CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Small Sample Properties of GARCH Estimates and Persistence

Soosung Hwang and Pedro L. Valls Pereira

CEA@Cass Working Paper Series

WP–CEA–10-2004

Small Sample Properties of GARCH Estimates and Persistence

Soosung Hwang[1] Pedro L. Valls Pereira[2]

Cass Business School, London Ibmec Business School, Brazil

October 2004

[1]Corresponding Author. Faculty of Finance, Cass Business School, 106 Bunhill Row, London, EC1Y 8TZ, UK. Tel: 44 (0)20 7040 0109. Fax: 44 (0)20 7040 8881. Email: s.hwang@city.ac.uk. We would like to thank Gordon Gemmill, Mark Salmon, Stephen E. Satchell, and the participants of the 2003 Forecasting Financial Markets Conference for their helpful comments. The second author acknowledges the financial support from CNPq, Ibmec Business School and PRONEX. All remaining errors are our responsibilities.

[2]Address: Ibmec Business School, Rua Maestro Cardim 1170, 01323-001 São Paulo, S.P., BRAZIL, Tel: 55 11 3175 2300, Fax: 55 11 3175 2315, Email: pvalls@ibmec.br

Small Sample Properties of GARCH Estimates and Persistence

Abstract

We show that the ML estimates of the popular GARCH(1,1) model are significantly negatively biased in small samples and that in many cases converged estimates are not possible with Bollerslev's non-negativity conditions. Our results also indicate that a high level of persistence in GARCH(1,1) models obtained using a large number of observations has a correlogram far less than these ML estimates suggest in small samples. Considering the size of biases and convergence errors, we propose at least 250 observations for ARCH(1) models and 500 observations for GARCH(1,1) models. We also propose a simple measure of how much GARCH conditional volatility explains squared returns. The measure indicates that for a typical index return volatility whose ARCH parameter is very small, the conditional volatility hardly explains squared returns.

Keywords: Small Sample, Volatility, GARCH, Persistence.

JEL Codes: C13

1 Introduction

Volatility is defined as the degree of unexpected price movement over time. However, it is not observable and the choice of volatility measure is controversial. There are many proxy volatilities such as absolute returns, squared returns, stochastic volatility, intra-day volatility and range volatility, but GARCH volatility is by far the most popular. In fact, GARCH models are so popular that many research papers in finance and economics apply them to observation sets of just a few hundreds without noticing the small sample problems of the estimates. For example, in small samples the correlogram of the squared returns in GARCH models is negatively biased and the maximum likelihood (ML) estimates of the ARCH and GARCH parameters are also downward biased.

One might guess that if 10 years' daily data (around 2500 observations) are appropriate to estimate a GARCH model, using a longer time series but lower frequency data such as 20 years' monthly data (240 observations) would not create any serious estimation problem. However, the GARCH estimates from low frequency data suffer from the temporal aggregation problem.[1] Even if this problem is solved, using the 240 monthly observations could still matter. The estimates obtained with a couple of hundred observations may not be reliable regardless of data frequencies because of the small number of observations. Merton (1980) shows that to obtain a better estimate of volatility one should use a large number of higher frequency data rather than a small number of low frequency data over a longer time period.

We investigate small sample properties of the ML estimates of the ARCH and GARCH models in terms of Bollerslev's (1986) non-negativity conditions.[2] More specifically, we show the effects of Bollerslev's (1986) non-negativity conditions on the ML estimates of the ARCH and GARCH parameters and how the conditions are related to small sample biases.

Our results indicate that Bollerslev's non-negativity conditions are serious restrictions in small samples and for GARCH(1,1) models rather than ARCH(1) models. In small samples the conditions cause a huge number of convergence errors, resulting in positive biases in ARCH and GARCH estimates. However, when we use much weaker non-negativity conditions the number of convergence errors decreases significantly but we find large negative biases which are consistent with severe negative biases in sample autocorrelations of squared returns. Considering the biases in the ML estimates and the convergence errors, we suggest that at least 250 observations are needed for ARCH(1) models and 500 observations are needed for GARCH(1,1) models.

Our results also show that the extreme persistence in the GARCH(1,1) model estimated for the large number of S&P500 daily index returns has a correlogram far less than the estimates suggest in small samples, and that this negative bias decreases very slowly with sample size. That is, the high persistence in GARCH models does not automatically imply that estimated autocorrelation coefficients of squared returns will show a similar level of persistence in small samples. The degree of persistence in squared returns could be overstated by GARCH estimates, which increase with the number of observations.

We propose a measure which shows how much GARCH(1,1) models explain squared returns. Our measure, the relative size of ARCH to GARCH parameters, indicates that extreme persistence with a large GARCH parameter which is common in a typical daily index volatility, does not explain squared returns very well and may not be economically meaningful. We suggest that the economically meaningful persistence in the GARCH(1,1) model should be re-investigated not only with the sum of ARCH and GARCH parameters but also with the relative size of ARCH parameter to GARCH parameter. For an extremely persistent GARCH process, it is the ARCH parameter that explains how much the process can explain squared returns.

In the next section, we investigate small sample properties of the ML estimates of ARCH(1) and GARCH(1,1) models using Monte Carlo simulations and the S&P500 index returns. Then, in section 3, we further investigate the economic meaning of persistence in GARCH(1,1) models. Finally, conclusions follow in section 4.

2 Small Sample Properties of the ML Estimates of ARCH(1) and GARCH(1,1) Models

We start our investigation of the small sample properties of the ML estimates in ARCH and GARCH models using Monte Carlo simulations. We show how ML estimates of ARCH(1) and GARCH(1,1) models are affected by Bollerslev's (1986) strict non-negativity conditions in small samples and how severely the sample autocorrelations of ARCH(1) and GARCH(1,1) models are biased. We then show that most of the small sample properties of the ML estimates for the S&P500 index volatility are explained by the simulation results.

The ARCH(1) model is

$$y_t = \xi_t h_t^{\frac{1}{2}}, \quad (1)$$

$$h_t = w + \alpha y_{t-1}^2 \quad (2)$$

where $\xi_t \sim N(0,1)$ and the GARCH(1,1) model is

$$h_t = w + \beta h_{t-1} + \alpha y_{t-1}^2, \quad (3)$$

where Bollerslev (1986) proposes $w > 0$, $\beta \geq 0$ and $\alpha \geq 0$ to guarantee positive conditional volatility. In our study, in order to allow for richer autocorrelation structures of y_t^2, we use non-negativity conditions weaker than those of Bollerslev (1986); we impose $w > 0$ and $h_t > 0$ rather than $w > 0$, $\beta \geq 0$ and $\alpha \geq 0$. Nelson and Cao (1992) argue that Bollerslev's non-negativity conditions are too restrictive and in some cases negative estimates may be obtained in practice. He and Teräsvirta (1999) further show that allowing negative parameters in GARCH models can give us various autocorrelation structures of y_t^2.

For the weak non-negativity conditions, we need different sets of starting values for maximum likelihood (ML) estimation because of the possible multimodality. Doornik and Ooms (2000) provide some examples that may have multiple maxima when sample sizes are small, and both p and q of GARCH models are more than 1. When there are outliers or structural breaks in samples, the likelihood function may become multimodal under the weak non-negativity conditions. In our study, four different sets of starting values were used for GARCH(1,1) models with the weak non-negativity conditions; the four sets of starting values for $\{\beta,\ \alpha\}$ are $\{0.7,\ 0.05\}$, $\{0.3,\ 0.05\}$, $\{-0.3,\ 0.05\}$, and $\{-0.7,\ 0.05\}$. We found the evidence of local maxima in many cases with the weak non-negativity conditions.[3]

2.1 Monte Carlo Simulations for the Small Sample Properties of the ML Estimates

We generate ARCH and GARCH errors as in equations (1), (2) and (3).[4] ARCH(1) and GARCH(1,1) errors of the sizes of 100, 250, 500, and 1000 are generated and the first 1000 observations are not used to avoid any problems from initial values. For each of the cases considered in this study, 1000 replications are used.

Sample autocorrelations for the squared ARCH(1) and GARCH(1,1) errors are calculated and reported with the theoretical autocorrelation in Bollerslev (1988). For squared GARCH(1,1) errors, the thoretical autocorrelations with lag τ, $\rho(\tau)$, are

$$\begin{aligned} \rho(1) &= \alpha\left(\frac{1-\alpha\beta-\beta^2}{1-2\alpha\beta-\beta^2}\right), \qquad (4) \\ \rho(\tau) &= \rho(1)\left(\alpha+\beta\right)^{\tau-1}, \end{aligned}$$

where α and β are defined in equations (1) to (3). For ARCH(1) errors, the thoretical autocorrelations can be obtained by setting $\beta = 0$ in the above equations.

2.1.1 ARCH(1) Process

The first three cases in tables 1A and 1B report the small sample properties of ARCH estimates. Cases 1 and 2 in table 1A show that when Bollerslev's non-negativity conditions are imposed for small ARCH parameters, the ARCH estimates are positively biased. This is because when α is small, there are many negative $\widehat{\alpha}$, which are eliminated by the strict non-negativity conditions. To see this, when we compare cases 1 and 2 in table 1A with those in table 1B, we find that most convergence errors in table 1A come from imposing the strict non-negativity constraints which do not allow negative ARCH estimates or negative autocorrelations in squared returns.[5] Therefore, when α is small, we find a positive bias with the strict non-negativity constraints by removing negative values of $\widehat{\alpha}$.

On the other hand, for the weak non-negativity conditions (cases 1 and 2 in table 1B) or a large α (case 3 in tables 1A and 1B), we have fewer convergence errors. In these cases we have negative biases in $\widehat{\alpha}$. The bias becomes larger for smaller samples.

<Tables 1A and 1B Here>

The negative bias can also be seen in the sample autocorrelations on the right of the tables. Compared with the theoretical autocorrelations, the sample autocorrelations of squared ARCH(1) errors are severely negatively biased, which is consistent with what Bollerslev (1988) found. The sample autocorrelations approach the true values slowly as sample size increases. This is clear for a large ARCH parameter; compare the sample autocorrelations of case 3 with those of the other cases.

Finally, the results in tables 1A and 1B suggest that the minimum sample size required for reasonable estimates depends on sample autocorrelations of squared returns. When autocorrelations are close to zero as in case 1, a large sample is needed for the ARCH(1) model, while for more persistent squared returns such as case 3, we only need 250 or 500. Considering convergence errors and biases in the estimates, we suggest that sample sizes should be at least larger than 250 for ARCH(1) models when we use the conventional Bollerslev non-negativity conditions.

2.1.2 GARCH(1,1) Process

The simulation results for the GARCH(1,1) model in tables 2A and 2B are similar to those for the ARCH(1) model. That is, the ARCH estimates are negatively biased with the weak non-negativity conditions in small samples, while they are positively biased with Bollerslev's strict non-negativity conditions for small values of α. Interestingly, the estimates of $\widehat{w}$ are all biased positively in small samples and approach the true value slowly with sample sizes.[6]

<Tables 2A and 2B Here>

Our concern here is the small sample properties of $\widehat{\beta}$. The first three cases in table 2B show that when sample sizes are small (e.g., 100) and we use the weak non-negativity conditions, $\widehat{\beta}$ becomes seriously negatively biased; in fact, most $\widehat{\beta}$s are negative for a small α. When we impose the strict non-negativity conditions, the negative bias of

$\widehat{\beta}$s becomes smaller but convergence errors increase significantly. The bias decreases slowly as sample sizes increase (see cases 2 and 3 in table 2A). Cases 2 and 3 also suggest that $\widehat{\beta}$ is affected by the size of α. When α is large, $\widehat{\beta}$ becomes less biased.

The negative bias is reflected in the sample autocorrelations in the last six columns of the table. One extreme example is the first case in table 2A where the theoretical autocorrelations are around 0.8 and highly persistent while the sample autocorrelations are far less than the true values. The comparison of the theoretical and simulation autocorrelations shows that squared returns are not as persistent as suggested by GARCH models.

When sample autocorrelations are close to zero we have serious problems in convergence (see cases 2 and 4 in table 2A). When $\beta = 0.6$, $\alpha = 0.1$ and the sample size is 100, 83.1% of cases suffer convergence errors with the Bollerslev non-negativity conditions. These results are far more serious than the ARCH(1) model. In the ARCH(1) model, even when $\alpha = 0$, we get converged estimates for more than 40% for the sample size of 100 (see case 1 of table 1A). The results suggest that when persistence level is low, we need a fairly large sample for GARCH models, probably more than 1000 in order to avoid the problem. Our study indicates that with Bollerslev's strict non-negativity conditions, at least 500 observations are required in GARCH(1,1) models to obtain estimates fairly close to the true parameters as well as to reduce convergence errors.

2.2 Small Sample ML Estimates for Index Returns

We estimate ARCH(1) and GARCH(1,1) models for the S&P500 index daily log-returns to investigate whether or not the above simulation results represent reality.[7] As in tables 1 and 2, the four different sets of starting values are used for the possible multimodality for GARCH(1,1) models. A total of two thousand daily log-returns from 22 September 1994 to 30 August 2002 is obtained from the Datastream, and for the mean-zero return series, y_t, obtained by taking out the sample mean of the log-returns, we estimate the ARCH(1) and GARCH(1,1) models for the entire set of 2000 observations. The same procedure is repeated for 100 observations (20 sub-periods), 250 observations (8 sub-periods), 500 observations (4 sub-periods) and 1000 observations (2 sub-periods) to investigate whether different sample sizes affect the estimates.

Tables 3 and 4 report the ML estimates of the ARCH(1) and GARCH(1,1) models with Bollerslev's non-negativity conditions and the weak non-negativity conditions, respectively. As in simulations, Bollerslev's non-negativity conditions cause many convergence errors in small samples, especially when sample sizes are 100 and 250, while we have fewer convergence errors with the weak non-negativity conditions. The number of convergence errors is more serious for the GARCH model than for the ARCH; there are 55% and 25% convergence errors for sample sizes of 100 and 200 in the GARCH model, respectively, while in the ARCH model, these are 45% and 12.5%, respectively.

<Table 3 Here>

We next investigate how the estimates change for different sample sizes. Panel

A of table 3 shows that the ML estimates of the ARCH(1) model obtained with the 2000 observations are $\widehat{w} = 1.042$ and $\widehat{\alpha} = 0.232$. The ML estimates for the various sub-periods do not seem to support a similar persistence level as the one we obtain from the entire sample. Many sub-periods' $\widehat{\alpha}$s are far less than 0.232. For example, when we divide the entire sample into the four sub-periods, only one sub-period's $\widehat{\alpha}$ is larger than 0.232. Thus, the average values of $\widehat{\alpha}$s are less than 0.232, especially in small samples.

The ML estimates of the GARCH(1,1) model with Bollerslev's non-negativity conditions (panel B of table 3) show that the value of $\widehat{\alpha} + \widehat{\beta}$ for the full sample is 0.993, which is highly persistent and a typical estimate for an index volatility. As in the ARCH(1) model, the ML estimates of the GARCH model show a large variation in small samples, and the extreme persistence in the entire sample period could not be observed in sub-periods. As in simulations, $\widehat{\beta}$ tends to increase with the sample size, while $\widehat{w}$ and $\widehat{\alpha}$ tend to decrease slightly with the sample size. Therefore, the persistence level measured by $\widehat{\alpha} + \widehat{\beta}$ tends to increase as sample size increases.

These patterns, i.e., less persistence in estimates of the ARCH(1) and GARCH(1,1) models in small samples, become clearer in table 4 where the weak non-negativity conditions are used to explain richer autocorrelation structures of squared returns. Panel A of table 4 shows that many ARCH estimates are negative in small samples. Therefore, we have smaller average values of $\widehat{\alpha}$s for the 100 and 250 sample sizes with the weak non-negativity conditions. The GARCH(1,1) estimates with the weak non-negativity conditions (panel B of table 4) report that for the sample size of 100, the mean value of $\widehat{\beta}$s is far from the estimate we obtained with the entire sample (i.e., 0.914) and actually it is close to zero, i.e., 0.01! When the number of observations is increased to 250, 500, and 1000, the mean values of $\widehat{\beta}$s become 0.459, 0.864, and 0.882, respectively. On the other hand, the average values of $\widehat{\alpha}$s show little change for different sample sizes.

<Table 4 Here>

In order to explain the empirical results in tables 3 and 4, we repeat the Monte Carlo simulations above under the assumption that the true parameter values of the ARCH(1) and GARCH(1,1) models are similar to those obtained with the entire sample of the S&P500 index in tables 3 and 4. Case 4 in tables 1 and 2 reports these simulation results. We find that the empirical results in tables 3 and 4 are consistent with the simulation results. That is, large negative biases values in $\widehat{\beta}$ with the weak non-negativity conditions and with Bollerslev's strict non-negativity conditions and small positive bias in $\widehat{\alpha}$ with the strict non-negativity conditions may be interpreted as small sample properties of the ML estimates of the GARCH(1,1) model. The extreme persistence in the GARCH model for the entire sample originates in the large estimate of β which depends on the large sample size.

Another property of the estimates of β is that they change significantly over time in small samples. This could be explained by a few 'major' events in the S&P500 index, which have a major influence on the index volatility. However, we also find similar changes in the above simulations (not reported) where the standard normal distribution is used for the standardised residuals. Therefore $\widehat{\beta}$ is highly sensitive to small changes in squared returns when sample sizes are small. When sample sizes increase, the variation decreases, especially when sample sizes are equal to or larger than 500.

3 Economic Meaning of Persistence in GARCH Models

The results in tables 1 to 4 raise a few questions. First, what is the meaning of persistence in GARCH(1,1) models? If estimates do not show persistence in small samples (for example, in table 2B, case 4, see the sample size of 100), is the extreme persistence estimated with a large sample economically meaningful? We also ask how much GARCH(1,1) volatility explains squared returns.

The extreme persistence of volatility in the full sample estimated with the GARCH(1,1) model, i.e., $\widehat{\alpha}+\widehat{\beta}=0.993$, is far from the persistence found with ARCH(1) model, i.e., $\widehat{\alpha}=0.232$. Those who believe that GARCH conditional volatility is close to the true latent volatility may argue that ARCH models are inappropriate because they do not explain the extreme persistence of volatility. If GARCH models represent the true volatility, then what are the properties of the extreme persistence, especially when β is very large and α is small?

To answer this question, we use ARMA presentation of GARCH models, since it is not easy to interpret the persistence of innovations in GARCH models.[8] Note that the

GARCH(1,1) process in (3) becomes the following ARMA(1,1) process

$$\begin{aligned} y_t^2 &= \omega + (\alpha+\beta)y_{t-1}^2 + (y_t^2 - h_t) - \beta(y_{t-1}^2 - h_{t-1}) \\ &= \omega + \phi y_{t-1}^2 + v_t - \theta v_{t-1} \end{aligned} \tag{5}$$

where $v_t = y_t^2 - h_t$, $\phi = \alpha + \beta$, and $\theta = \beta$.

Schwert (1989) shows that the finite sample ARMA(1,1) series with $\phi = 1$ becomes indiscernible from white noise as the MA coefficient increases to 1.[9] In general, when $\phi \approx \theta$, the ARMA(1,1) process is not different from white noise because of the common factor in the two polynomials. Note that when ϕ is larger than θ, the process has positive autocorrelations. Thus for y_t^2 to have a small but highly persistent positive autocorrelation, ϕ should be close to 1, but θ should be large and close to ϕ. In the GARCH(1,1) model, when β is far larger than α, the autocorrrelation of y_t^2 is positive. However the persistence that the sum of α and β suggests may not have economic meaning, because as $\alpha \to 0$, $\theta \to \phi$ and the process in (5) is not different from white noise.

The next question we are interested in is how much the GARCH(1,1) model explains squared returns. We propose a simple measure using the stochastic volatility (SV) model introduced by Taylor (1986) and Hull and White (1987) and further developed by Harvey and Shephard (1993, 1996) and by Harvey, Ruiz and Shephard (1994). The SV model for the log-squared returns, $\log(y_t^2)$, is

$$\begin{aligned} \log(y_t^2) &= \mu + V_t + \varepsilon_t, \\ V_t &= \varphi V_{t-1} + \eta_t, \end{aligned} \tag{6}$$

where V_t is an unobserved volatility at time t, $Var(\varepsilon_t) = \sigma_\varepsilon^2$, and $Var(\eta_t) = \sigma_\eta^2$.[10] Using more formal correspondence between the GARCH(1,1) model and the SV model, we propose the following theorem.

Theorem 1 *The asymptotically approximated relationship between the parameters of the SV model in (6) and the GARCH(1,1) model in (3) is*

$$\frac{\alpha}{\alpha+\beta} \approx \frac{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2}-\sqrt{\sigma_\varepsilon^2}}{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2}}. \tag{7}$$

Proof. See the Appendix. ■

The right-hand-side component in (7) is the proportion of signal included in the volatility process, which is similar to the signal to noise ratio in state-space models. Therefore, $\frac{\alpha}{\alpha+\beta}$ provides information on how much GARCH(1,1) models explain squared returns. When $\alpha+\beta$ is close to 1, α represents the proportion of squared returns explained by the GARCH(1,1) model. As in our example, when α is relatively small, the GARCH(1,1) process hardly explains squared returns even though $\alpha+\beta$ is close to 1. In extreme cases, when $\alpha = 0$, the GARCH(1,1) volatility is not heteroskedatic any more.[11]

The above explanation becomes clearer with the following example. Suppose that there are two volatility series, A and B, whose GARCH(1,1) parameters are $\alpha_A = 0.04$ and $\beta_A = 0.95$, and $\alpha_B = 0.95$ and $\beta_B = 0.04$, respectively. The conventional interpretation is that both volatility series are equally highly persistent.[12] However, our discussion above suggests the two volatilities are quite different. Volatility B can be interpreted as being persistent and can explain squared returns well, while the persistence of volatility A, although very high, may not be meaningful.

4 Conclusions

In this study we investigated small sample properties of the ML estimates of popular ARCH(1) and GARCH(1,1) models. Using S&P500 index daily returns and Monte Carlo simulations, we showed that a high level of persistence in GARCH(1,1) models obtained using a large number of observations has a correlogram far less than these estimates suggest in small samples. This is consistent with negative biases in small samples in the ML estimates of ARCH and GARCH parameters. Therefore, in small samples there may be serious convergence errors from the negatively biased ARCH and GARCH estimates, which make the ML estimates positively biased by eliminating these negative estimates when Bollerslev's (1986) non-negativity conditions are applied. The results suggest that we should use the weak non-negativity conditions rather than Bollerslev's non-negativity conditions at least as a pre-test. Although the weak non-negativity conditions cannot guarantee positive volatility, they reduce convergence errors significantly. In addition, these estimates can be investigated further, for example, for small sample bias or significance of the negative estimates.

Our results suggest that volatility may be far less persistent than generally thought. In other words, ARCH and GARCH estimates tend to suggest much higher persistence than actual sample autocorrelations do. In partciular when the process is extremely persistent with a small ARCH but a much larger GARCH parameter, sample autocorrelations of squared returns tend to be smaller. We suggest that persistence in the GARCH(1,1) model should be re-investigated not only with $\alpha + \beta$, but also with the individual sizes of α and β. All our results in this study show that the ARCH parameter, α, could be more important than the GARCH parameter, β. If α is small, the persistence ($\alpha + \beta$) has less meaning. For an extremely persistent GARCH process, it is α that explains how much the process can explain squared returns.

We also were able to suggest how many observations should be used for ARCH(1) and GARCH(1,1) models. When the ARCH parameter, α, is large, sample sizes such as 100 observations may give us unbiased estimates without large convergence errors. However, when the ARCH parameter is small, large samples such as 1000 observations may not be enough for unbiased estimates and convergence errors. In general, for the typical sample autocorrelations of squared returns, our results suggest that sample sizes for ARCH(1) and GARCH(1,1) models should be at least 250 and 500 respectively.

Notes

[1]See Drost and Nijman (1993) for further discussion of this topic.

[2]The general properties of small sample estimates in GARCH models are known in econometrics – Engle, Hendry and Trumble (1985), Bollerslev (1988), Diebold and Pauly (1989), and Baillie and Chung (2001). However, they are not well known in applied areas such as finance and economics. In addition, the effects of Bollerslev's (1986) non-negativity conditions on bias and convergence errors in small samples are rarely investigated. Many variations of GARCH models (IGARCH, EGARCH, GARCH with t-distribution, etc.) could be used. However, we believe that the estimates of these variations would show similar properties in small samples to those of the standard ARCH and GARCH models.

[3]We also tried $\{0.7, 0.05\}$, $\{0.3, 0.05\}$, $\{0.05, 0.7\}$, and $\{0.05, 0.3\}$ with Bollerslev's non-negativity conditions, but did not find any local maxima.

[4]We only consider the cases that the standardised residuals are normal. Simulations with other probability density functions could be considered. See Baillie and Chung (2001), for example.

[5]We used a few different algorithms to see if there were any changes in convergence errors, but found little difference.

[6]The unconditional volatilities calculated with the estimates are similar to the true unconditional volatility. Therefore when there is a bias in the ARCH and GARCH

parameters, it is the contant term that is adjusted so that the unconditional volatility calculated with the ML estimates is unbiased.

[7]We also estimated ARCH(1) and GARCH(1,1) models for the FTSE100 daily log-returns using the same method as described in this section and found similar results. The results are not reported here, but can be obtained from the authors upon request.

[8]See Nelson (1991), who raises this point.

[9]This is a reason why unit root tests are affected by the presence of MA components; see the simulation results of Phillips and Perron (1988) and Schwert (1989).

[10]The same experiments as those for GARCH models were carried out for SV models for the S&P500 index log-squared returns. The results are consistent with what we found in table 2 and can be obtained from authors by request.

[11]Andersen and Bollerslev (1998) calculate the population R^2 of the ex-post squared return – GARCH(1,1) volatility regression which is $\frac{\alpha^2}{1-\beta^2-2\alpha\beta}$. Their results are not different from Theorem 1 in the sense that for realistic parameter values of α and β, R^2 is very low and GARCH(1,1) does not explain suqared returns well.

[12]See equation (4) in the previous section for the autocorrelations of ARCH(1) and GARCH(1,1) models.

Appendix Proof of Theorem 1

The reduced form of the SV model in (6) is

$$\log(y_t^2) = \mu^* + \varphi\ \log(y_{t-1}^2) + \xi_t - \theta^*\ \xi_{t-1} \tag{A1}$$

where $\mu^* = \mu(1-\varphi)$, $\xi_t = \varepsilon_t + \eta_t$ and $\theta^* = \frac{\varphi\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2}}$.

Note that asymptotically the GARCH(1,1) model is equivalent to the SV model, since consistent estimates of a stochastic volatility model can be obtained with GARCH models under certain conditions. See Nelson and Foster (1994), Nelson (1996), and Taylor (1994) for examples. Therefore, from equation (5) we have

$$\begin{aligned}
\alpha &= \phi - \theta \\
&\approx \varphi - \theta^* \\
&= \varphi\left(\frac{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2} - \sqrt{\sigma_\varepsilon^2}}{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2}}\right)
\end{aligned}$$

where the second equation comes from the comparison of (5) and (A1), and the third equation can be obtained with $\theta^* = \frac{\varphi\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2+\sigma_\eta^2}}$. Using the approximation $\varphi \approx \phi = \alpha + \beta$, we obtain the theorem.

References

Andersen, T. and T. Bollerslev (1998) Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39(4), 885-905.

Baillie, R. T. and H. Chung (2001) Estimation of GARCH Models from the Autocorrelations of the Squares of a Process, *Journal of Time Series Analysis*, 22, 631-650.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1988) On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process, *Journal of Time Series Analysis*, 9, 121-131.

Diebold, F. X. and P. Pauly (1989) Small Sample Properties of Asymptotically Equivalent Tests for Autoregressive Conditional Heteroscedasticity, *Statistical Papers*, 30, 105-131.

Doornik, J. and M. Ooms (2000) Multimodality and the GARCH Likelihood, mimeo, Nuffield College, University of Oxford.

Drost, F. C. and T. E. Nijman (1993) Temporal Aggregation of GARCH Processes, *Econometrica,* 61, 909-927.

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

Harvey, A. C., E. Ruiz and N. Shephard (1994) Multivariate Stochastic Variance Models, *Review of Economic Studies*, 61, 247-264.

Harvey, A. C. and Shephard, N., 1993. Estimation and Testing of Stochastic Variance Models. Econometrics discussion paper EM/93/268, London School of Economics.

Harvey, A. C. and N. Shephard (1996) Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns, *Journal of Business and Economic Statistics*, 14, 429-34.

He, C and T. Teräsvirta (1999) Properties of the Autocorrelation Function of Squared Observations for Second-order GARCH Processes under Two Sets of Parameter Constraints, *Journal of Time Series Analysis*, 20, 23–30.

Hull, J. and A. White (1987) The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance*, 42, 281-301.

Merton, R. C. (1980) On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics*, 8, 323-361

Nelson, D. B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

Nelson, D. B. and C. Q. Cao (1992) Inequality Constraints in the Univariate GARCH Model, *Journal of Business and Economic Statistics*, 10, 229–235.

Nelson, D. B. and D. P. Foster (1994) Asymptotic Filtering Theory for Univariate ARCH Models, *Econometrica*, 62, 1-41.

Phillips, B. P., and P. Perron (1988) Testing for a Unit Root in Time Series Regression, *Biometrika*, 75(2), 335-346.

Schwert, G. W. (1989) Test for Unit Roots: a Monte Carlo Investigation, *Journal of Business and Economic Statistics*, 7(2), 147-159.

Taylor, S. J. (1986) *Modelling Financial Time Series.* Chichester: John Wiley

Taylor, S. J. (1994) Modeling Stochastic Volatility: A Review and Comparative Study, *Mathematical Finance*, 4, 183-204.

Table 1A Properties of ML Estimates for Different Sample Sizes with Bollerslev's (1986) Non-negativity Conditions

Case 1. When the Data Generating Process is ARCH(1) with w=1 and α=0

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.000				0.00	0.00	0.00	0.00	0.00	0.00
100	0.904	0.090	0.146	0.074	41.9%	0.00	-0.01	-0.01	0.00	-0.01	-0.01
250	0.942	0.054	0.087	0.043	45.5%	-0.01	0.00	0.00	-0.01	0.00	0.00
500	0.960	0.037	0.066	0.030	45.6%	0.00	0.00	0.00	0.00	0.00	0.00
1000	0.970	0.027	0.050	0.020	47.2%	0.00	0.00	0.00	0.00	0.00	0.00

Case 2. When the Data Generating Process is ARCH(1) with w=1 and α=0.1

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.100				0.10	0.01	0.00	0.00	0.00	0.00
100	0.961	0.143	0.173	0.111	70.0%	0.07	-0.01	-0.02	-0.01	-0.01	-0.01
250	0.990	0.112	0.109	0.072	86.8%	0.09	0.00	-0.01	0.00	0.00	-0.01
500	1.001	0.099	0.082	0.052	95.8%	0.09	0.01	0.00	0.00	0.00	0.00
1000	1.002	0.099	0.059	0.042	99.5%	0.10	0.01	0.00	0.00	0.00	0.00

Case 3. When the Data Generating Process is ARCH(1) with w=1 and α=0.5

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.500				0.50	0.25	0.13	0.03	0.00	0.00
100	1.039	0.455	0.241	0.203	98.8%	0.31	0.09	0.02	-0.02	-0.02	-0.02
250	1.019	0.477	0.141	0.126	100.0%	0.37	0.13	0.04	0.00	-0.01	-0.01
500	1.008	0.490	0.101	0.090	100.0%	0.39	0.15	0.06	0.01	0.00	0.00
1000	1.001	0.497	0.068	0.063	100.0%	0.40	0.16	0.06	0.01	0.00	0.00

Case 4. When the Data Generating Process is ARCH(1) with w=1.042 and α=0.232

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.042	0.232				0.21	0.04	0.01	0.00	0.00	0.00
100	1.043	0.229	0.205	0.142	90.0%	0.16	0.02	-0.01	-0.02	-0.02	-0.01
250	1.048	0.225	0.134	0.097	98.6%	0.20	0.03	0.00	-0.01	0.00	-0.01
500	1.047	0.229	0.093	0.071	100.0%	0.21	0.04	0.01	0.00	0.00	0.00
1000	1.044	0.231	0.065	0.049	100.0%	0.22	0.05	0.01	0.00	0.00	0.00

Notes: The simulation results are obtained with 1000 replications for the following ARCH(1) model;

$$y_t = \varepsilon_t h_t^{0.5} \qquad \varepsilon_t \sim N(0,1)$$

$$h_t = w + \alpha y^2_{t-1}$$

where w>0 and α>0. The convergence rate represents the proportion of converged estimates out of the 1000 replications.

Table 1B Properties of ML Estimates for Different Sample Sizes with Weak Non-negativity Conditions

Case 1. When the Data Generating Process is ARCH(1) with w=1 and α=0

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.000				0.00	0.00	0.00	0.00	0.00	0.00
100	0.999	-0.001	0.176	0.111	99.1%	0.00	-0.01	-0.01	0.00	-0.01	-0.01
250	1.005	-0.006	0.109	0.064	99.7%	-0.01	0.00	0.00	-0.01	0.00	0.00
500	0.999	-0.002	0.079	0.045	99.8%	0.00	0.00	0.00	0.00	0.00	0.00
1000	1.000	0.000	0.055	0.032	100.0%	0.00	0.00	0.00	0.00	0.00	0.00

Case 2. When the Data Generating Process is ARCH(1) with w=1 and α=0.1

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.100				0.10	0.01	0.00	0.00	0.00	0.00
100	1.008	0.075	0.195	0.131	99.7%	0.07	-0.01	-0.02	-0.01	-0.01	-0.01
250	1.005	0.093	0.122	0.084	100.0%	0.09	0.00	-0.01	0.00	0.00	-0.01
500	1.004	0.094	0.082	0.057	100.0%	0.09	0.01	0.00	0.00	0.00	0.00
1000	1.002	0.097	0.060	0.042	100.0%	0.10	0.01	0.00	0.00	0.00	0.00

Case 3. When the Data Generating Process is ARCH(1) with w=1 and α=0.5

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.000	0.500				0.50	0.25	0.13	0.03	0.00	0.00
100	1.046	0.454	0.241	0.202	99.9%	0.31	0.09	0.02	-0.02	-0.02	-0.02
250	1.015	0.483	0.144	0.126	100.0%	0.37	0.13	0.04	0.00	-0.01	-0.01
500	1.008	0.493	0.104	0.090	100.0%	0.39	0.15	0.06	0.01	0.00	0.00
1000	0.999	0.498	0.068	0.064	100.0%	0.40	0.16	0.06	0.01	0.00	0.00

Case 4. When the Data Generating Process is ARCH(1) with w=1.402 and α=0.232

Sample Size	Average Value of Estimates		Standard Error of Estimates		Percent of Convergence	Sample Autocorrelations					
	w	α	w	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	1.042	0.232				0.21	0.04	0.01	0.00	0.00	0.00
100	1.069	0.202	0.213	0.162	100.0%	0.16	0.02	-0.01	-0.02	-0.02	-0.01
250	1.045	0.223	0.133	0.100	100.0%	0.20	0.03	0.00	-0.01	0.00	-0.01
500	1.040	0.228	0.090	0.071	100.0%	0.21	0.04	0.01	0.00	0.00	0.00
1000	1.042	0.230	0.066	0.050	100.0%	0.22	0.05	0.01	0.00	0.00	0.00

Notes: The simulation results are obtained with 1000 replications for the following ARCH(1) model;

$$y_t = \varepsilon_t h_t^{0.5} \qquad \varepsilon_t \sim N(0,1)$$

$$h_t = w + \alpha y^2_{t-1}$$

where w>0 and h_t>0. The convergence rate represents the proportion of converged estimates out of the 1000 replications.

Table 2A Properties of ML Estimates for Different Sample Sizes with Bollerslev's (1986) Non-negativity Conditions

Case 1. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.74, and α=0.25

Sample Size	Average Value of Estimates			Standard Error of Estimates			Convergence Rate	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.740	0.250					0.81	0.80	0.80	0.78	0.74	0.67
100	0.039	0.650	0.247	0.058	0.162	0.112	85.4%	0.21	0.19	0.16	0.12	0.05	-0.01
250	0.017	0.721	0.238	0.014	0.084	0.079	98.8%	0.28	0.25	0.23	0.19	0.13	0.05
500	0.013	0.732	0.238	0.007	0.050	0.049	99.4%	0.33	0.31	0.28	0.24	0.17	0.09
1000	0.012	0.736	0.243	0.004	0.034	0.040	99.6%	0.36	0.34	0.32	0.28	0.21	0.12

Case 2. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.60, and α=0.10

Sample Size	Average Value of Estimates			Standard Error of Estimates			Convergence Rate	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.600	0.100					0.11	0.08	0.05	0.03	0.00	0.00
100	0.014	0.471	0.168	0.008	0.222	0.100	16.9%	0.07	0.05	0.03	0.01	-0.02	-0.02
250	0.012	0.520	0.128	0.007	0.231	0.085	42.7%	0.10	0.06	0.04	0.02	0.00	-0.01
500	0.011	0.549	0.104	0.007	0.225	0.050	88.9%	0.10	0.06	0.04	0.02	0.00	0.00
1000	0.011	0.560	0.104	0.005	0.173	0.035	93.1%	0.11	0.07	0.05	0.02	0.00	0.00

Case 3. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.60, and α=0.20

Sample Size	Average Value of Estimates			Standard Error of Estimates			Convergence Rate	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.600	0.200					0.26	0.21	0.17	0.11	0.03	0.00
100	0.015	0.470	0.241	0.009	0.220	0.124	57.4%	0.15	0.12	0.09	0.03	-0.01	-0.02
250	0.012	0.544	0.202	0.007	0.179	0.083	93.9%	0.20	0.14	0.11	0.06	0.01	-0.01
500	0.012	0.565	0.201	0.005	0.136	0.064	98.4%	0.22	0.17	0.14	0.08	0.02	0.00
1000	0.011	0.585	0.200	0.003	0.092	0.043	99.9%	0.23	0.18	0.14	0.09	0.03	0.00

Case 4. When the Data Generating Process is GARCH(1,1) with w=0.005, β=0.943, and α=0.05

Sample Size	Average Value of Estimates			Standard Error of Estimates			Convergence Rate	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.005	0.943	0.050					0.19	0.19	0.19	0.19	0.18	0.17
100	0.223	0.682	0.115	0.337	0.238	0.082	59.1%	0.06	0.05	0.04	0.04	0.02	0.00
250	0.090	0.837	0.085	0.189	0.148	0.042	86.7%	0.11	0.10	0.10	0.09	0.08	0.05
500	0.029	0.891	0.080	0.048	0.071	0.027	95.5%	0.14	0.14	0.14	0.13	0.12	0.09
1000	0.016	0.909	0.077	0.017	0.028	0.017	98.1%	0.17	0.17	0.17	0.16	0.15	0.12

Notes: The simulation results are obtained with 1000 replications for the following GARCH(1,1) model;

$$y_t = \varepsilon_t h_t^{0.5} \qquad \varepsilon_t \sim N(0,1)$$

$$h_t = w + \beta h_{t-1} + \alpha y^2_{t-1}$$

where w>0, β>0, and α>0. The convergence rate represents the proportion of converged estimates out of the 1000 replications.

Table 2B Properties of ML Estimates for Different Sample Sizes with Weak Non-negativity Conditions

Case 1. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.74, and α=0.25

Sample Size	Average Value of Estimates			Standard Error of Estimates			Percent of Convergence	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.740	0.250					0.81	0.80	0.80	0.78	0.74	0.67
100	0.153	0.190	0.211	0.234	0.680	0.155	95.0%	0.21	0.19	0.16	0.12	0.05	-0.01
250	0.033	0.646	0.235	0.090	0.325	0.082	98.5%	0.28	0.25	0.23	0.19	0.13	0.05
500	0.015	0.722	0.238	0.032	0.128	0.054	99.1%	0.33	0.31	0.28	0.24	0.17	0.09
1000	0.011	0.736	0.244	0.004	0.032	0.040	99.9%	0.36	0.34	0.32	0.28	0.21	0.12

Case 2. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.60, and α=0.10

Sample Size	Average Value of Estimates			Standard Error of Estimates			Percent of Convergence	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.600	0.100					0.11	0.08	0.05	0.03	0.00	0.00
100	0.050	-0.620	0.085	0.021	0.524	0.127	94.0%	0.07	0.05	0.03	0.01	-0.02	-0.02
250	0.043	-0.384	0.079	0.023	0.627	0.089	98.0%	0.10	0.06	0.04	0.02	0.00	-0.01
500	0.034	-0.107	0.081	0.024	0.683	0.069	99.5%	0.10	0.06	0.04	0.02	0.00	0.00
1000	0.025	0.143	0.087	0.022	0.640	0.055	99.7%	0.11	0.07	0.05	0.02	0.00	0.00

Case 3. When the Data Generating Process is GARCH(1,1) with w=0.01, β=0.60, and α=0.20

Sample Size	Average Value of Estimates			Standard Error of Estimates			Percent of Convergence	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.010	0.600	0.200					0.26	0.21	0.17	0.11	0.03	0.00
100	0.060	-0.410	0.146	0.038	0.643	0.159	95.5%	0.15	0.12	0.09	0.03	-0.01	-0.02
250	0.037	0.063	0.165	0.036	0.668	0.116	98.9%	0.20	0.14	0.11	0.06	0.01	-0.01
500	0.022	0.363	0.179	0.028	0.517	0.084	99.9%	0.22	0.17	0.14	0.08	0.02	0.00
1000	0.016	0.481	0.189	0.021	0.386	0.063	99.9%	0.23	0.18	0.14	0.09	0.03	0.00

Case 4. When the Data Generating Process is GARCH(1,1) with w=0.005, β=0.943, and α=0.05

Sample Size	Average Value of Estimates			Standard Error of Estimates			Percent of Convergence	Sample Autocorrelations					
	w	β	α	w	β	α		Lag 1	Lag 2	Lag 3	Lag 5	Lag 10	Lag 20
True Values	0.005	0.943	0.050					0.19	0.19	0.19	0.19	0.18	0.17
100	0.914	-0.001	0.076	0.981	0.674	0.136	92.1%	0.06	0.05	0.04	0.04	0.02	0.00
250	0.291	0.574	0.081	0.522	0.574	0.062	91.5%	0.11	0.10	0.10	0.09	0.08	0.05
500	0.052	0.856	0.079	0.161	0.242	0.027	94.2%	0.14	0.14	0.14	0.13	0.12	0.09
1000	0.021	0.906	0.078	0.178	0.063	0.017	98.1%	0.17	0.17	0.17	0.16	0.15	0.12

Notes: The simulation results are obtained with 1000 replications for the following GARCH(1,1) model;

$$y_t = \varepsilon_t h_t^{0.5} \qquad \varepsilon_t \sim N(0,1)$$

$$h_t = w + \beta h_{t-1} + \alpha y^2_{t-1}$$

where w>0 and h_t>0. The convergence rate represents the proportion of converged estimates out of the 1000 replications.

Table 3 ML Estimates of ARCH(1) and GARCH(1,1) Models with Bollerslev's (1986) Non-negativity Conditions for Various Numbers of Observations: S&P500 Daily Index Log-Returns

A. ARCH(1) Estimates

Number of Observations	ARCH(1) Model		AR(1) Estimates Obtained from the ARCH(1) Estimates (See Equations (2) and (5))	
	w	α	w	$\phi\ (\alpha)$
Full Sample (2000 Observations)	1.042 (0.061) *	0.232 (0.059) *	1.042	0.232
2 Subperiods (1000 Observations)	0.554 (0.066) *	0.360 (0.224)	0.554	0.360
(All converged)	1.614 (0.102) *	0.130 (0.053) *	1.614	0.130
Mean of Estimates	1.084 (0.749)	0.245 (0.163)	1.084	0.245
	0.391 (0.042) *	0.013 (0.036)	0.391	0.013
4 Sub-Periods (500 Observations)	0.890 (0.160) *	0.279 (0.334)	0.890	0.279
(All converged)	1.518 (0.119) *	0.070 (0.050)	1.518	0.070
	1.698 (0.172) *	0.186 (0.096) *	1.698	0.186
Mean of Estimates	1.124 (0.599)	0.137 (0.527)	1.124	0.137
	0.509 (0.074) *	0.006 (0.039)	0.509	0.006
8 Sub-periods (250 Observations)	0.789 (0.101) *	0.014 (0.063)	0.789	0.014
(7 Convergences out of 8)	0.964 (0.410) *	0.539 (0.854)	0.964	0.539
	1.475 (0.152) *	0.043 (0.044)	1.475	0.043
	1.530 (0.179) *	0.099 (0.102)	1.530	0.099
	1.684 (0.226) *	0.092 (0.117)	1.684	0.092
	1.720 (0.253) *	0.265 (0.143) *	1.720	0.265
Mean of Estimates	1.239 (0.480)	0.151 (0.192)	1.239	0.151
	0.169 (0.038) *	0.238 (0.265)	0.169	0.238
	0.389 (0.137) *	0.090 (0.352)	0.389	0.090
20 Sub-periods (100 Observations)	1.005 (0.172) *	0.010 (0.081)	1.005	0.010
(11 Convergences out of 20)	1.590 (0.524) *	0.116 (0.091)	1.590	0.116
	0.720 (0.148) *	0.064 (0.168)	0.720	0.064
	1.369 (0.566) *	0.261 (0.823)	1.369	0.261
	1.672 (0.324) *	0.108 (0.078)	1.672	0.108
	1.385 (0.277) *	0.249 (0.239)	1.385	0.249
	1.835 (0.393) *	0.117 (0.237)	1.835	0.117
	1.982 (0.398) *	0.049 (0.111)	1.982	0.049
	2.572 (0.617) *	0.292 (0.273)	2.572	0.292
Mean of Estimates	1.335 (0.716)	0.145 (0.097)	1.335	0.145

Notes : The table reports the Maximum Likelihood Estimates for the ARCH(1) model for the daily log-returns of the S&P500 indx from 22 September 1994 to 30 August 2002 for a total of 2000 observations Numbers in parentheses are robust standard errors. The ARCH(1) model is

$$y_t = \xi_t h_t^{1/2},$$
$$h_t = w + \alpha y^2_{t-1}$$

where y_t is residuals and $\xi_t \sim N(0,1)$. The estimates of the AR(1) model are obtained using equation (5) with β=0. * represents significance at the 10% level.

B. GARCH(1,1) Estimates

Number of Observations	GARCH(1,1) Model			ARMA(1,1) Estimates Obtained from the GARCH(1,1) Estimates (See equations (3)and (5))		
	w	β	α	w	ϕ $(\alpha+\beta)$	θ (−β)
Full Sample (2000 Observations)	0.010 (0.006) *	0.914 (0.025) *	0.079 (0.023) *	0.010	0.993	-0.914
2 Subperiods (1000 Observations)	0.009 (0.007)	0.910 (0.044) *	0.082 (0.040) *	0.009	0.992	-0.910
(All converged)	0.107 (0.043) *	0.854 (0.036) *	0.087 (0.022) *	0.107	0.941	-0.854
Mean of Estimates	0.058 (0.070)	0.882 (0.039)	0.084 (0.004)	0.058	0.966	-0.882
4 Sub-Periods (500 Observations)	0.007 (0.006)	0.941 (0.029) *	0.043 (0.021) *	0.007	0.985	-0.941
(All converged)	0.071 (0.041) *	0.797 (0.085) *	0.153 (0.076) *	0.071	0.949	-0.797
	0.073 (0.055)	0.906 (0.040) *	0.048 (0.020) *	0.073	0.954	-0.906
	0.138 (0.051) *	0.814 (0.042) *	0.117 (0.033) *	0.138	0.931	-0.814
Mean of Estimates	0.072 (0.053)	0.864 (0.070)	0.090 (0.054)	0.072	0.955	-0.864
	0.047 (0.047)	0.860 (0.105) *	0.048 (0.040)	0.047	0.908	-0.860
8 Sub-periods (250 Observations)	0.135 (0.081) *	0.682 (0.132) *	0.273 (0.149) *	0.135	0.955	-0.682
(6 Convergences out of 8)	0.090 (0.057)	0.894 (0.043) *	0.045 (0.020) *	0.090	0.939	-0.894
	0.306 (0.323)	0.694 (0.246) *	0.127 (0.122)	0.306	0.821	-0.694
	0.142 (0.091)	0.815 (0.080) *	0.106 (0.053) *	0.142	0.921	-0.815
	0.083 (0.155)	0.855 (0.111) *	0.113 (0.036) *	0.083	0.968	-0.855
Mean of Estimates	0.134 (0.091)	0.800 (0.090)	0.119 (0.083)	0.134	0.919	-0.800
	0.019 (0.031)	0.883 (0.096) *	0.066 (0.047)	0.019	0.949	-0.883
20 Sub-periods (100 Observations)	0.514 (0.258) *	0.501 (0.088) *	0.256 (0.361)	0.514	0.757	-0.501
(9 Convergences out of 20)	0.131 (0.269)	0.760 (0.330) *	0.180 (0.234)	0.131	0.940	-0.760
	0.120 (0.080)	0.850 (0.045) *	0.079 (0.036) *	0.120	0.930	-0.850
	0.312 (0.456)	0.780 (0.277) *	0.038 (0.109)	0.312	0.818	-0.780
	0.849 (1.349)	0.441 (0.702)	0.156 (0.220)	0.849	0.597	-0.441
	0.122 (0.154)	0.838 (0.108) *	0.095 (0.059)	0.122	0.933	-0.838
	0.222 (0.129) *	0.760 (0.087) *	0.106 (0.072)	0.222	0.867	-0.760
	0.288 (0.188)	0.769 (0.084) *	0.140 (0.073) *	0.288	0.910	-0.769
Mean of Estimates	0.286 (0.256)	0.732 (0.155)	0.124 (0.067)	0.286	0.856	-0.732

Notes : The table reports the Maximum Likelihood Estimates for the GARCH(1,1) model for the daily log-returns of the S&P500 indx from 22 September 1994 to 30 August 2002 for a total of 2000 observations. Numbers in parentheses are robust standard errors. The GARCH(1,1) model is

$$y_t = \xi_t h_t^{1/2},$$
$$h_t = w + \beta h_{t-1} + \alpha y^2_{t-1}.$$

where y_t is residuals and $\xi_t \sim N(0,1)$. The estimates of the ARMA(1,1) model are obtained using the relationship with GARCH(1,1) in equation (5). * represents significance at the 10% level.

Table 4 ML Estimates of ARCH(1) and GARCH(1,1) Models with Weak Non-negativity Conditions for Various Numbers of Observations: S&P500 Daily Index Log-Returns

A. ARCH(1) Estimates

Number of Observations	ARCH(1) Model		AR(1) Estimates Obtained from the ARCH(1) Estimates (See Equations (2) and (5))	
	w	α	w	$\phi\ (\alpha)$
Full Sample (2000 Observations)	1.042 (0.061)	0.232 (0.059) *	1.042	0.232
2 Sub-Periods	0.554 (0.066)	0.360 (0.224)	0.554	0.360
(All Converged)	1.614 (0.102)	0.130 (0.053) *	1.614	0.130
Mean of Estimates	1.084 (0.749)	0.245 (0.163)	1.084	0.245
4 Sub-Periods	0.391 (0.042)	0.013 (0.036)	0.391	0.013
(500 Observations)	0.890 (0.160)	0.279 (0.334)	0.890	0.279
(All Converged)	1.518 (0.119)	0.070 (0.050)	1.518	0.070
	1.698 (0.172) *	0.186 (0.096) *	1.698	0.186
Mean of Estimates	1.124 (0.599)	0.137 (0.119)	1.124	0.137
	0.304 (0.036)	-0.079 (0.010) *	0.304	-0.079
	0.509 (0.074)	0.006 (0.039)	0.509	0.006
	0.789 (0.101)	0.014 (0.063)	0.789	0.014
8 Sub-periods	0.964 (0.410)	0.539 (0.854)	0.964	0.539
(250 Observations)	1.475 (0.152)	0.043 (0.044)	1.475	0.043
(All Converged)	1.530 (0.179)	0.099 (0.102)	1.530	0.099
	1.684 (0.226) *	0.092 (0.117)	1.684	0.092
	1.720 (0.253) *	0.265 (0.143) *	1.720	0.265
Mean of Estimates	1.122 (0.554)	0.122 (0.196)	1.122	0.122
	0.374 (0.068)	-0.106 (0.020) *	0.374	-0.106
	0.289 (0.051)	-0.095 (0.018) *	0.289	-0.095
	0.169 (0.038)	0.238 (0.265)	0.169	0.238
	0.623 (0.134)	-0.005 (0.033)	0.623	-0.005
	0.567 (0.145)	-0.028 (0.129)	0.567	-0.028
	0.389 (0.137)	0.090 (0.352)	0.389	0.090
20 Sub-periods	1.005 (0.172)	0.010 (0.081)	1.005	0.010
(100 Observations)	1.590 (0.524)	0.116 (0.091)	1.590	0.116
(All Converged)	0.720 (0.148)	0.064 (0.167)	0.720	0.064
	1.369 (0.566)	0.261 (0.823)	1.369	0.261
	1.672 (0.324) *	0.108 (0.078)	1.672	0.108
	1.606 (0.248)	-0.130 (0.090)	1.606	-0.130
	1.377 (0.211)	-0.062 (0.030) *	1.377	-0.062
	1.901 (0.380) *	-0.085 (0.017) *	1.901	-0.085
	1.385 (0.277)	0.249 (0.239)	1.385	0.249
	1.835 (0.393) *	0.117 (0.237)	1.835	0.117
	1.982 (0.398) *	0.049 (0.111)	1.982	0.049
	1.837 (0.352) *	-0.062 (0.018) *	1.837	-0.062
	0.982 (0.173)	-0.001 (0.072)	0.982	-0.001
	2.572 (0.617) *	0.292 (0.273)	2.572	0.292
Mean of Estimates	1.212 (0.677)	0.051 (0.131)	1.212	0.051

Notes : The table reports the Maximum Likelihood Estimates for the ARCH(1) model for the daily log-returns of the S&P500 indx from 22 September 1994 to 30 August 2002 for a total of 2000 observations Numbers in parentheses are robust standard errors. The ARCH(1) model is

$$y_t = \xi_t h_t^{1/2},$$
$$h_t = w + \alpha y^2_{t-1}$$

where y_t is residuals and $\xi_t \sim N(0,1)$. The estimates of the AR(1) model are obtained using equation (5) with β=0. * represents significance at the 10% level.

B. GARCH(1,1) Estimates

Number of Observations	GARCH(1,1) Model			ARMA(1,1) Estimates Obtained from the GARCH(1,1) Estimates (See equations (3) and (5))		
	w	β	α	w	ϕ $(\alpha+\beta)$	θ $(-\beta)$
Full Sample (2000 Observations)	0.010 (0.006) *	0.914 (0.025) *	0.079 (0.023) *	0.010	0.993	-0.914
2 Sub-Periods (All Converged)	0.009 (0.007)	0.910 (0.044) *	0.082 (0.040) *	0.009	0.992	-0.910
	0.107 (0.043) *	0.854 (0.036) *	0.087 (0.022) *	0.107	0.941	-0.854
Mean of Estimates	0.058 (0.070)	0.882 (0.039)	0.084 (0.004)	0.058	0.966	-0.882
4 Sub-Periods (500 Observations) (All Converged)	0.007 (0.006)	0.941 (0.029) *	0.043 (0.021) *	0.007	0.985	-0.941
	0.071 (0.041) *	0.797 (0.085) *	0.153 (0.076) *	0.071	0.949	-0.797
	0.073 (0.055)	0.906 (0.040) *	0.048 (0.020) *	0.073	0.954	-0.906
	0.138 (0.051) *	0.814 (0.042) *	0.117 (0.033) *	0.138	0.931	-0.814
Mean of Estimates	0.072 (0.053)	0.864 (0.070)	0.090 (0.054)	0.072	0.955	-0.864
8 Sub-periods (250 Observations) (All Converged)	0.361 (0.417)	-0.198 (1.422)	-0.081 (0.019) *	0.361	-0.280	0.198
	0.047 (0.047)	0.860 (0.105) *	0.048 (0.040)	0.047	0.908	-0.860
	1.515 (0.168) *	-0.932 (0.075) *	0.033 (0.029)	1.515	-0.898	0.932
	0.135 (0.081) *	0.682 (0.132) *	0.273 (0.149) *	0.135	0.955	-0.682
	0.090 (0.057)	0.894 (0.043) *	0.045 (0.020) *	0.090	0.939	-0.894
	0.306 (0.323)	0.694 (0.246) *	0.127 (0.122)	0.306	0.821	-0.694
	0.142 (0.091)	0.815 (0.080) *	0.106 (0.053) *	0.142	0.921	-0.815
	0.083 (0.155)	0.855 (0.111) *	0.113 (0.036) *	0.083	0.968	-0.855
Mean of Estimates	0.335 (0.490)	0.459 (0.666)	0.083 (0.101)	0.335	0.542	-0.459
20 Sub-periods (100 Observations) (19 Convergences out of 20)	0.207 (0.057) *	0.486 (0.179) *	-0.104 (0.019) *	0.207	0.382	-0.486
	0.341 (0.085) *	-0.206 (0.197)	-0.091 (0.017) *	0.341	-0.297	0.206
	0.411 (0.056) *	-0.993 (0.008) *	0.110 (0.053) *	0.411	-0.884	0.993
	1.224 (0.236) *	-0.929 (0.059) *	-0.034 (0.023)	1.224	-0.963	0.929
	0.082 (0.044) *	0.782 (0.066) *	0.068 (0.071)	0.082	0.850	-0.782
	0.821 (0.120) *	-1.032 (0.003) *	0.057 (0.023) *	0.821	-0.975	1.032
	1.969 (0.262) *	-1.036 (0.003) *	0.078 (0.020) *	1.969	-0.958	1.036
	0.514 (0.258) *	0.501 (0.088) *	0.256 (0.361)	0.514	0.757	-0.501
	0.756 (0.448) *	-0.047 (0.505)	0.064 (0.167)	0.756	0.017	0.047
	0.131 (0.269)	0.760 (0.330) *	0.180 (0.234)	0.131	0.940	-0.760
	0.120 (0.080)	0.850 (0.045) *	0.079 (0.036) *	0.120	0.930	-0.850
	1.877 (0.350) *	-0.197 (0.236)	-0.125 (0.087)	1.877	-0.322	0.197
	2.018 (0.406) *	-0.496 (0.178) *	-0.065 (0.028) *	2.018	-0.561	0.496
	2.090 (1.086)	-0.104 (0.505)	-0.086 (0.021) *	2.090	-0.191	0.104
	0.849 (1.349)	0.441 (0.702)	0.156 (0.220)	0.849	0.597	-0.441
	0.122 (0.154)	0.838 (0.108) *	0.095 (0.059)	0.122	0.933	-0.838
	0.222 (0.129) *	0.760 (0.087) *	0.106 (0.072)	0.222	0.867	-0.760
	1.921 (0.292) *	-0.951 (0.171) *	-0.013 (0.025)	1.921	-0.964	0.951
	0.288 (0.188)	0.769 (0.084) *	0.140 (0.073) *	0.288	0.910	-0.769
Mean of Estimates	0.840 (0.759)	0.010 (0.734)	0.046 (0.107)	0.840	0.056	-0.010

Notes : The table reports the Maximum Likelihood Estimates for the GARCH(1,1) model for the daily log-returns of the S&P500 indx from 22 September 1994 to 30 August 2002 for a total of 2000 observations. Numbers in parentheses are robust standard errors. The GARCH(1,1) model is

$$y_t = \xi_t h_t^{1/2},$$

$$h_t = w + \beta h_{t-1} + \alpha y^2_{t-1}.$$

where y_t is residuals and $\xi_t \sim N(0,1)$. The estimates of the ARMA(1,1) model are obtained using the relationship with GARCH(1,1) in equation (5). * represents significance at the 10% level.