors Engle and Bollerslev, I was struck by the sheer quantity of high quality research that has emerged since Professor Engle's original paper was published in 1982 -- the authors cite thirty-six papers that either apply or extend the ARCH framework. This is, however, just the tip of the iceberg as the LM-test for ARCH errors has become a routinely applied regression diagnostic. Therefore, even though the literature is still in its infancy, it seems an appropriate time to take stock of what has been accomplished with this model and to set a course for future research in the area.

The paper provides a very readable survey of both the statistical aspects of modelling conditional variances and the economic implications of these models for simple equilibrium asset pricing equations. Further, the paper explores a number of interesting extensions to the familiar ARCH and GARCH specifications: (i) a new class of models that are integrated in variance is proposed as a way of parameterizing conditional variance processes with long memories, (ii) models with conditional Student-t densities are employed to allow the degree of conditional kurtosis to be estimated rather than imposed, and finally, (iii) parameterizations for the conditional variance process that allow for nonlinearities in lagged innovations are also proposed. I will focus my discussion on the first of these extensions concentrating on some of the unresolved issues that the authors have raised.

0747-4938/86/0501-0075$3.50/0

The authors motivate their analysis of integrated variance models in part by noting an empirical regularity in the conditional variance process of a GARCH(1,1) model applied to weekly changes in the spot Dollar/Swiss Franc exchange rate. The sample autocorrelations and partial autocorrelations of the squared innovations process remain large for long lags. Further, the parameter estimates for the variance process in the fitted GARCH(1,1) model have a sum close to one which is a necessary condition for an integrated variance model. This property of exchange rates does not appear to be specific to the particular currencies or data set chosen in their example. McCurdy and Morgan (1985, 1986) present similar evidence for GARCH(1,1) models applied to both futures and spot prices for a wide variety of currencies measured both daily and weekly. These inferences, as noted by the authors, are strictly casual since a formal testing procedure has not yet been developed. The methodology would obviously be more complete if the theoretical issues surrounding tests of this hypothesis were resolved. The message in the data, however, seems sufficiently strong to indicate that conditional variances demonstrate a degree of persistence that is well approximated by an integrated variance process. If we, therefore, assume that persistent variance models are an important consideration when characterizing risk and uncertainty in economics, a number of statistical and economic issues arise as a result. I will briefly address two such issues.

One of the most popular applications of ARCH and GARCH models has been in estimating time-varying risk premiums in financial data. As noted in the paper this application is a natural implication of simple equilibrium asset pricing theory. By including conditional second moments in the mean of a regression, risk premiums are modelled exactly, as in the asset pricing theory provided in the paper, or at the very least they are well approximated. The authors cite numerous applied papers that exploit these time varying second moments to capture market returns to bearing risk. Unfortunately, in the empirical examples provided in the paper there is no explicit role played by risk premiums. One could, however, envision applications in which the innovations follow an integrated GARCH(1,1) process, and the conditional variance enters the regression mean through a time-varying risk premium. Two fundamental problems arise in this situation. First, the identification of the order of the conditional variance process and the search for integration in this process through the inspection of the correlation structure of the squared innovations may prove unsuccessful since the innovations or forecast errors cannot be distinguished from risk premiums until the GARCH process is itself specified. This problem is evident in specifications such as equation (14) of the paper. The magnitude of the effect this has on the estimated correlations is difficult to anticipate but is sure to be problem specific. A second (and perhaps more serious) problem involves the statistical properties of a maximum likelihood estimator for an integrated GARCH-M model. As noted by the authors with respect to the estimated model in equation (24), under the null hypothesis of an integrated GARCH process the unconditional fourth moment of the innovation process is infinite. Therefore, including as a regressor the conditional second moment of the innovation process, which is a function of the squared innovation process, implies a regressor with an infinite population second moment, i.e., the fourth moment of the innovation. This clearly violates the classical conditions most commonly employed to justify asymptotic normality of maximum likelihood parameter estimators. For these reasons, it would seem that an applied researcher must approach the identification and estimation of a GARCH-M model when integration in variance is a consideration with extreme caution.

A second issue that I would like to address relates to the economic environment that could in fact be generating time series that exhibit integration in variance. At the present time it is difficult to point to an economic theory that has integration in variance as a direct implication as is the case with integration in the mean [for example Hall (1978) derives integration in the mean of consumption as a consequence of intertemporal optimization by households]. This is obviously a difficult problem well beyond the scope of this short comment. However, in an attempt to shed some light on this issue, I will proceed in a somewhat different direction. By generating data from a fairly well received asset pricing model we can examine the correlation structure of squared innovation processes and perhaps learn something about the type of economic behaviour that could generate persistent conditional variances. I have, therefore, constructed a simple example from a dynamic stochastic general equilibrium model of the term structure of interest rates based on the work of Backus, Gregory and Zin (1986). A representative agent maximizes time additive expected utility over his infinite lifetime consuming a stochastic endowment of a single commodity in each period. Relative prices of state-contingent claims are derived from marginal rates of substitution and the size of the endowment each period is determined by a finite state, stationary Markov chain. For complete details of the model and its calibration, the interested reader is referred to Backus, Gregory and Zin (1986). For the purpose at hand, it is sufficient to note that when calibrated with actual time series data this asset pricing model generates price data that closely mimic the behaviour of actual asset price series. It would be of some interest to examine whether such a model is capable of generating the persistent conditional variances that are so prevalent in foreign exchange data.

In the following example the artificial economy had three states for the growth rate of the consumption endowment with a high degree of persistence specified for the conditional process generating these growth rates of endowments. The average sample autocorrelations and partial autocorrelations for the square of the innovation to the spot price for a two-period discount bond were computed. The experiment was replicated 1000 times and 500 time series observations were generated at each replication. The average correlation structure is as follows:

Lag	1	2	3	4	5	8	10
ρk	.23	.11	.05	.02	.007	.0004	-.001
φ_{kk}	.23	.06	.01	-.001	-.002	-.002	-.003

$$\sigma_a = T^{-\frac{1}{2}} = .047$$

It is clear from these results that this economy failed to generate the type of persistence in variance that would require an integrated variance specification. Combining the risk premium with the innovation process did not contribute appreciably to the persistence in the variance. Further, these autocorrelations are indicative of the behaviour of economies with different Markov forcing processes and different degrees of risk aversion.

It is extremely difficult to generalize from such a stylized model, but perhaps two obvious points can be made. First, the persistence of the conditional variance seems to depend directly on the serial dependence of the forcing process. In this case, a first order process was assumed based on the observed time series behaviour of growth rates in aggregate consumption. A higher order process could conceivably generate conditional variances with more persistence, however, the model would be poorly calibrated. Second, even though this model is capable of generating data with first moment properties that closely resemble actual term structure data, the second moment properties of the artificial data do not reflect the behaviour of second moments in exchange markets. The obvious implications being that this model is inappropriate for describing foreign exchange behaviour on a weekly or daily basis and also that consumption based models of asset pricing are not very likely to generate conditional variance processes with long memories. In general, it seems that little is known about the source of persistent variances and how they are related to the structure and the underlying sources of uncertainty in an equilibrium asset pricing model. Theoretical guidance would seem to be especially valuable in this context given the unresolved statistical issues surrounding testing the integrated variance hypothesis.

Stanley E. Zin

Queen's University

ADDITIONAL REFERENCES

Backus, D.K., Gregory, A.W., & Zin, S.E., (1986). Risk Premiums in the Term Structure: Evidence from Artificial Economies. Manuscript, Queen's University.

Hall, R.E., (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. Journal of Political Economy, 86, 971-87.

McCurdy, T.H. & Morgan, I.G., (1986). Tests of the Martingale Hypothesis for Foreign Exchange Futures. Manuscript, Queen's University.

ECONOMETRIC REVIEWS, 5(1), 81-87 (1986)

REPLY

We would like to thank the discussants for their careful and insightful comments on our paper and in fact on the whole development of conditionally heteroskedastic models in economics. Several themes reappeared in most of the discussions and these we would like to treat first. We would like to discuss the economic underpinning of GARCH or Integrated GARCH models, the simulation properties of IGARCH models, and the appropriateness of the log GARCH formulation.

Most of the discussants particularly Hendry, Diebold and Zin, sought an economic model which could motivate the GARCH model or more specifically , the IGARCH model. We agree entirely with this objective. Our first response however is that it is not heteroskedasticity but homoskedasticity which needs to be justified as it is a restriction on the possible conditional densities. Once heteroskedasticity is taken as a potential complication of real data, the question of the parameterization and modelling of the process becomes important. A simple approach is to assume that there is an unmeasured variable which causes the heteroskedasticity but that it is only slowly evolving. In this context, it can be shown that an optimal predictor of the variance

0747-4938/86/0501-0081$3.50/0

would depend only on past residuals, not on the full information set. The form of such dependence hinges on the process of this unobservable cause. Possibly if it has a unit root, the variance process would also be integrated.

Hendry also tacitly suggested another approach which is to consider the variance process as composed of permanent and transitory components. Such a model in time series is an unobserved components model with a unit root in the permanent component. The observable process therefore also has a unit root. Applying this approach to the GARCH models might lead to more complex dynamics with a unit root in variance.

A complimentary justification can be found in the continuous time stochastic differential equations used so often in analysis of financial data.

$$dy = f(y,\sigma)dt + g(y,\sigma)dz_y$$

$$d\sigma = k(y,\sigma)dt + l(y,\sigma)dz_\sigma$$

where dz are standard Brownian motion, y is the asset price and $g(y,\sigma)$ is its instantaneous variance. The expectation of σ at some point T+h in the future made at time T, is given by a complicated Ito stochastic integral which will depend upon the state variables y_T and σ_T. Although the process is not a discrete time IGARCH, it has the persistence property of such processes as well as slowly changing variance.

A second point raised very persuasively by John Geweke is that a process with a constant second moment and a fourth moment which goes to infinity, must have most of its probability

distribution centered about zero. In the simplest integrated ARCH(1) model,

$$\epsilon_t|\psi_{t-1} \sim N(0,h_t), \quad h_t = \epsilon^2_{t-1}$$

it is clear that whenever ϵ equals zero, all subsequent ϵ's will be zero. Thus zero is an absorbing state. Furthermore since zero is the mean of the distribution and the variance is finite, it follows from Chebyschev's inequality that there is a positive probability of ϵ_{t+s} being arbitrarily close to zero and thereby getting a whole string of ϵ's close to zero. Geweke's simulations indicate that the density for this process masses about zero for very short series. After six steps the theoretical variance remains unity, but 90% of the distribution lies below .56 in absolute value while 99% lies below 4. Thus there are very long tails but the bulk of the probability lies very close to zero. For the GARCH(1,1)

$$h_t = \alpha\epsilon^2_{t-1} + (1-\alpha)h_{t-1}$$

with parameter α=.5, the corresponding numbers after 18 steps are .84 and 3.35 respectively. In this case it takes a string of small ϵ's to drive the conditional variance close to its absorbing state.

The estimated value of α from the foreign exchange data was much smaller than the values used by Geweke. Taking the estimate of .1 as the true value some additional simulations were carried out. Visual inspection of 1000 steps of the first three simulations confirms the general tendency but notes that it occurs much later. All three showed bursts of volatility as one expects from GARCH processes for the first 400-800 observations, however the amplitude in each case decreased substantially by number 1000. As our data set was 600 observations long, this is exactly the relevant region. A histogram of the 600^{th} step from 2000 simulations revealed a concentrated distribution with a 90% point of .41 and a 99% point of 2.46. Thus again we see a concentrated distribution but not one which has collapsed to zero. Although many of the simulations decrease in amplitude, a few become very large. Interestingly, there is no natural starting point for such a series; a scaled picture of the distribution starting at any point in its history will look the same. The problem is much like the random walk model which has theoretically infinite unconditional variance, but conditional on a starting value, the variance is finite and the distribution is centered.

A possible implication of this analysis is that the drift term in the IGARCH formulation should be allowed. With any drift, zero is no longer an absorbing state and further simulations suggested that a very small intercept would eliminate the tendency of the distribution to collapse to zero. Furthermore, the empirical results we report suggested that the data also support a drift parameter for the model.

The third point raised in both Geweke and Pantula's discussions was an alternative parameterization of the ARCH process to deal with the non-negativity restrictions inherent in variance models. They suggest a log-GARCH formulation:

$$\epsilon_t|\psi_{t-1} \sim N(0,h_t), \quad \log h_t = \omega + \alpha \log \epsilon_{t-1}^2 + \beta \log h_{t-1}$$

which of course could have its integrated version. No longer must the parameters ω,α,β be positive. Geweke shows that the log likelihood is globally concave with respect to ω and α, and Pantula shows that $\ln \epsilon_t^2$ may be assumed to be a regular stationary process with an absolutely continuous spectral density which may be approximated by an ARMA or ARIMA process with independent and identically distributed innovations. Consequently, standard identification, estimation and testing procedures can be used, including Dickey Fuller types of tests for the unit root in variance. We think this line of modelling should be pursued but have several reasons to believe it will not provide a useful class of models. These reasons are again centered on the possibility that ϵ will be close or exactly equal to zero and therefore $\log \epsilon$ will go to minus infinity.

In the log-GARCH model with $\alpha > 0$, zero is again an absorbing state since any single zero in the ϵ process will make the subsequent h equal to zero and thus all succeeding ϵ's will be zero. Notice that one zero collapses the subsequent distribution regardless of the size of α. Of course if $\alpha < 0$ any $\epsilon_t = 0$ will lead to an infinite variance next period. The tendency for the integrated log model to decay to zero is even more pronounced in Geweke's simulations with $\alpha > 0$. Although the probability of $\epsilon_t = 0$ is zero, in practical applications with observable and inherently discrete data, that is not the case. Indeed for the exchange rate data studied in the paper, nine of the changes were exactly zero. When ϵ is a residual from a linear regression model such as

$$\epsilon = y - x\theta$$

then the log likelihood with respect to θ will be particularly ill-behaved. For any value of θ where one or more ϵ's are zero, the log likelihood will be minus infinity. For a single explanatory variable there will generally be a pole for each observation so that the log likelihood will be as rough as a saw blade. It is our belief that these complexities will outweigh the previously mentioned advantages of the log model.

Diebold, Hendry and Zin each felt the need for additional guidance in specifying a GARCH model. Specification tests of various kinds such as tests for serial correlation in the presence of ARCH, for constantancy of the ARCH parameters, for the lag length and structure of the GARCH process as well as tests for the degree of integration in variance were requested. In fact, all of these are easily available as Lagrange Multiplier tests and several were calculated in the paper. Tests for serial correlation and tests for the order of the GARCH model were regularly computed and are reported. Both Diebold and Zin were concerned that asymptotic theory would break down in the presence of the IGARCH model. In fact, Diebold went so far as to call it "dangerous". The IGARCH-M model would indeed have a regressor with infinite fourth moment as an independent variable and therefore the test statistics might be non-Gaussian. This however remains to be shown since the model is essentially transformed by the same variable for heteroskedasticity. In any case a damped or bounded function of the variance could turn out to be the preferred

functional form. Finally, we would agree with Diebold that the multivariate version of this model possibly with factor structure or integrated or co-integrated multivariate variance processes is an interesting next step. It should serve to better deal with the portfolio issues which are so important in financial analysis.

In conclusion, we would like to thank the discussants for extremely interesting and constructive comments.

Robert F. Engle
University of California, San Diego

Tim Bollerslev
Northwestern University