

## NEW FRONTIERS FOR ARCH MODELS

Robert Engle  
Professor of Finance, NYU  
Professor of Economics, UCSD

Prepared for conference

Modeling and Forecasting Financial Volatility  
Perth, Australia  
September 2001  
This Draft June 6, 2002

### ABSTRACT

In the 20 years following the publication of the ARCH model, there has been a vast quantity of research uncovering the properties of competing volatility models. Wide-ranging applications to financial data have discovered important stylized facts and illustrated both the strengths and weaknesses of the models. There are now many surveys of this literature.

This paper looks forward to identify promising areas of new research. The paper lists five new frontiers. It briefly discusses three – high frequency volatility models, large-scale multivariate ARCH models, and derivatives pricing models. Two further frontiers are examined in more detail – application of ARCH models to the broad class of non-negative processes, and use of Least Squares Monte Carlo to examine non-linear properties of any model that can be simulated. Using this methodology, the paper analyzes more general types of ARCH models, stochastic volatility models, long memory models and breaking volatility models. The volatility of volatility is defined, estimated and compared with option implied volatilities.

Keywords: ARCH, GARCH, volatility, non-linear process, non-negative process, option pricing, stochastic volatility, long memory, Least Squares Monte Carlo, ACD, Multiplicative Error Model, MEM

Who could imagine 20 years ago, the flowering of research and applications that would develop around the ARCH model? It certainly was not an instant success. Years went by before anyone except my students and I wrote a paper on ARCH. But as applications shifted to financial markets, and as richer classes of models were developed, researchers saw how volatility models could be used to investigate the fundamental questions in finance. How are assets priced and what is the tradeoff between risk and return? ARCH models offered new tools for measuring risk, and its impact on return. They also provided new tools for pricing and hedging non-linear assets such as options.

This conference and this paper are designed to reflect on these developments and look forward to the next important areas of research. In this paper I will give a rather brief idiosyncratic assessment of the important accomplishments of the last 20 years in volatility modeling. Then I will point to five frontiers on which I think new developments can be expected in the next few years. For two of these, I will give some new results to show the directions I see developing.

## I. WHAT WE HAVE LEARNED IN 20 YEARS

The number of new models proposed, estimated and analyzed has been dramatic. The alphabet soup of volatility models continually amazes. The most influential models were the first: the GARCH model of Bollerslev(1986), and the EGARCH of Nelson(1991). Asymmetric models of Glosten, Jaganathan Runkle(1993) Rabemananjara and Zakoian(1993), Engle and Ng(1993) and power models such as Higgins and Bera(1992), Engle and Bollerslev(1986), and Ding Granger and Engle(1993) joined models such as SWARCH, STARCH, QARCH and many more. The linguistic culmination might be that of Figlewski(1996), the YAARCH model – an acronym for Yet Another ARCH model. Coupled with these models was a sophisticated analysis of the stochastic process of data generated by such models as well as estimators of the unknown parameters. Theorems for the autocorrelations, moments and stationarity and ergodicity of these processes have been developed for many of the important cases; see for example Nelson(1990), and Ling and McAleer(2002a,2002b). Work continues and new models are continually under development, but this is a well-studied frontier. The limiting distribution of the MLE for GARCH models waited for Lumsdaine(1996) and Lee and Hansen(1994) for rigorous treatments. There are now a collection of survey articles that give a good appreciation of the scope of the research. See for example, Bollerslev, Chou and Kroner(1992), Bollerslev Engle and Nelson(1994), Bera and Higgins(1993), and recent pedagogical articles by Engle(2001) and Engle and Patton(2001). A very recent survey is Li, Ling and McAleer(2002)

Another topic for ARCH models is their usefulness in trading options. It was initially supposed that volatility models could give indications of mispricing in options markets leading to trading opportunities. Early studies such as Engle, Kane and Noh(1994) suggested the profitability of such strategies. More recent data fails to find evidence of significant trading opportunities, at least in the US index options market.

This is not surprising since GARCH models have a limited information set and are available to all traders today. The same question is often asked in terms of forecast accuracy. Do GARCH models out-forecast implied volatility models? The answer is complex depending upon the statistical approach to forecast evaluation, but generally it is found that implied volatilities are more accurate forecasts of future volatility than are GARCH models. See for example Granger and Poon(2001) for a survey.

The theory of asset pricing is based upon the reward for bearing risk. ARCH models have been developed to measure the price of risk. The first such model was the univariate ARCH-M model of Engle, Lilien and Robins(1987). Estimation of the CAPM began with Bollerslev, Engle and Wooldridge(1988) and has been extended and improved by a series of interesting papers including McCurdy and Stengos(1992), Engel Frankel, Froot and Rodriguez(1995), and de Santis, Gerard and Hillion(1997).

With the introduction of Value at Risk, a new role for ARCH models emerged. A variety of studies examined the usefulness of volatility models in computing VaR and comparing these methods with the exponential smoothing approach favored by Riskmetrics. See for example Christoffersen and Diebold(2000), Christoffersen, Hahn and Inoue(2001) and Alexander(1998). GARCH methods proved successful but suffered if errors were assumed to be Gaussian.

These chapters of research on ARCH models are full and may have reached the point of diminishing returns. However, new directions are always available and these are the main focus of this paper.

## II. FIVE NEW FRONTIERS

Five new frontiers are identified below. These are areas where substantial research can be expected over the next few years. The problems are important, soluble, and already have some important new papers. For two of the areas, I will give some new results suggesting a possible direction for future research.

### A. HIGH FREQUENCY VOLATILITY MODELS

The study of volatility models within the day is in its infancy yet is a natural extension of the daily models examined so widely. Several alternative formulations have been introduced including Andersen and Bollerslev (1997) and Ghosh and Kroner(1995). Such models focus on the time of day or “diurnal” effect and have the requirement that they be useful forecasting many days in the future. These models have regularly spaced observations in calendar time but ultimately it will be desirable to find models based on irregularly spaced data as this is the inherent limit of high frequency data. Engle(2000) calls such tick data “ultra high frequency” data and gives some models which indicate that the arrival rate of trades, spreads and other economic variables may be important variables for forecasting volatility at this frequency. Such a model could give a continuous record of instantaneous volatility where events such as trades and quote revisions as well as time itself, modify the volatility estimate.

Continuous time models are ubiquitous in financial theory and derivative pricing. However most estimation of these models begins with equally observed prices and focuses on the mean process possibly with jumps. Continuous time stochastic

volatility models, possibly with volatility jumps are a new class of models with interesting derivative implications.

In addition to these models, there is now increasing interest in using intra-daily data to estimate better daily models. Andersen Bollerslev Diebold and Labys(2001) for example build models based upon “realized volatility” and Andersen and Bollerslev(1998) use this measure to evaluate traditional GARCH specifications.

## B. MULTIVARIATE MODELS

Although the research on multivariate GARCH models has produced a wide variety of models and specifications, these have not yet been successful in financial applications as they have not been capable of generalization to large covariance matrices. As computation becomes cheaper, and new parsimonious models are formulated, the potential for building ever larger time varying conditional covariance and correlation matrices increases. Models such as the vec and BEKK model of Engle and Kroner(1995) have attractive properties as linear systems. The constant conditional correlation (CCC) model of Bollerslev(1990) has the attraction of computational simplicity. A new model called Dynamic Conditional Correlation (DCC) by Engle(2002) combines some of these features to introduce a parsimonious correlation model to go with a conventional volatility model. Engle and Sheppard(2001) estimate and test models of up to 100 assets. Ledoit and Santa-Clara(1998) combine bivariate models to form multivariate models in a way which can be greatly expanded.

Correlation models can be estimated directly on intraday data. However as the frequency increases, the asynchronicity of trades and returns leads to a serious underestimate of comovements. This has been observed since Epps(1979) and the solutions of Scholes and Williams(1977) are widely employed in spite of both theoretical and empirical difficulties. These are not appropriate for ultra high frequency data and new solutions must be found.

## C. OPTIONS PRICING AND HEDGING

The pricing of options when the underlying asset follows a GARCH model is a topic of future research. Most approaches are based on simulation but the appropriate approach to risk neutralization must be investigated. This was barely addressed in Engle and Mustafa(1992) when they simply simulated GARCH returns with the riskless rate as mean, calculating an “implied GARCH” model. Duan(1995)(1997) in a series of papers has proposed a local risk neutralization which is based on the assumption that quadratic utility is a good local approximation to the representative agent’s risk preferences. Engle and Rosenberg(2000) develop hedging parameters for GARCH models and Rosenberg and Engle(2001) jointly estimate the pricing kernel and the empirical GARCH density. This paper is in the line of several which use both options and underlying data to estimate both the risk neutral and objective densities with ever more complex time series properties in an attempt to understand the enormous skew in index options volatilities.

An alternative strategy is the GARCH tree proposed by Richken and Trevor (1999) which adapts binomial tree methods for the path-dependence of GARCH

models. These methods involve a variety of assumptions that must be examined empirically. They are computationally much faster than the simulation estimators discussed above.

### III. MULTIPLICATIVE ERROR MODELS FOR MODELING NON-NEGATIVE PROCESSES

#### A. INTRODUCTION

GARCH type models have been introduced for other variables. Most notable is the Autoregressive Conditional Duration model ACD of Engle and Russell(1998) that surprisingly turns out to be isomorphic to the GARCH model. In this section, I explore a much wider range of potential applications of GARCH type models for any non-negative time series process.

Frequently we seek to build models of time series that have non-negative elements. Such series are generally common and are particularly common in finance. For example, one could model the volume of shares traded over a 10-minute period. Or one might want to model the high price minus the low price over a time period, or the ask price minus the bid price, or the time between trades, or the number of trades in a period, or many other series. There are two conventional approaches to this problem: the first is to ignore the non-negativity, and the second is to take logs. We discuss the disadvantages of these approaches.

Consider a time series  $\{x_t\}, t = 1, \dots, T$ , where  $x_t \geq 0$  for all  $t$ . Suppose in addition that

$$P(x_t < \mathbf{x} | x_{t-1}, \dots, x_1) > 0, \text{ for all } \mathbf{x} > 0, \text{ and for all } t \quad (1)$$

which says that the probability of observing zeros or near zeros in  $x$  is greater than zero. Let the conditional mean and variance of the process be defined as:

$$\mathbf{m}_t \equiv E(x_t | x_{t-1}, \dots, x_1), \quad \mathbf{s}_t^2 \equiv V(x_t | x_{t-1}, \dots, x_1) \quad (2)$$

A linear model is given by

$$x_t = \mathbf{m}_t + \mathbf{e}_t, \quad \mathbf{e}_t | \mathfrak{S}_{t-1} \sim D(0, \mathbf{s}_t^2) \quad (3)$$

It is clear that the distribution of the disturbances must be carefully specified. Since the mean is positive and  $x$  is non-negative, the disturbances cannot be more negative than the mean. Thus the range of the disturbance will be different for every observation. The variance and other higher moments are unlikely to be constant. Efficient estimation via Maximum Likelihood is going to be very difficult, although least squares will remain consistent. The probability of a near zero is given by

$$P_{t-1}(x_t < \mathbf{x}) = P_{t-1}(\mathbf{e}_t < \mathbf{x} - \mathbf{m}_t)$$

hence the error distribution must be discontinuous at  $-\mathbf{m}_t$  in order to satisfy (1).

The second conventional solution is to take logs. The model might then be written as

$$\log(x_t) = m_t + u_t, \quad (4)$$

where

$$\mathbf{m}_t = e^{m_t} E(e^{u_t}), \quad \mathbf{s}_t^2 = e^{2m_t} V(e^{u_t}). \quad (5)$$

This solution will not work if there are any exact zeros in  $\{x_t\}$ . Sometimes a small constant is added to eliminate the zeroes. However, this is more of a theoretical solution than a practical one since the finite sample estimates are typically heavily influenced by the size of this constant. Furthermore, the assumption in (1) that observations very near zero are possible, requires that  $P(u < -A) > 0$ , for all  $A > 0$ . This is only true of very peculiar distributions.

Estimation of (4) requires the specification of both  $m$  and  $u$ . Clearly the relation between  $m_t$  and  $\mathbf{m}_t$  depends upon the distribution of  $u$ . Thus even one step forecasts require knowing the distribution of  $u$ .

## B. THEORETICAL MODEL

The proposed model, which solves all these problems, is the multiplicative error model that could be abbreviated as MEM. This model specifies an error that is multiplied times the mean. The specification is

$$x_t = \mathbf{m}_t \mathbf{e}_t, \quad \mathbf{e}_t | \mathcal{F}_{t-1} \sim D(1, \mathbf{f}_t^2) \quad (6)$$

thereby automatically satisfying (2). The range of the disturbance would naturally be from zero to infinity thereby satisfying (1). If the disturbance is *i.i.d.*, then the variance of  $x$  is proportional to the square of its mean. This is a strong restriction but is not in conflict with other parts of the model. If it is not *i.i.d.* then a non-negative distribution with a unit mean and time varying variance can be specified. There are many candidates. The residual in  $x$  is naturally measured as the proportional deviation from the estimated mean as the standardized residual,  $x_t / \hat{\mathbf{m}}_t$ . This would be homoskedastic although the additive residual  $x_t - \hat{\mathbf{m}}_t = \hat{\mathbf{m}}_t (\mathbf{e}_t - 1)$  would not.

Vector models can be formulated in just the same way. Let an arrow represent a column vector and let  $\odot$  represent the Hadamard product of two matrices, which is element-by-element multiplication.

$$\vec{x}_t = \vec{\mathbf{m}}_t \odot \vec{\mathbf{e}}_t, \quad \text{and} \quad V(x_t) = \vec{\mathbf{m}}_t \vec{\mathbf{m}}_t' \odot V(\vec{\mathbf{e}}_t) = \text{diag}(\vec{\mathbf{m}}_t) V(\vec{\mathbf{e}}_t) \text{diag}(\vec{\mathbf{m}}_t) \quad (7)$$

Thus the positive definiteness of the covariance matrix is automatically guaranteed by the MEM structure.

Estimation of the MEM can typically proceed by maximum likelihood once the distribution of the disturbance has been specified. A natural choice of distribution is the exponential as it has non-negative support. An exponential random variable with mean one is called a unit exponential. Assuming that the disturbance is a unit exponential, then the univariate log likelihood is simply

$$L(\mathbf{q}) = \sum_{t=1}^T \left[ -\log(\mathbf{m}_t(\mathbf{q})) - \frac{x_t}{\mathbf{m}_t(\mathbf{q})} \right] \quad (8)$$

where  $\mathbf{q}$  represents the vector of parameters to be estimated. The first order conditions for a maximum of this likelihood function are

$$\frac{\partial L}{\partial \mathbf{q}} = \sum_{t=1}^T \left[ \left( \frac{x_t - \mathbf{m}_t}{\mathbf{m}_t^2} \right) \frac{\partial \mathbf{m}_t}{\partial \mathbf{q}} \right] \quad (9)$$

By the law of iterated expectations, the expected value of the first order condition when evaluated at the true parameter value will be zero regardless of whether the density of  $x$  is truly unit exponential. This implies that the log likelihood in (8) can be interpreted as a Quasi Likelihood function and that parameters that maximize this are QMLE. Application of the theorem originally given in White(1980) requires regularity conditions on the mean function and its determinants, and gives general expressions for the covariance matrix.

A fairly general class of mean functions can be entertained for this problem. Suppose the mean is linear in lagged  $x$  and in a  $k \times 1$  vector of predetermined or weakly exogenous variables  $z_t$ . Then a  $(p,q)$  mean specification would be

$$\mathbf{m}_t = \mathbf{w} + \sum_{j=1}^p \mathbf{a}_j x_{t-j} + \sum_{j=1}^q \mathbf{b}_j \mathbf{m}_{t-j} + \mathbf{g}' z_t \quad (10)$$

The parameters of this model may be restricted to ensure positive means for all possible realizations, and to ensure stationary distributions for  $x$ . If  $z$  are positive variables, then sufficient conditions for non-negativity are clearly that all parameters are positive. However these are not necessary. See Nelson and Cao(1992) for an examination of sufficient conditions. Sufficient conditions for the covariance stationarity of  $x$  from Bollerslev, Engle and Nelson(1994) are that  $z$  is covariance stationary and

$$\sum_{j=1}^p \mathbf{a}_j + \sum_{j=1}^q \mathbf{b}_j < 1. \quad (11)$$

This result can be formalized from Engle and Russell(1998) based upon a theorem in Lee and Hansen(1994) for GARCH models. In their case,  $x$  is the duration between successive events, but the theorem applies to any non-negative process. In this theorem, the process is assumed to be a first order GARCH type model possibly with unit or explosive roots.

#### Corollary to Lee and Hansen(1994)

If

- 1)  $E_{t-1}(x_t) \equiv \mathbf{m}_{0,t} = \mathbf{w}_0 + \mathbf{a}_0 x_{t-1} + \mathbf{b}_0 \mathbf{m}_{0,t-1}$ ,  
 $\mathbf{e}_t \equiv x_t / \mathbf{m}_{0,t}$  is i) strictly stationary and ergodic  
ii) nondegenerate  
iii) has bounded conditional second moments  
iv)  $\sup_t E[\ln(\mathbf{b}_0 + \mathbf{a}_0 \mathbf{e}_t) | \mathcal{F}_{t-1}] < 0 \quad a.s.$
- 2)
- 3)  $\mathbf{q}_0 \equiv (\mathbf{w}_0, \mathbf{a}_0, \mathbf{b}_0)$  is in the interior of  $\Theta$
- 4)  $L(\mathbf{q}) = -\sum_{t=1}^T \left( \log(\mathbf{m}_t) + \frac{x_t}{\mathbf{m}_t} \right)$  where  $\mathbf{m}_t = \mathbf{w} + \mathbf{a} x_t + \mathbf{b} \mathbf{m}_{t-1}$  for  $t > 1$   
 $\mathbf{m}_t = \mathbf{w} / (1 - \mathbf{b})$  for  $t = 1$

Then:

- a) the maximizer of  $L$  will be consistent and asymptotically normal with a covariance matrix given by the familiar robust standard errors as in Lee-Hansen, and
- b) the model can be estimated with GARCH software by taking  $\sqrt{x}$  as the dependent variable and setting the mean to zero.
- c) the robust standard errors of Bollerslev and Wooldridge(1996) coincide with those in Lee and Hansen

From this corollary it is apparent that even mildly explosive models may be estimated consistently by QMLE. From an examination of the mean specification in (10) it is apparent that the (p,q) version of this MEM model with exogenous variables can also be estimated using GARCH software by making  $\sqrt{x_i}$  the dependent variable, specifying it to have zero mean and an error process assumed normal GARCH(p,q) with exogenous variables  $z$ . The estimated “conditional variance” is then the conditional mean of  $x$ . Multi-step forecasts of  $x$  are computed simply by multi-step forecasts of the conditional variance.

The clear advantage of the exponential error assumption is that estimation is consistent regardless of the correctness of this distribution. The disadvantage is that it is not fully efficient. However, it is perfectly straight-forward to formulate more general likelihood functions which allow more flexible shapes of density function or time variation in higher moments of this density function.

### C. EMPIRICAL MODELS

Several papers have already developed models based on the multiplicative error structure. Of course the first to mention is the family of ARCH and GARCH models themselves. The basic model common to all these processes<sup>1</sup> and its square are

$$\begin{aligned} r_t &= \sqrt{h_t} \mathbf{e}_t \\ r_t^2 &= h_t \mathbf{e}_t^2 \end{aligned} \tag{12}$$

In the squared version, the dependent variable is non-negative with mean  $h$  and a non-negative multiplicative *i.i.d.* error with unit mean. This can be estimated directly by taking the absolute value of returns as the dependent variable of a GARCH model.

The second paper in this direction is the ACD model of Engle and Russell(1998) where the dependent variable is modeled as the time between events. The model proposed is

$$\begin{aligned} x_i &= \mathbf{y}_i \mathbf{e}_i \\ \mathbf{y}_i &= \mathbf{w} + \sum_{j=1}^p \mathbf{a}_j x_{i-j} + \sum_{j=1}^q \mathbf{b}_j \mathbf{y}_{i-j} + \mathbf{g}' z_i \end{aligned} \tag{13}$$

essentially a GARCH(p,q) with exogenous variables for the square root of durations.

---

<sup>1</sup> Although the dependent variable does not need to be signed, the lagged variables in the conditional variance can still include sign information if asymmetric models are sought.



Manganelli(2000) has used multiplicative error models for volume in a transactions model of market microstructure. He estimated models of returns, duration and volume as a trivariate system of equations and then examined impulse responses through this system.

Engle and Gallo(2001) and Chou(2001) estimated models on realized volatility and high low ranges to obtain new more efficient volatility estimators. These models all had this form.

I will now present some illustrative results using realized volatilities of dollar/DM exchange rates from Andersen, Bollerslev, Diebold and Labys(2000). They construct a series of daily variances from squaring and averaging 5 minute returns obtained from Olsen and associates for a period from 12/1986 to 4/1996. For almost 10 years of daily data, we have a return and a “realized variance” and we want to use these to model volatilities.

The data show that the average of the squared returns is close to the average realized variance. The squared return however has a much bigger range and a bigger standard deviation.

TABLE 1.  
Sample: 12/02/1986 4/18/1996

	R <sup>2</sup>	V
Mean	0.4904	0.5287
Median	0.1501	0.3984
Maximum	12.441	5.2454
Minimum	0	0.0518
Std. Dev.	1.0140	0.4837

Estimation of a GARCH(1,1) model with robust standard errors gives the rather conventional results in Table 2.

TABLE 2.  
Dependent Variable: R

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.007	0.003	2.386	0.017
ARCH(1)	0.046	0.010	4.676	0.000
GARCH(1)	0.940	0.012	75.55	0.000

Introducing lagged realized variance,  $v(-1)$  into this equation produces the results in table 3.

TABLE 3.  
Dependent Variable: R

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.0122	0.0069	1.7525	0.0797
ARCH(1)	0.0077	0.0144	0.5322	0.5946
GARCH(1)	0.8698	0.0324	26.8602	0.0000
V(-1)	0.0908	0.0293	3.1006	0.0019

Realized variance does have explanatory power beyond past squared returns to predict squared returns. In fact, lagged squared returns is no longer significant in this model.

Now I will estimate a model for realized volatility which is the square root of realized variance. Again, a GARCH(1,1) will be specified but in this case it should be recognized that this is a multiplicative error model with an exponential error assumption. Standard errors are computed using the Bollerslev-Wooldridge formulation. Two models are estimated; the second also includes past squared returns.

TABLE 4.

Dependent Variable: SQR(V)				
	Coefficient	Std. Error	z-Statistic	Prob.
C	0.0325	0.0075	4.3470	0.0000
ARCH(1)	0.2860	0.0248	11.5519	0.0000
GARCH(1)	0.6538	0.0307	21.2816	0.0000
C	0.0304	0.0066	4.5854	0.0000
ARCH(1)	0.2014	0.0228	8.8291	0.0000
GARCH(1)	0.6912	0.0276	25.0348	0.0000
R(-1) <sup>2</sup>	0.0546	0.0081	6.7344	0.0000

As can be seen the first equation is similar to a conventional GARCH model except that the coefficient on the arch term is much bigger and the coefficient on the garch term much smaller than usual. This process is still highly persistent but is more volatile; it has a higher volatility of volatility. This will be discussed in much detail later in this paper. In the second set of results, the previous squared return is introduced. It has a small coefficient but a very small standard error<sup>2</sup> and is quite significant in relation to the asymptotic normal distribution. In fact, the coefficient is quite similar to that in Table 2. There is apparently information in the squared return that helps to predict realized volatility.

This estimator is potentially inefficient as it assumes an exponential density when another density could be better. In fact the squared standardized residuals which in this case are estimates of the disturbances, have a mean of one but a mode which is close to one as well and a standard deviation which is .75, revealing under dispersion. A plot is given in Figure 1. A more flexible set of density functions for non-negative disturbances is the gamma density, which is a continuous version of a Chi Square. Setting the mean to unity leaves one shape parameter in the density, which is the degrees of freedom/2 in a Chi Square. The results are in Table 5 for both specifications. These densities are strongly preferred over the exponential, which achieved a log likelihood of only -619. A chi square of 7 degrees of freedom divided by 14, which is its mean, has a plot similar to that in Figure 1. The significance of the squared returns is supported again with very similar coefficients and standard errors.

Many further extensions are possible including more flexible densities and time varying densities. Preliminary efforts to estimate time varying gamma densities for this data set indicated that the assumption of *i.i.d.* gamma was quite acceptable. For other data, this may not be the case.

<sup>2</sup> Interestingly, when conventional standard errors are used, the t-statistic is only 1. It is not clear why the robust standard errors are so much smaller.

The potential applications of this set of statistical methods are enormous. As experience grows with various types of data it will become clearer which densities will be most useful and how best to parameterize any changing density. The forecast performance of such models can be investigated with both one step and many step forecasts.

TABLE 5.

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.0325	0.0025	12.7904	0.0000
ARCH(1)	0.2856	0.0129	22.0913	0.0000
GARCH(1)	0.6541	0.0146	44.7447	0.0000
DF/2	3.4768	0.0824	42.2075	0.0000
Log likelihood	132.1059	Akaike info criterion		-0.1046
Avg. log likelihood	0.0539	Schwarz criterion		-0.0951
Number of Coefs.	4	Hannan-Quinn criter.		-0.1012
C	0.0303	0.0022	13.6604	0.0000
ARCH(1)	0.2010	0.0140	14.3499	0.0000
GARCH(1)	0.6918	0.0135	51.3626	0.0000
R(-1)^2	0.0546	0.0071	7.7330	0.0000
DF/2	3.5527	0.0828	42.8839	0.0000
Log likelihood	160.9815	Akaike info criterion		-0.1274
Avg. log likelihood	0.0658	Schwarz criterion		-0.1156
Number of Coefs.	5.0000	Hannan-Quinn criter.		-0.1231

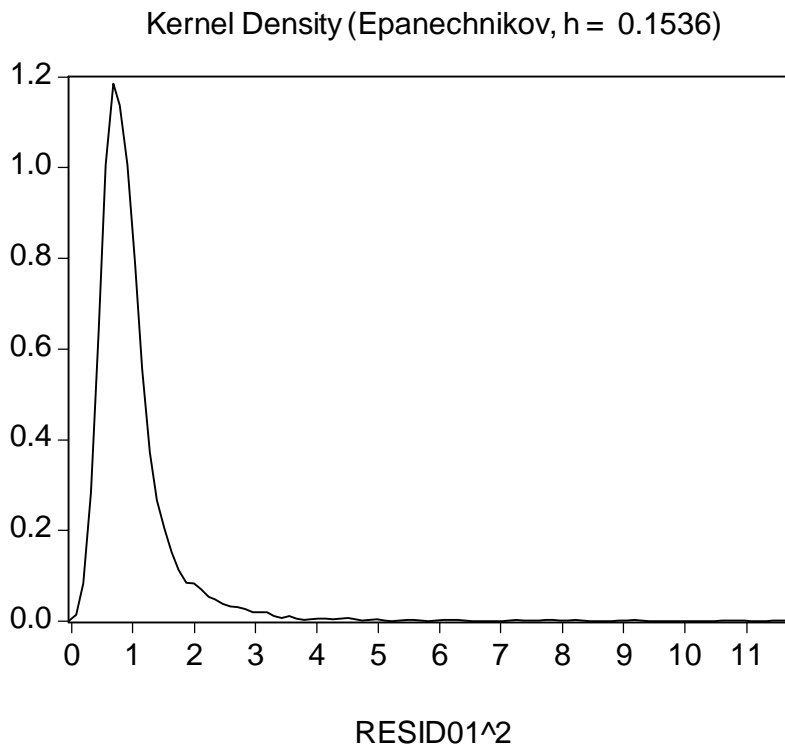


FIGURE 1.

## IV. SIMULATION METHODS FOR CONDITIONAL EXPECTATIONS

### A. LEAST SQUARES MONTE CARLO

A common problem is calculating the expected value of some future outcome given state variables that we know today. This expectation depends upon the stochastic process generating the data but only in the simplest cases are these expectations available analytically. Examples of this problem in time series analysis are the expectation of a random variable many steps in the future when the process is non-linear or the expectation of a non-linear function of a random variable even when the process is linear. In finance, an example might be the payoff of an option as a function of today's underlying price or in a more complex setting, the payoff of an option strategy that could include path dependent decisions, as a function of today's state variables. In the ARCH context, we would like to know the expectation of future volatility as a function of today's information or perhaps the expectation of the log of future volatility. For risk management, we may want to know the expected value of a density function below a threshold or the probability of a joint tail event.

All these problems can be formulated as the following problem.

For a stochastic process

$$\{r_1, \dots, r_{T+M}\} \sim F(\cdot), \quad (14)$$

find

$$E\left(y_T \mid z_{1,t}, \dots, z_{k,t}\right), \quad \text{where } y_T = u(r_{t+1}, r_{t+2}, \dots, r_{t+T}), \quad (15)$$

$$z_{j,t} = v_j(r_t, r_{t-1}, \dots, r_{t-M}), \quad j = 1, \dots, k$$

A potential solution to this problem is a simulation of simulations, which is, unfortunately, unworkable. This consists of simulating the process from time 1 to  $t$  and computing the  $z$ 's. For each of these simulations, a set of future  $r$ 's is simulated and the resulting  $y_T$ 's are averaged to estimate the expectation given the  $z$ 's. Then the expected values of  $y$  given  $z$  are tabulated, or approximated by a generous parameterization or a non-parametric regression. If  $N$  replications are needed, this requires  $N^2$  simulations which is typically too time consuming.

The alternative strategy is Least Squares Monte Carlo as proposed by Longstaff and Schwartz(2001). Their approach requires only  $N$  simulations and is thus computationally far more efficient. They used the method to price American style options by simulation. This has been thought to be impossible because it requires knowing the value of the option before the expiration point – hence the need for simulations of simulations. Their method can be used to find the early exercise frontier for the options and then price these options.

The Least Squares Monte Carlo method estimates a cross sectional regression with each simulation being one observation with the dependent variable  $y_T$ . The

independent variables are a non-parametric or generously parameterized function of the  $z_{j,t}$ 's. The result is a consistent estimate of (15) by the standard properties of ordinary least squares. Since each observation is independent, then the efficiency of this estimation will depend upon the distribution of the errors. If in addition

$$V\left(y_T \mid z_{1,t}, \dots, z_{k,t}\right) = s^2\left(z_{1,t}, \dots, z_{k,t}\right) \quad (16)$$

where  $s$  is a known function, then weighted least squares will be the best linear unbiased estimator of this model.

This solution is closely related to the method of “reprojection” introduced by Gallant and Tauchen(1998) to find properties of simulated series. Their approach however initially constructs a very long time series and thus does not deal easily with overlapping observations.

## B. ESTIMATING THE VOLATILITY OF VOLATILITY

To illustrate the problem we will define and estimate measures of the volatility of volatility that can be estimated for a wide range of conditional volatility models as well as stochastic volatility models and compared with implied volatilities. Consider the volatility of a price process that follows geometric Brownian motion.

$$\begin{aligned} dp &= \mathbf{m}pdt + \mathbf{s}pdw \\ d \log p &= \left(\mathbf{m} - \mathbf{s}^2 / 2\right)dt + \mathbf{s}dw \end{aligned} \quad (17)$$

The volatility is generally measured by the standard deviation of  $d \log p$ . Even in more general diffusion processes, it is natural to measure the volatility as the unconditional standard deviation of  $d \log p$ . Now replace  $p$  with an instantaneous volatility called  $v$ , then the standard deviation of  $d \log v$  would be called the volatility of volatility or  $VoV$ . Similarly,  $v$  could be the realized volatility over a fixed time period, given in discrete time by  $v_T^2 = \left(r_{T-k}^2 + \dots + r_T^2\right)$ . Again  $VoV$  would be the standard deviation of  $d \log v$ .

The same definition can be used for expected volatility. If for example  $v_{T/t}^2$  is the forecast of the variance from time  $T-k$  to  $T$ , based on information at time  $t$ , then the  $VoV$  can be constructed for this volatility letting  $t$  increase while holding  $T$  fixed. Again the definition gives  $VoV$  as the standard deviation of  $d \log v_{T/t}$ , or in discrete time<sup>3</sup>:

$$VoV = \sqrt{V\left(\log v_{T/t} - \log v_{T/t-1}\right)} \approx \sqrt{V\left[E_t\left(\log\left(v_T\right)\right) - E_{t-1}\left(\log\left(v_T\right)\right)\right]} \quad (18)$$

This measure depends on the conditional variances but is calculated unconditionally over all realizations. A conditional version could also be examined.

A particular application of this formula is to implied volatilities, which, under some conditions, can be interpreted as the square root of a forecast of variance from today to the expiration of the option at  $T$ . This measure can be constructed for any time series on implied volatilities; there may be minor issues associated with the rollover from one contract to another, as maturity will not be exactly constant. Such a number can be compared with forecasts from GARCH or other volatility models. However, it is not easy to calculate this value for most volatility models. For example, the VIX implied

<sup>3</sup> This approximation is exact if the Jensen inequality terms cancel. This will be true in a second order taylor expansion if the coefficient of variation is the same at  $t$  and  $t-1$ .

volatility series is approximately the 22-day Black Scholes at-the-money implied volatility of the S&P100 options. The estimated  $VoV$  for the sample period of the 90's as used below, is

$$VoV^{VIX} = stdev\left(\log(VIX / VIX(t-1))\right) = 0.060$$

The standard error of an autoregression in  $\log(VIX)$  is almost the same number.

To apply the least squares monte carlo method to estimation of the  $VoV$  for a process, estimate a cross sectional regression:

$$y_{T,i} = f(r_{t,i}, z_{t-1,i}) + \mathbf{e}_i \quad (19)$$

where

$$y_{T,i} = \log\left(\sum_{j=1}^k r_{n,t+j}^2\right)^{1/2} \quad (20)$$

is the log realized standard deviation of returns in the  $i^{\text{th}}$  simulation. This is the log volatility regression. An alternative is the variance regression when logs are not taken. The regressors are returns at an intermediate time  $t$  and a set of variables based on information prior to time  $t$ . The fitted value is an estimate of the expectation of  $y$  made at time  $t$ . A similar regression can be estimated with information only to time  $t-1$ . Let this be

$$y_{T,i} = g(z_{t-1,i}) + \mathbf{e}_i \quad (21)$$

The difference between  $f$  and  $g$  is the improvement in forecast attributed to information on  $r_t$ . It is the vector of innovations and its standard deviation is the estimated  $VoV$ .

$$VoV = \mathbf{s}_{\hat{f}-\hat{g}} \quad (22)$$

The plot of  $f$  against  $r$  is interpreted as the News Impact Curve of Engle and Ng(1994) for the statistic  $y_T$ . Engle and Ng considered only the one step ahead conditional variance but this can be extended to many statistics. Clearly the NIC will depend in general on the values of  $z$ , and therefore it is common to evaluate this plot at an average value of  $z$  although any values can be used.

The  $VoV$  in (22) is most easily calculated as the standard deviation of the difference between the residuals of (19) and (21). To form an asymptotic confidence interval, let  $u$  be the squared difference between the two residuals. Since each residual has mean zero and observations are independent, applying the central limit theorem and the delta method,

$$N^{1/2}\left(\bar{u}^{1/2} - VoV\right) \xrightarrow{N \rightarrow \infty} N\left(0, \mathbf{s}_u^2 / VoV^2\right) \quad (23)$$

where  $\mathbf{s}_u^2$  is the sample variance of  $u$ .

The regressors assume that the model and its parameters are known but only information up to time  $t$  or  $t-1$  is observed. The choice of regressors for (19) and (21) is simple in only a few special cases. For all types of GARCH models, the innovation in

the variance equation is the difference between the squared return and its conditional variance. The expectation at time  $t-1$  is based just upon the conditional variance at  $t$ . For GARCH, all forecasts of variances are linear combinations of these two variables. In the log volatility equation, it might be expected that logs of conditional variances and innovations will occur. In general, this should be a non-parametric specification. The parametric approximation follows some experimentation. The results are only very slightly altered by a wide range of specifications. The following model was used for all conditional volatility models.

$$\log(\hat{\mathbf{s}}_T) = c + \mathbf{a}_1 r_t^2 + \mathbf{a}_2 |r_t| + \mathbf{a}_3 |r_t|/\sqrt{h_t} + \mathbf{b}_1 h_t + \mathbf{b}_2 \log(h_t) + \mathbf{b}_3 \sqrt{h_t} + \mathbf{e}_i \quad (24)$$

where  $\hat{\mathbf{s}}_T^2 = (r_{t+1}^2 + \dots + r_{t+k}^2)$  which will be positive with probability 1. Typically the most important variables are  $(|r_t|/\sqrt{h_t}, \log(h_t))$ , but for some models, others are also significant.

### C. APPLICATION TO GARCH(1,1\_)

To examine the performance of this methodology as well as to estimate the previously incomputable *VoV*, a series of GARCH models was simulated. The model is defined to be

$$h_t = \mathbf{w} + \mathbf{a}h_{t-1}\mathbf{e}_{t-1}^2 + \mathbf{b}h_{t-1} \quad (25)$$

The design parameters are given in Table 6. There are some models with unusually high alphas and others with very low alpha but high persistence as measured by the sum of alpha and beta.

Horizons of one month (22 days) and one year (252 days) are considered with 100000 simulations. The log volatility equation is estimated for each of the 11 models to compute the volatility of volatility. The volatility of volatility is given in Table 6 along with the standard errors of the estimates from (23).

TABLE 6.  
GARCH VOLATILITY OF VOLATILITY

alpha	beta	Monthly VoV	Std Err	Annual VoV	Std Err
0.400	0.400	0.04977	0.00094	0.00812	0.00054
0.300	0.600	0.06176	0.00073	0.01085	0.00113
0.200	0.700	0.04584	0.00066	0.00553	0.00016
0.100	0.800	0.02619	0.00029	0.00317	0.00004
0.100	0.850	0.03674	0.00046	0.00531	0.00013
0.100	0.870	0.04398	0.00044	0.00990	0.00017
0.050	0.900	0.02010	0.00022	0.00253	0.00004
0.050	0.940	0.02842	0.00030	0.01162	0.00014
0.030	0.960	0.01849	0.00020	0.00791	0.00010
0.020	0.970	0.01219	0.00013	0.00514	0.00007
0.040	0.955	0.02545	0.00025	0.01395	0.00021

It is interesting to note that at the monthly horizon, the values of volatility of volatility are largest for the models with larger  $\alpha$ . However at the annual horizon, the most persistent models generally are the most volatile. The standard errors in all cases are quite small. The daily VoV's of the monthly forecasts from the models we are accustomed to seeing, are well below the .06 for the option data.

#### D. APPLICATION TO OTHER GARCH MODELS

The same approach can be applied to a wide range of GARCH models for which analytic results are not available. In each case, the models have been estimated by maximum likelihood for a decade of S&P500 daily returns ending on August 2, 2001. The GARCH(1,1), Component GARCH and symmetric EGARCH(1,1) are the first three models considered. For these models, most analytical results are known<sup>4</sup>. However for the remaining models, this is not true.

The first new model is the SQGARCH or square root garch model of Engle and Isao(2001). This model parameterizes the volatility of volatility so that the variance of the variance is linear in the variance. It is a discrete time analogue of square root diffusion models used by Cox Ingersoll and Ross(1085), Heston(1993) and many others. The conditional variance equation is given by

$$h_{t+1} = w + \mathbf{b}h_t + \mathbf{a}h_t^{1/2} \left( \frac{r_t^2}{h_t} - 1 \right) \quad (26)$$

It is easy to see that if the exponent on the innovation term is 1, it is simply a GARCH(1,1). The estimated parameters are in Table 7.

<sup>4</sup> The log likelihood for these three models is : -3741.492, -3730.631, and -3737.322 respectively. The symmetric EGARCH is the same as the conventional EGARCH but omits the term  $r_t / \sqrt{h_t}$ .



TABLE 7.  
SQGARCH PARAMETERS

	Coefficient	Std. Error	z-Statistic
<i>omega</i>	0.008874	0.001596	5.560236
<i>alpha</i>	0.041878	0.003685	11.36383
<i>beta</i>	0.990080	0.001990	497.5850
<i>Log likelihood</i>	-3747.891	Akaike info criterion	2.562960

A second model is a non-linear GARCH model that is a hybrid between an integrated model and a mean reverting model. This integrated model has no intercept and therefore will implode eventually as shown by Nelson(1990). However if the variance gets too small, it converts to a mean reverting GARCH in a piecewise linear fashion. Thus it behaves like an integrated GARCH (or Riskmetrics style Exponentially weighted moving average model) for large conditional variances and like a conventional GARCH for small series. The formula is:

$$h_{t+1} = h_t + \mathbf{a}(r_t^2 - h_t) + \mathbf{g}(d - h_t)I_{h < d} \quad (27)$$

where  $I$  is an indicator function. Notice that it is an IGARCH for variances bigger than delta, and mean reverting below delta. The estimated values are given in Table 8.

TABLE 8.  
NON-LINEAR GARCH PARAMETERS

	Coefficient	Std. Error	z-Statistic
<i>alpha</i>	0.054196	0.004749	11.41158
<i>gamma</i>	0.001208	0.001935	0.624299
<i>delta</i>	3.194072	4.471226	0.714362
<i>Log likelihood</i>	-3741.520	Akaike info criterion	2.558606

The daily variance of the S&P is about 1%. The estimated value of delta is not very significant but is considerably above the mean variance indicating that the model is primarily mean reverting except for large volatilities.

Several existing models were also considered such as the power model of Engle and Bollerslev(1986)

$$h_{t+1} = \mathbf{w} + \mathbf{a}r_t^p + \mathbf{b}h_t \quad (28)$$

which has parameter estimates:

TABLE 9.  
ENGLE - BOLLERSLEV POWER MODEL

	coefficient	std. err	z-statistic
<i>omega</i>	0.000587	0.002340	0.251006
<i>alpha</i>	0.067071	0.009086	7.381861
<i>p</i>	1.712818	0.117212	14.61294
<i>beta</i>	0.941132	0.005546	169.7078
<i>Log likelihood</i>	-3739.091		

Notice that  $p$  is significantly below 2 by conventional criteria. Consequently, no analytical formula for multi-step forecast can be found, even if a particular distributional assumption is made on the innovations. This model might appear explosive since alpha and beta sum to more than unity, however that configuration implies explosive behavior only when  $p=2$ .

Ding, Granger and Engle(1993) proposed the Power ARCH or PARCH model. In its symmetric form it is given by

$$h_{t+1}^g = \mathbf{w} + \mathbf{a}r_t^{2g} + \mathbf{b}h_t^g \quad (29)$$

and the estimated parameters are in Table 10. In this case the exponent applies to all the variables and is estimated at close to 1/2 suggesting a standard deviation style model.

TABLE 10  
PARCH MODEL

	coefficient	std. err	z-statistic
<i>omega</i>	0.006680	0.001653	4.041563
<i>alpha</i>	0.064930	0.005608	11.57887
<i>gamma</i>	0.665636	0.082814	8.037719
<i>beta</i>	0.941625	0.005211	180.704
<i>Log likelihood</i>	-3738.040		

A model written entirely in square root form was proposed by Taylor(1986) and used by Schwert(1989). It is defined by

$$\sqrt{h_{t+1}} = \mathbf{w} + \mathbf{a}|r_t| + \mathbf{b}\sqrt{h_t} \quad (30)$$

and has parameter estimates given in Table 11.

TABLE 11  
TAYLOR-SCHWERT MODEL

	coefficient	std. err	z-statistic
<i>omega</i>	0.007678	0.001667	4.605529
<i>alpha</i>	0.065232	0.005212	12.51587
<i>beta</i>	0.942517	0.005104	184.6524
<i>Log likelihood</i>	-3739.032		

Notice that this fits slightly better than the conventional GARCH(1,1) for this data set.

These models can all be compared from the perspective of volatility of volatility since they can all be simulated. Using the same strategy as employed for the GARCH(1,1) models, the following results are found.

TABLE 12  
VOLATILITY OF VOLATILITY WITH GAUSSIAN SHOCKS

	Monthly VoV	Standard Error	Annual VoV	Standard Error
GARCH(1,1)	0.01683	0.00058	0.00244	0.00017
COMPONENT	0.02791	0.00056	0.01578	0.00056
NL-GARCH	0.02570	0.00042	0.02102	0.00187
SQ-GARCH	0.02448	0.00064	0.00944	0.00059
TAYLOR/SCHWERT	0.03318	0.00047	0.01790	0.00055
EB-POWER	0.02765	0.00057	0.01209	0.00067
PARCH	0.03199	0.00051	0.01831	0.00051
EGARCH	0.03191	0.00046	0.01264	0.00054

TABLE 13.  
VOLATILITY OF VOLATILITY WITH NON-GAUSSIAN SHOCKS

	Monthly t(6)	Annual t(6)	Monthly t(4)	Annual t(4)	Monthly Crashes	Annual Crashes
GARCH(1,1)	0.0299	0.0033	0.0316	0.0092	0.0364	0.0068
COMPONENT	0.0427	0.0199	0.0437	0.0239	0.0513	0.0388
NL-GARCH	0.0395	0.0213	0.0413	0.0275	0.0491	0.0447
SQ-GARCH	0.0409	0.0140	0.0477	0.0194	0.0393	0.0176
TAYLOR/SCHWERT	0.0409	0.0169	0.0374	0.0169	0.0472	0.0272
EB-POWER	0.0422	0.0169	0.0451	0.0210	0.0440	0.0192
PARCH	0.0423	0.0180	0.0403	0.0198	0.0494	0.0331
EGARCH	0.0418	0.0163	0.0407	0.0187	0.0465	0.0183

This wide range of GARCH type models have monthly VoV that range from .016 to .033 per day. These would correspond to annualized volatilities from 25 to 50%, although one might not want to annualize these in the conventional way since volatilities are mean reverting. The most volatile models are the Taylor/Schwert, the PARCH and the EGARCH. The annual volatilities are considerably lower ranging from .002 to .02. The largest VoV are the NLGARCH, the PARCH and Taylor/Schwert.

Table 13 calculates the same measures when the simulated data are generated with non-Gaussian innovations. Shocks are generated as  $t(6)$ ,  $t(4)$  and crash, where a crash innovation is standard Gaussian with probability .995 and Gaussian  $(-5,2)$  otherwise. The fatter tailed innovations increase the  $VoV$  but not dramatically. If the implied volatility series VIX is a standard for the monthly volatility forecast, then all of these models are too smooth. The  $VoV$  of the VIX is .06, which exceeds all of these cases. Of course the VIX refers to the S&P100 which is more volatile than the S&P500, and it has variability attributable to time variation in risk premia and mispricing. Either the multi-step forecasts of these models are insufficiently responsive to news, or options prices incorporate important additional sources of volatility.

## E. STOCHASTIC VOLATILITY MODELS

The techniques developed for the conditional volatility models can be extended to any volatility model that can be simulated. In this section, we develop this approach and apply it to several types of stochastic volatility model. Several recent surveys of this material should be examined by the interested reader: The main difference, is that the past information set can no longer be summarized by a conditional variance, since this is no longer defined by the model. Instead, a more general conditioning data set is introduced.

The natural generalization of equation (24) considers simply long lag distributions of absolute and squared returns. Thus the equations that are estimated are:

$$\log(\hat{s}_T) = c + \mathbf{a}_1 |r_t| + \mathbf{a}_2 r_t^2 + \mathbf{a}_3 \sqrt{|r_t|} + \sum_{j=1}^{50} (\mathbf{b}_{1j} |r_{t-j}| + \mathbf{b}_{2j} r_{t-j}^2) \quad (31)$$

and the parameters labeled alpha measure the impact of news.

Three stochastic processes are simulated. The first is a conventional stochastic volatility process, the second is a “breaking volatility” model that has changes in volatilities at random times, and the third is a long memory stochastic volatility model. The generating equation for the SV model is

$$\begin{aligned} r_t &= \mathbf{s}_t \mathbf{e}_t \\ \log(\mathbf{s}_t) &= \mathbf{m} + \mathbf{r} \log(\mathbf{s}_{t-1}) + \mathbf{kn}_t \\ (\mathbf{e}_t, \mathbf{v}_t) &\sim IN(0, I) \end{aligned} \quad (32)$$

as originally proposed by Taylor(1982) and later Nelson(1988). When this model is estimated by Bayesian methods using the approach of Kim, Shephard & Chib (1998), the parameters are given by the middle column of table 14.

TABLE 14  
STOCHASTIC VOLATILITY PARAMETERS

PARAMETER	KSC	JPR
<i>Rho</i>	.987	.97
<i>Kappa</i>	.133	.15
<i>Mu</i>	-.0024	0.0

When the same model was estimated using Shephard's Maximum Likelihood routine, almost identical parameters were obtained. For a different sample period of the S&P, Jaquier, Poulson and Rossi(1994) (JPR) obtained similar parameter values given in the last column.

The "breaking volatility" model is also a type of stochastic volatility model since it models the volatility as a latent variable. In this model, volatility is constant until a random event occurs. At this time a new volatility is drawn from a fixed distribution. The process can be defined as:

$$\begin{aligned}
 r_t &= s_t e_t \\
 s_t^2 &= \begin{cases} s_{t-1}^2 & \text{with probability } p \\ \exp(kn_t + m) & \text{otherwise} \end{cases} \\
 (e_t, n_t) &\sim IN(0, I)
 \end{aligned}
 \tag{33}$$

Two sets of parameters were used for this process based on a very casual matching of moments.

TABLE 15  
BREAKING VOLATILITY PARAMETERS

PARAMETER	BREAKING VOL I	BREAKING VOL II
<i>p</i>	.99	.999
<i>Kappa</i>	1	1
<i>Mu</i>	-.5	-.5

The third process is the Long Memory Stochastic Volatility model as proposed by Breidt, Crato and de Lima(1998) and studied by Hurvich and Deo(2001).

$$\begin{aligned}
r_t &= \mathbf{b} \exp(\sqrt{u_t}) \mathbf{e}_t \\
(1 - B)^d u_t &= \mathbf{k} n_t \\
(\mathbf{e}_t, n_t) &\sim IN(0, I)
\end{aligned}
\tag{34}$$

Parameter estimates used for this model are:

TABLE 16  
LONG MEMORY STOCHASTIC VOLATILITY PARAMETERS

PARAMETER	VALUE
<i>D</i>	.47
<i>Kappa</i>	.6
<i>Beta</i>	.4

18. The results from estimating the *VoV* for these five models are tabulated in table

TABLE 17  
VOV FOR STOCHASTIC VOLATILITY MODELS

	Monthly <i>VoV</i>	Std Err	Annual <i>VoV</i>	Std Err
Breaking Vol I	0.05300	0.00091	0.02054	0.00053
Breaking Vol II	0.02559	0.00093	0.02356	0.00035
Stochastic Vol KSC	0.04315	0.00067	0.01535	0.00045
Stochastic Vol JPR	0.04322	0.00082	0.00818	0.00037
Long Memory SV	0.04797	0.00085	0.01743	0.00033

From these results it appears that all these models have moderately high volatility of volatility for long forecasts. Presumably this is because the persistence is governed by a different parameter from the volatility of volatility. A related point has been made by Carnero, Pena and Ruiz(2001) in the context of the stochastic volatility model. The difficulty of estimation and forecasting for these models has made it impossible to observe this property before now. The long memory models have the property that the long maturity *VoV* is high and rather closer to that of the shorter horizon. The breaking volatility models are not systematically calibrated to the data and therefore it is not clear which is to be considered most representative of this class of models. However, the more frequently breaking model does have the highest *VoV* of all the Gaussian models at the monthly horizon.

The use of the least squares Monte Carlo method can illuminate these and many other features of time series models. In fact the applications of this general structure seem unlimited. The forecast equation for these stochastic volatility models can be estimated, the expected payoff or utility from path dependent hedging strategies can be assessed, the structure of derivative prices can be analyzed, and many other tasks can be

rather simply evaluated, reducing the need to have analytically simple models in order to obtain closed form forecasting results.

## V CONCLUSIONS

ARCH models have come a long way from 1982 to the present. In some respects their very success has made it less interesting to continue research on volatility models. There is a sense in which this is a known body of econometrics. In other ways, new horizons are opening all the time. The five frontiers discussed in this paper are all fresh new areas of research with important applications and lots of room for both theoretical and empirical developments. These five areas – high frequency volatility, high dimension correlation, derivative pricing, modeling non-negative processes, and analyzing conditional simulations by Least Squares Monte Carlo – could well occupy our research efforts for another decade, and would pay handsome dividends in terms of useful and reliable research and practitioner tools. But probably, even brighter frontiers lie ahead and we can only await them with anticipation.

## References

- Alexander, Carol,(1998), *Risk Management and Analysis*. New York : Wiley,
- Andersen T G, Bollerslev T, Diebold F X and Labys P (2001), “The distribution of exchange rate volatility”, *Journal of the American Statistical Association*, 96 42–55
- Andersen, T. G. and T. Bollerslev(1998), “Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts,” *International Economic Review*, 39,4, 885-905
- Andersen, T. G. and T. Bollerslev(1997), “Intraday Periodicity and Volatility Persistence in Financial Markets,” *Journal of Empirical Finance*, 4:2, 115-158
- Baillie R T, Bollerslev T and Mikkelsen H O (1996), “Fractionally integrated generalized autoregressive conditional heteroskedasticity”, *Journal of Econometrics*, 74 3–30
- Bera, Anil and M.L. Higgins(1993), “ARCH Models: Properties, Estimation and Testing”, *Journal of Economic Surveys*,7, 305-366
- Bollerslev, Tim,(1990), “Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach”, *Review of Economics and Statistics*, 72, 498-505
- Bollerslev, Tim, 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- Bollerslev, Tim and Wooldridge, Jeffrey M., 1992, Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances, *Econometric Reviews*, 11(2), 143-172.
- Bollerslev, Tim(1987),”A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return”, *Review of Economics and Statistics*, 69,542-547
- Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988), "A Capital Asset Pricing Model With Time-Varying Covariance," *Journal of Political Economy*, Vol. 96, No. 1, 116-131.
- Bollerslev, T. , R.Y. Chou and K.F. Kroner(1992),”ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence”, *Journal of Econometrics*, 52, 5-59
- Bollerslev, T. R.F. Engle and D. B. Nelson,(1994) “ARCH Model”, Chapter 49 of Handbook of Econometrics Volume IV, (ed. Engle and McFadden), Elsevier Science, pp2961-3031



- Breidt, F.J., N. Crato and P. deLima(1998) "The detection and estimation of long memory in stochastic volatility", *Journal of Econometrics*, 83, pp325-348
- Cox, J., J. Ingersoll, and S. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-408.
- Carnero, M. Angeles, Daniel Pena and Esther Ruiz,(2001), "Is Stochastic Volatility more flexible than GARCH?" manuscript Department of Statistics and Econometrics Universidad Carlos III de Madrid
- Christoffersen,-Peter-F.; Diebold,-Francis-X. (2000),How Relevant Is Volatility Forecasting for Financial Risk Management?". *Review-of-Economics-and-Statistics*; 82(1), pages 12-22..
- Christoffersen,-Peter; Hahn,-Jinyong; Inoue,-Atsushi, (2001), "Testing and Comparing Value-at-Risk Measures", *Journal-of-Empirical-Finance*; 8(3), July, pages 325-42.
- Chou, Ray Y.(2001)," Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model", manuscript Academia Sinica
- De Santis, Giorgio, Bruno Gerard and Pierre Hillion,(1997) "How Big is the Premium for Currency Risk", *Journal of Financial Economics*, 49(3), pp 375-412
- Deo, Rohit S. and C. M. Hurvich,(2000) "Estimation of Long Memory in Volatility", manuscript New York University
- Ding, Zhuanxin, C. W. J. Granger and R. F. Engle (1993), "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83-106.
- Duan, Jin-Chuan, (1995), The GARCH option pricing model, *Mathematical Finance* 5, 13-32.
- Duan Jin-Chuan, Augmented GARCH(p,q) process and its diffusion limit, *J. Econometrics* (79)1 (1997) pp. 97-127
- Durbin, Jim and Siem Jan Koopman,(1997)"Monte Carlo maximum likelihood estimation of non-Gaussian state space model", *Biometrika*, Vol 84,pp669-84.
- Engel, Charles; Jeffrey A. Frankel; Kenneth A. Froot; and, Anthony P. Rodrigues, (1995), "Tests of Conditional Mean-Variance Efficiency of the U.S. Stock Market," **Journal of Empirical Finance** 2, March 1995, pages 3-18.

- Engle, Robert F., 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50(4), 987-1007.
- Engle, Robert F., (2000), "The Econometrics of Ultra High Frequency Data", *Econometrica*, pp 1-22
- Engle, Robert,(2002), "Dynamic Conditional Correlation: A New Simple Class of Multivariate GARCH Models", forthcoming *JBES*
- Engle, Robert and Tim Bollerslev,(1986) "Modeling the Persistence of Conditional Variances," *Econometric Reviews* 5 (1986): 1-50.
- Engle, Robert and Isao Ishida, (2001) "Modeling Variance of Variance: The Square-Root, the Afine, and the CEV GARCH Models", manuscript UCSD
- Engle, Robert and Chaudhury Mustafa, (1992) "Implied ARCH Models from Options Prices," *Journal of Econometrics*, 52, 289-311.
- Engle, Robert F., and Ng, Victor, 1993, Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48, 1749-1778.
- Engle, Robert, Alex Kane and Jaesun Noh (1994) "Forecasting Volatility and Option Prices of the S&P 500 Index," *Journal of Derivatives*, 217-30.
- Engle, R., D. Lilien, and R. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55,391-407.
- Engle, Robert and Joshua Rosenberg(2000), "Testing the Volatility Term Structure Using Option Hedging Criteria", *Journal of Derivatives*,8,10-28
- Engle, Robert and Jeffrey R. Russell,(1998) "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica* 66: 1127-1162.
- Engle, Robert and Kevin Shephard,(2001) "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH", UCSD Discussion Paper 2001-15, NYU Discussion Paper 2001.
- Engle,Robert (2001) "GARCH 101: An Introduction to the Use of ARCH/GARCH models in Applied Econometrics", forthcoming *Journal of Economic Perspectives*
- Engle, Robert and Andrew Patton(2001), "What Good is a Volatility Model?" *Quantitative Finance*, pp 237-245
- Engle, Robert and Giampiero Gallo(2002), "A Model for Intra-Daily Volatility with Multiple Indicators" manuscript

- Epps, Thomas W.(1979) “Comovements in Stock Prices in the Very Short Run” *Journal-of-the-American-Statistical-Association*; 74(366), pages 291-98..
- Figlewski, Steve, (1995) “Conference Presentation at UCSD”
- French K. R., G. W. Schwert and R.F. Stambaugh(1987) “Expected Stock Returns and Volatility”, *Journal of Financial Economics*, 19, 3-30
- Gallant, A.R., and G. E. Tauchen,(1998) “Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions,” *Journal of the American Statistical Association*, **93**, pp10-24
- Ghose, Devajyoti, and Kenneth F. Kroner,(1995),”Components of Volatility in Foreign Exchange Markets: An Empirical Analysis of High Frequency Data”, manuscript
- Glosten, L.R., R. Jaganathan and D. Runkle(1993) “On the Relationi between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”, *Journal of Finance*, 48, 1779-1801
- Poon, S.H. and C.W.J.Granger,(2001) “Forecasting Financial Market Volatility: A Review”, manuscript, Department of Economics, UCSD
- Heston, S., (1993), “A closed form solution for options with stochastic volatility with applications bond and currency options”, *Review of Financial Studies* 6, 327-343.
- Higgins, M. L. and A. K. Bera (1992), A Class of Nonlinear Arch Models *International Economic Review*, Vol. 33, No. 1,pp. 137-158.
- Hull, J., and A. White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42,281-300.
- Kim Sangjoon, Neil Shephard and Siddhartha Chib,(1998) "Stochastic Volatility: likelihood inference and comparison with ARCH models", *Review of Economic Studies*, 65,pp361--393
- Jacquier, E., N.G. Polson and P.E. Rossi(1994) “Bayesian Analysis of Stochastic Volatility Models” *Journal of Business and Economic Statistics*, 12,371-417
- Lee, S., and B. Hansen, (1994), Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10,29-52.
- Ledoit, Olivier, and Pedro Santa -Clara(1998) “Estimating large conditional covariance matrices with an application to international stock markets”, Manuscript, Andersen Graduate School, Department of Finance, UCLA

- Lee, S., and B. Hansen, (1994), Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory* 10, 29-52
- Lumsdaine, R., (1996), Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, *Econometrica* 64, 575-596.
- Ling, Shiqing and Michael McAleer (2002a) “Necessary and Sufficient Moment Conditions for the GARCH(r,s) and Asymmetric Power GARCH(r,s) Models”, *Econometric Theory*, 18,722-729
- Ling, Shiqing and Michael McAleer (2002b) “Stationarity and the existence of moments of a family of GARCH processes”, *Journal of Econometrics*, 106,109-117
- Manganelli, Simone(2000) “Volume, Volatility and the Price Impact of Trades”, PhD Dissertation, UCSD.
- McCurdy, Tom and T. Stengos(1992) “A Comparison of Risk Premium Forecasts Implied by Parametric and Nonparametric conditional Mean Estimators,” *Journal of Econometrics*,52, 225-244
- Nelson, Daniel B. (1988) “Time Series Behavior of Stock Market Volatility and Returns,” PhD thesis, MIT.
- Nelson, Daniel B., 1991, Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica*, 59(2), 347-370.
- Nelson, Daniel (1990) “Stationarity and Persistence in the GARCH(1,1) Model”, *Econometric Theory*, 6, 318-334
- Nelson,-Daniel-B.; Cao,-Charles-Q. (1992), “Inequality Constraints in the Univariate GARCH Model” *Journal-of-Business-and-Economic-Statistics*; 10(2), pages 229-35.
- Rosenberg, Joshua and Robert Engle,(2001) “Empirical Pricing Kernels”, *Journal of Financial Economics*, forthcoming
- Sandmann, G and S J Koopman, (1998) "Estimation of stochastic volatility models via Monte Carlo maximum likelihood", *Journal of Econometrics*, 87, pp271--301
- Schwert G W 1989 Why does stock market volatility change over time? *J. Finance* 44 1115–53
- Scholes, Myron; Joseph T. Williams,(1977) “ Estimating Betas from Nonsynchronous Data” *Journal-of-Financial-Economics*; 5(3), pages 309-27..

- Shephard, Neil and Michael K Pitt,(1997),"Likelihood analysis of non-Gaussian measurement time series", *Biometrika*, Vol 84,pp653-67
- Styan, G. P. H. (1973), "Hadamard products and multivariate statistical analysis," *Linear Algebra and Its Applications*, 6, 217-240
- Taylor, Steven J. (1982), "Financial Returns Modelled by the Product of Two Stochastic Processes- A Study of Daily Sugar Prices, 1961-1979", in Time Series Analysis: Theory and Practice, (ed. O.E.Anderson), North Holland, pp203-226
- Taylor, Stephen J, (1986), Modelling Financial Time Series, John Wiley and Sons Ltd.
- Ritchken, Peter and Trevor, Rob, (1999) "Pricing Options under Generalized GARCH and Stochastic Volatility Processes", *Journal-of-Finance*; 54(1), pages 377-402..
- R. Rabemananjara, J. M. Zakoian Threshold Arch Models and Asymmetries in Volatility,1993, *Journal of Applied Econometrics*, Vol. 8, No. 1. (Jan. - Mar., 1993), pp. 31-49
- White, Halbert, (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*; 48(4), May, pages 817-38.