Recursive Modelling of Symmetric and Asymmetric Volatility in the Presence of Extreme Observations*

Hock Guan Ng

Department of Accounting and Finance
University of Western Australia

Michael McAleer

Department of Economics
University of Western Australia

Revised: June 2002

* The authors wish to thank two referees for useful comments and suggestions, and Felix Chan and Thierry Jeantheau for helpful discussions. The second author is grateful for the financial support of the Australian Research Council.

Corresponding author: Hock Guan Ng, Department of Accounting and Finance, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia, telephone: +61 8 9380 2971, facsimile: +61 8 9380 1047, e-mail: Hock.Guan.Ng@uwa.edu.au.

Recursive Modelling of Symmetric and Asymmetric Volatility in the Presence of Extreme Observations

Abstract

This paper is concerned with recursive estimation, testing and forecasting of the asymmetric volatility of daily returns in Standard and Poor's 500 Composite Index and the Nikkei 225 Index in the presence of extreme observations, or significant spikes in the volatility of daily returns. For each of the two data sets, the empirical analysis increases the sample size up to 12000 observations recursively to examine the effects of extreme observations on: (i) the Quasi Maximum Likelihood Estimates (QMLE) of the GARCH(1,1) and asymmetric GJR(1,1) parameters; (ii) the associated asymptotic and robust t-ratios of the QMLE; (iii) recursive statistical testing of the asymmetry parameter in GJR(1,1); (iv) the sufficient second and fourth moment conditions for consistency and asymptotic normality, respectively, of the QMLE of GARCH(1,1) and GJR(1,1); and (v) the forecast performance of the GARCH(1,1) and GJR(1,1) models for periods with significant spikes in volatility and for periods of relative calm.

Keywords: Outliers, extreme observations, time-varying volatility, symmetry, asymmetry, leverage, moment conditions, recursive modelling, structural change.

1 Introduction

Given the importance of risk in economic and financial markets, and the use of volatility in evaluating risks, asymmetric shocks and leverage effects, it is not surprising that time-varying volatility has become an active area of research in finance in recent years. Engle (1982) captured the time-varying nature of volatility by developing the autoregressive conditional heteroscedasticity (ARCH(p)) model. Bollerslev (1986) generalized the ARCH model to GARCH (p, q), which has subsequently become the most popular model of time-varying symmetric volatility in practice. The GARCH specification has several attractive features, namely the ability to accommodate key stylised facts of volatility in financial data, such as the persistence of volatility and volatility clusters, and leptokurtic data, as well as mathematical and computational simplicity. Glosten, Jagannathan and Runkle (1993) modified the GARCH(p,q) model to GJR(p,q) by accommodating the asymmetric responses of volatility to positive and negative shocks. The ease of interpretation and application has also made the GJR(p,q) model very popular among financial practitioners.

A number of important structural properties of the models and the asymptotic theory underlying a variety of estimation methods have recently been established for GARCH and GJR. These theoretical developments provide a solid foundation for applying the various models in practice (see Li, Ling and McAleer (2002) for a survey of recent theoretical results associated with GARCH models).

Extreme observations and outliers, or significant spikes in volatility, are commonly observed in high frequency financial time series. Such observations can adversely affect the estimates of the parameters and forecasts of volatility. Questions arise as to how these aberrant observations should be accommodated in estimation, testing and forecasting. In this paper, we investigate the optimal number of observations to be used from two large data sets that include extreme observations.

Using the daily returns in Standard and Poor's Composite 500 Index (S&P 500) and the Nikkei 225 Index, the paper is concerned with recursive estimation, testing and forecasting of the symmetric and asymmetric volatility of daily returns in the presence of extreme observations. The empirical analysis increases the sample size recursively up to 12000 observations in order to examine the effects of extreme observations in the data on: (i) the Quasi Maximum Likelihood Estimates (QMLE) of the GARCH(1,1) and asymmetric GJR(1,1) parameters; (ii) the associated asymptotic and robust t-ratios of the QMLE of GARCH(1,1) and GJR(1,1); (iii) recursive statistical testing of the asymmetry parameter in GJR(1,1); (iv) the sufficient second and fourth moment conditions for consistency and asymptotic normality, respectively, of the QMLE of GARCH(1,1) and GJR(1,1); and (v) the forecast performances of GARCH(1,1) and GJR(1,1) for periods with significant spikes in volatility and for periods of relative calm.

Several interesting results emerge from the empirical analysis, namely: expanding the sample size recursively and including an extreme observation does not necessarily improve the accuracy of predicting future extreme observations; the parameter estimates of the GARCH(1,1) and GJR(1,1) processes, their associated asymptotic and robust t-ratios, the second and fourth moment regularity conditions, and various forecast performance measures, are all highly volatile in small samples, but stabilise when an extreme observation is included in the estimation period at sample sizes in excess of 2000; increasing the sample size recursively beyond an extreme observation is unnecessary; the robust t-ratios are, in general, dramatically superior to their asymptotic counterparts; the second moment condition is always satisfied in the case of S&P 500, but not so in the case of Nikkei 225; if the conditional (or standardised) error is normal, the fourth moment condition is generally satisfied for S&P 500 but not for Nikkei 225; if the conditional error follows a fatter-tailed distribution such as the $t(5)$ distribution, the fourth moment condition is generally not satisfied; increasing the sample size recursively does not necessarily lead to the moment conditions being satisfied; increasing the sample size recursively does not necessarily lead to improved forecasts; the GARCH(1,1) and GJR(1,1) models are superior to the RiskMetrics model in forecasting volatility; and neither GARCH(1,1) nor GJR(1,1) dominates the other.

The plan of the paper is as follows. Section 2 presents the structural properties of the GARCH(1,1) and GJR(1,1) models and the associated asymptotic theory. Section 3 describes the data. The empirical estimates and forecasts are analysed in Section 4. Some concluding remarks are given in Section 5.

2 The Symmetric GARCH and Asymmetric GJR Models

Both volatility models to be estimated are associated with a stationary AR(1) conditional means (for the logarithmic returns of the S&P 500 and Nikkei 225 Indexes) given by

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t , \qquad |\phi| < 1 . \tag{1}$$

2.1 GARCH(1,1)

For the GARCH(1,1) model, the conditional variance of the unconditional shock ε_t is given by

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{2}$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{3}$$

where η_t is a sequence of normally, independently and identically distributed random variables with zero mean and unit variance. Sufficient conditions for h_t to be positive, and hence for the GARCH process to exist, are that $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$. The ARCH (or α) effect indicates the short run persistence of shocks, while the GARCH (or β) effect indicates the contribution of shocks to long run persistence (namely, $\alpha + \beta$).

Several structural properties have been established for the GARCH(1,1) process in order to define the unconditional moments of ε_t. The second moment of ε_t exists, that is $E\varepsilon_t^2 < \infty$, if $\alpha + \beta < 1$, which is sufficient to ensure that the GARCH(1,1) process is strictly stationary and ergodic (see Bollerslev (1986) and Ling and Li (1997)). A sufficient condition for the existence of the fourth moment of ε_t is $k\alpha^2 + 2\alpha\beta + \beta^2 < 1$ (see Bollerslev (1986)), where k is the conditional fourth moment of η_t. Under the assumption of conditional normality, $k \equiv E_t(\eta_t^4) = 3$, so that the condition becomes

$$S_{G(N)} \equiv 3\alpha^2 + 2\alpha\beta + \beta^2 < 1. \tag{4}$$

A common alternative assumption is that η_t is distributed according to the *t* distribution with $\nu > 4$ degrees of freedom, in which case $k = 3(\nu - 2)/(\nu - 4)$, with $3 \leq k \leq 9$. In the extreme case $\nu = 5$, the condition becomes

$$S_{G(t)} \equiv 9\alpha^2 + 2\alpha\beta + \beta^2 < 1. \tag{5}$$

More generally, Ling and McAleer (2002a) derived the necessary and sufficient conditions for the existence of all the moments of the GARCH(p,q) model.

For the GARCH(1,1) model, Nelson (1991) obtained the necessary and sufficient log-moment condition for strict stationarity and ergodicity as:

$$E(\ln(\alpha\eta_t^2 + \beta)) < 0. \tag{6}$$

A difficulty in applying the necessary and sufficient condition in (4) is that it is a function of a random variable and unknown parameters, and hence needs to be simulated or estimated. Unlike the second moment condition, the log-moment condition allows $\alpha + \beta > 1$, in which case $E\varepsilon_t^2 = \infty$. The condition for a finite variance of the GARCH(1,1) process is $\alpha + \beta < 1$ and, as given above, the condition for finite fourth moment under normality is (4). The fourth moment condition is clearly more stringent than its second moment counterpart, which in turn is stronger then the log-moment condition.

In the absence of normality of η_t, the parameters of the GARCH(1,1) model are typically estimated by the maximum likelihood method to obtain Quasi-Maximum Likelihood Estimators (QMLE). Ling and McAleer (2002b) showed that the QMLE for GARCH(p,q) is consistent if the second moment of the unconditional shocks is finite. For GARCH(p,q), Ling and Li (1997) demonstrated that the local QMLE is asymptotically normal if the fourth moment of the unconditional shocks is finite, while Ling and McAleer (2002b) proved that the global QMLE is asymptotically normal if the sixth moment is finite.

2.2 GJR(1,1)

For the GARCH model, positive shocks are assumed to have the same effect on conditional volatility as negative shocks. In order to accommodate asymmetric behaviour, the GJR(1,1) model incorporates a stochastic indicator variable I_{t-1} in the conditional variance equation, as follows:

$$h_t = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{7}$$

where I_{t-1} takes the value 1 when $\varepsilon_{t-1} < 0$, and 0 otherwise. The indicator variable differentiates between positive and negative shocks, so that asymmetric effects in the data are captured by γ, with $\gamma > 0$. The asymmetric effect, γ, measures the contribution of shocks to both short run persistence, $\alpha + \frac{\gamma}{2}$, and long run persistence, $\alpha + \beta + \frac{\gamma}{2}$.

Ling and McAleer (2002c) established the sufficient conditions for the second moment of ε_t (under symmetry of the standardised shock) and fourth moment of ε_t to exist as $\alpha + \beta + \frac{\gamma}{2} < 1$ and $k\alpha^2 + 2\alpha\beta + \beta^2 + \beta\gamma + k\alpha\gamma + \frac{k\gamma^2}{2} < 1$, respectively. If it is assumed that η_t is distributed as $N(0,1)$, the fourth moment condition becomes

$$S_{GJR(N)} \equiv 3\alpha^2 + 2\alpha\beta + \beta^2 + \beta\gamma + 3\alpha\gamma + \frac{3\gamma^2}{2} < 1. \tag{8}$$

However, if η_t is distributed as $t(5)$, then the fourth moment condition is given as

$$S_{GJR(t)} \equiv 9\alpha^2 + 2\alpha\beta + \beta^2 + \beta\gamma + 9\alpha\gamma + \frac{9\gamma^2}{2} < 1. \tag{9}$$

These conditions make it clear that the admissible region of (α, β) for second- and fourth-order stationarity of the asymmetric GJR(1,1) model is smaller than that for its symmetric GARCH(1,1) counterpart, as the asymmetry of the model increases its uncertainty.

Although the regularity conditions for the existence of moments for the GJR model are now known, there are as yet no theoretical results regarding the statistical properties of the model. In practice, it is assumed that the QMLE are consistent and asymptotically normal.

3 Data

The daily closing values of the S&P 500 Index for the period 3 January 1950 to 5 May 1998, and of the Nikkei 225 Index for the period 10 May 1951 to 22 April 1998, were extracted from the Datastream database. The daily return for each index was calculated as the ratio of the close-to-close change in the index to the previous trading day's close.

These two indexes were chosen for the availability of daily observations over an extended period. The sample period for each index was chosen such that each series has 12000 observations covering approximately the same time period as the other. The long sample periods include many significant spikes in the volatility of each set of daily returns, as well as many episodes of relative calm. An important date in the sample period is 19 October 1987 in the USA, or 20 October 1987 in Japan, as this is when the largest volatility spikes for both series occurred[1]. Consequently, this data set offers an invaluable opportunity to study the effects of extreme observations on the estimation, testing and forecasting of volatility over an extended period.

[1] We shall henceforth refer to this observation by the US date of 19 October 1987.

Various subsets of the data are used for estimation, testing and forecasting. In order to evaluate the effects of extreme observations on estimation, 12000 observations of each series are used, with the sample period for each series ending on 7 May 1997. For the evaluation of forecasting performance, various measures based on one-period ahead forecasts over two separate "out-of-sample" periods are used, each consisting of 250 observations; the first of these, starting from 8 May 1997 for both series, includes some significant spikes in the volatility of daily returns, while the second, ending on 7 May 1997 for both series, is a period of relative calm.

4 Empirical Results

4.1 Estimation Results

In order to evaluate the effects of increasing the sample size and including extreme observations, the GARCH(1,1) and GJR(1,1) models are estimated recursively. In each set of estimates, the end observation of the sample is fixed at 7 May 1997. For the S&P 500 series, the sample begins with 200 observations from 23 July 1996 to 7 May 1997, and is then expanded backward recursively until it reaches 12000 observations at 3 January 1950. For Nikkei 225, the sample expands from the 200 observations over the period 1 August 1996 to 7 May 1997 to the 12000 observations over the period 10 May 1951 to 7 May 1997.

Figures 1A and 1B show the estimated values of the ARCH parameter α of the GARCH(1,1) model as the sample size is increased recursively using the S&P 500 and Nikkei 225 data sets, respectively[2]. The actual volatility of the daily returns is shown in the lower half of each figure to indicate where the volatility spikes occur. It is clear that the estimates of α are highly volatile when the sample sizes are below 2400. Significant spikes in the actual volatility correspond to huge variations in the estimates of α. The most obvious feature common to both figures is the huge shift in the α estimates with the 19 October 1987 spike in volatility, after which the variations in the α estimates are much smaller in magnitude. Other notable features are the general U-shape in the middle of Figure 1A and the cyclical pattern in Figure 1B, both of which indicate that α is not constant over time.

Figures 2A and 2B present the asymptotic t-ratios, as well as the robust t-ratios of Bollerslev and Wooldridge (1992), for estimates of α in the GARCH(1,1) model. The robust t-ratios are designed to be insensitive to departures from normality, especially extreme observations. Both sets of t-ratios in each figure are somewhat erratic at small sample sizes and are more sensitive to extreme observations before the inclusion of the 19 October 1987 spike, but when the sample size exceeds 400 both t-ratios exceed the critical value for the null hypothesis that $\alpha = 0$. The effects of significant spikes in volatility on the two sets of t-ratios are also dramatically different. Each spike in volatility increases the asymptotic t-ratios but decreases the robust t-ratios, with the magnitudes of the shifts being far greater for the asymptotic t-ratios. It is worth noting the huge increase in the asymptotic t-ratios when the 19 October 1987 spike is included. In contrast, the impact of this extreme observation on the robust t-ratios is barely visible.

[2] Throughout the rest of the paper, the "A" figures represent results for S&P 500 and the "B" figures represent results for Nikkei 225.

Estimates of the GARCH parameter β of the GARCH(1,1) model are given in Figures 3A and 3B. These are virtually mirror images of the estimates of α, with the β estimates moving in the opposite direction to those of α. There is also much variability in the β estimates at sample sizes below 2400. As the sample size is increased beyond 2400, Figure 3A shows an inverted U-shape while Figure 3B shows a cyclical pattern. The spikes in volatility also have larger impacts on the β estimates when the sample size is small.

Figures 4A and 4B show the t-ratios for the β estimates in the GARCH(1,1) model. Both the asymptotic and robust t-ratios in each of these figures show greater variability for sample sizes below 2500, prior to the inclusion of the 19 October 1987 spike in volatility. After the inclusion of this extreme observation, both t-ratios become much smoother, especially the robust t-ratios. Throughout all sample sizes, both t-ratios for each data set exceed the critical value for the null hypothesis that $\beta = 0$.

The second moment condition for stationarity and consistency of the GARCH(1,1) model, as discussed above, is $\alpha + \beta < 1$. Figures 5A and 5B show the value of the estimated $\alpha + \beta$. Spikes in the volatility of returns have large impacts on this value when the sample size is below 2400. For large sample sizes, and with the inclusion of the 19 October 1987 spike, this value is less volatile, but it is also not constant. It is significant to note that the second moment condition is satisfied for all sample sizes in the backward recursions with the S&P 500 data, but the same does not hold true with the Nikkei 225 data.

Since the fourth moment condition for asymptotic normality is condition (4) when η_t is distributed as $N(0,1)$, and condition (5) when η_t is distributed as $t(5)$, Figures 6A and 6B show the values of $S_{G(N)}$ and $S_{G(t)}$. As discussed previously, $S_{G(N)}$ is smaller and hence is more likely to be less than unity. Its value in the recursions follows a pattern that is identical to that of the second moment condition, and is less than unity for most of the S&P 500 sample, but is greater than unity for most of the Nikkei 225 sample. While following the same pattern in fluctuations, $S_{G(t)}$ is greater in value and exceeds unity for all sample ranges in both data sets that include the 19 October 1987 volatility spike.

Figures 7A and 7B show the α estimates for the GJR(1,1) model having severe fluctuations at small sample sizes, with negative values occurring for sample sizes below 1400 observations for both S&P 500 and Nikkei 225. As in the GARCH(1,1) model, the fluctuations are less severe at sample sizes above 2400, after the inclusion of the 19 October 1987 volatility spike. The estimated value of α does not appear to be constant in this model.

The graphs of the t-ratios for the α estimates of the GJR(1,1) model in Figure 8A and 8B are dramatically different from those of their GARCH(1,1) counterparts. Absent from both Figures 8A and 8B are the dramatic shifts in the asymptotic t-ratios when the 19 October 1987 volatility spike is included. Instead the figures show volatility spikes in the distant past having much larger impacts on the α t-ratios, especially the asymptotic t-ratios, of the GJR(1,1) model. Another significant difference is that the t-ratios do not exceed the critical value for the hypothesis that $\alpha = 0$ until the sample size exceeds 2500 for both S&P 500 and Nikkei 225.

The general pattern of the β estimates of the GJR(1,1) model shown in Figures 9A and 9B are similar to those of the GARCH(1,1) model. There are large fluctuations for sample sizes below 2400, but they disappear once the sample size expands beyond the inclusion of the 19

October 1987 observation. The persistent variability of the β estimates as the sample is expanded is also present in each of these two figures.

In Figures 10A and 10B, the graphs of the t-ratios for the β estimates of the GJR(1,1) model are almost identical to those of the GARCH(1,1) models in Figures 4A and 4B, except when the sample size is below 1000. Both t-ratios are volatile for small sample sizes, especially with the Nikkei 225 data, but are less so when the samples are expanded beyond the 19 October 1987 extreme observation.

Figures 11A and 11B display the asymmetry, or γ, estimates of the GJR(1,1) model. In both cases, large fluctuations in the estimates of γ are observed for small sample sizes. When the samples exceed 2500 observations, the fluctuations are greatly reduced, but the estimated values of γ continue to change over time.

The graphs of the t-ratios for the γ estimates of the GJR(1,1) model in Figures 12A and 12B show volatility spikes having significantly different impacts on the two t-ratios. Each extreme observation has a large and positive impact on the asymptotic t-ratio, but a smaller and negative impact on its robust counterpart. While the asymptotic t-ratios always exceed the critical value for the null hypothesis that $\gamma = 0$ when the sample size exceeds 500, the robust t-ratios do not always exceed the same critical value.

The results of testing the second moment condition for the GJR(1,1) model, as shown in Figures 13A and 13B, are very similar to those for the GARCH(1,1) model. The second moment condition requires $\alpha + \beta + \frac{\gamma}{2} < 1$. Figure 13A show that the estimate of $\alpha + \beta + \frac{\gamma}{2}$

for S&P 500 is volatile for small sample sizes, more stable for large sample sizes, and does not exceed unity at any stage. The estimates for Nikkei 225 show a similar tendency to be volatile for small sample sizes, but they differ in that the second moment condition is not always satisfied.

Figures 14A and 14B show the values of $S_{GJR(N)}$ and $S_{GJR(t)}$, which are required to be less than unity by conditions (8) and (9), respectively. With S&P 500, both $S_{GJR(N)}$ and $S_{GJR(t)}$ are very volatile for small sample sizes, but much less volatile for samples with more than 2500 observations. Moreover, $S_{GJR(N)}$ does not exceed unity for most sample sizes, except for some sample sizes between 8000 and 9000 observations, while $S_{GJR(t)}$ exceeds unity for all sample sizes above 2500 observations. With Nikkei 225, there is also greater volatility in both $S_{GJR(N)}$ and $S_{GJR(t)}$ when the sample size is small, and significant increases in the values of $S_{GJR(N)}$ and $S_{GJR(t)}$ when there is a large volatility shock. The values of $S_{GJR(N)}$ and $S_{GJR(t)}$, however, both exceed unity for most sample sizes.

4.2 Forecasting Results

In order to evaluate the effects of increasing sample sizes and including extreme observations on the forecast performance of the GARCH(1,1) and GJR(1,1) models, similar backward recursions are used. For each model, two sets of forecasts are performed. In the first set of forecasts, the forecast period is the 250 trading days starting from 8 May 1997, which includes an extreme observation at 27 October 1997. Estimation of the parameters to obtain these forecasts is in the same manner as the backward recursions explained above, with sample sizes ranging from 200 observations to 5000 observations. For each sample size, 250 one-day ahead forecasts are made, covering the period 8 May 1997 to 5 May 1998 in the case of S&P 500, and the period 8 May 1997 to 22 April 1998 in the case of Nikkei 225. The prediction errors from these 250 forecasts are then combined in three measures of forecast performance, namely mean absolute prediction error (MAPE), mean absolute percentage prediction error (MAPPE), and root mean square prediction error (RMSPE).

The forecast performance measures of the GARCH(1,1) model for the 250 days starting from 8 May 1997 are graphed in Figures 15A to 17B. Not surprisingly, MAPE and MAPPE show very similar patterns. They both vary substantially for small sample sizes and both reach their respective minima at sample sizes below 2500. The effect of including the 19 October 1987 extreme observation is to increase both measures substantially, and then to stabilise at higher levels. This leads to the important and useful conclusion that expanding the sample size for estimation by including an extreme observation does not necessarily improve the accuracy of predicting future extreme observations. This conclusion applies to both the S&P 500 and Nikkei 225 data sets.

Figures 17A and 17B show that RMSPE is also highly volatile for small sample sizes. Again, the inclusion of the 19 October 1987 observation spike leads to a deterioration in forecast performance, especially in the case of Nikkei 225. There does, however, seem to be increased stability in forecast performance after the volatility spike.

The second set of forecasts is for the 250-trading day period ending on 7 May 1997, which does not contain any large spikes in the volatility of returns. The same backward recursion procedure is used for estimation, with the samples for estimation recurring backwards from 2

May 1996. The same three procedures for averaging of one-day ahead forecasts are also used to obtain the forecast performance measures.

Figures 18A and 18B show that MAPE for the GARCH(1,1) model is relatively stable and reaches its minimum at a small sample size, namely about 850 for S&P 500, and 350 for Nikkei 225. The inclusion of the 19 October 1987 extreme observation spike shifts up both trends and smoothes them, so that the inclusion of extreme observations in the estimation period does not necessarily help in prediction for a relatively calm period.

Figures 19A and 19B show that MAPPE reaches its minimum at the same sample size as the MAPE measure, but is more volatile, especially at sample sizes below 1000. Again, the effect of including the 19 October 1987 extreme observation in the estimation sample is to shift the measure up, more so in the case of Nikkei 225.

Figures 20A and 20B for RMSPE show a less consistent pattern, with RMSPE for S&P 500 and for Nikkei 225 reacting differently to the 19 October 1987 observation spike. With the inclusion of this extreme observation, RMSPE for S&P 500 stabilises and commences on a slight but clear downward trend, while RMSPE for Nikkei 225 shifts up and stabilizes at a relatively constant level.

Figures 21A to 23B show the forecast performances of the GJR(1,1) model for the forecast period beginning on 8 May 1997. All three measures achieve their respective minima at very low sample sizes of around 300 observations. These, however, are not indicative of the forecast performances over other small sample ranges. In fact, all three measures are highly volatile when sample sizes are below 2500 observations. The major impact of including the

19 October 1987 volatility spike on each of the three measures is to reduce the fluctuations significantly. The other important point to note is that the forecast performance measures stabilise at higher levels when the sample sizes increase beyond 2500 observations.

The forecast performances of the GJR(1,1) model for the relatively calm forecast period ending on 7 May 1997, as shown in Figures 24A to 26B, are also highly volatile for small sample sizes. The three measures for S&P 500 exhibit significantly different trends as sample sizes are increased. In Figures 24A and 25A, MAPE and MAPPE both reach their respective minima at around 1000 observations, while RMSPE in Figure 26A follows a general downward trend. The performance measures for Nikkei 225, however, are closer to each other.

While the 19 October 1987 volatility spike reduces the fluctuations of all three measures, it affects the trends differently. The inclusion of this extreme observation shifts up MAPE for both S&P 500 and Nikkei 225, as well as MAPPE for Nikkei 225, but not the other measures.

Comparing each forecasting performance graph for the GARCH(1,1) model in Figures 15A to 20B against the corresponding graph for the GJR(1,1) model in Figures 21A to 26B, it is not straightforward to determine which model is superior. In order to examine the results more closely, forecast performance measures for 250 one-day ahead forecasts, based on a fixed sample size of 5000, are given in Table 1[3]. Moreover, the RiskMetrics procedure[4] is used to calculate the corresponding forecast performance measures over the same period to serve as a benchmark.

[3] Comparisons of forecast performance based on sample sizes of 1000, 2000, 3000, and 4000 are also available, with the results being qualitatively similar to those reported in Table 1.
[4] Explained in Kim and Mina (2001), pp. 8-10.

[Insert Table 1 here]

The much smaller values in the GARCH(1,1) and GJR(1,1) columns in Table 1, as compared with the corresponding values in the RiskMetrics column, indicate that the GARCH(1,1) and GJR(1,1) models perform much better than the RiskMetrics benchmark. Comparing GARCH(1,1) against GJR(1,1), the two models perform equally well, with neither model being consistently better than the other. The GJR(1,1) model generally appears to perform better with the S&P 500 data, whereas the GARCH(1,1) model is better in some cases with Nikkei 225.

5 Concluding Remarks

This paper has investigated the effects of increasing sample sizes recursively, both with and without the inclusion of extreme observations, on the parameter estimates, t-tests, moment conditions and forecasts of the GARCH(1,1) and GJR(1,1) models, using S&P 500 and Nikkei 225 data. The results indicate that for these sets of data, the ARCH and GARCH parameter estimates, their asymptotic and robust t-ratios, the second and fourth moment regularity conditions, and various forecast performance measures for both models, are all highly volatile for small sample sizes. However, when an extreme observation is included in the estimation period for sample sizes above 2000, all the sample estimates and their associated statistics seem to stabilise. An important implication of these results is that increasing the sample sizes recursively beyond the extreme observation is unnecessary.

Another important result is that the robust t-ratios are dramatically superior to the asymptotic t-ratios, especially in the presence of high volatility in the returns. The second moment condition for stationarity is always satisfied for both the GARCH(1,1) and GJR(1,1) models in the case of S&P 500, but not so in the case of Nikkei 225. Similar results hold for the fourth moment condition for asymptotic normality under the assumption of normality of the conditional errors. For S&P 500, this condition is generally satisfied for both models, but for Nikkei 225, the same condition is usually violated. If it is assumed that the conditional error follows a fatter-tailed distribution such as $t(5)$, then the fourth moment condition is generally not satisfied for both models and both data sets when an extreme observation such as 19 October 1987 is included, regardless of the sample sizes used.

For most measures of forecasting performance, the inclusion of an extreme observation in the sample used for estimation leads to a marked deterioration in forecasting performance of both models, especially if the forecasting period is volatile. Increasing the sample sizes recursively does not necessarily improve the forecasting performance of either model. Both the GARCH(1,1) and GJR(1,1) models show superior forecasting performance to the RiskMetrics model. In choosing between the two models, however, superiority in forecasting performance depends on the data set used.

References

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

Bollerslev, T. and J.M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews*, 11, 143-172.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1801.

Franses, P.H. and H. Ghijsels (1999). Additive outliers, GARCH and forecasting volatility, *International Journal of Forecasting*, 15, 1-9.

Kim, J. and J. Mina (2001). *RiskGrades*^{TM} *Technical Document* [Online]. Available HTTP: http://www.riskmetrics.com.

Li, W.K., S. Ling and M. McAleer (2002). Recent theoretical results for time series models with GARCH errors, *Journal of Economic Surveys*, 16, 245-269.

Ling, S. and W.K. Li (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroskedasticity, *Journal of the American Statistical Association*, 92, 1184-1194.

Ling, S. and M. 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, S. and M. McAleer, (2002b), Asymptotic theory for a vector ARMA-GARCH model, to appear in *Econometric Theory*.

Ling, S. and M. McAleer (2002c). Stationarity and the existence of moments of a family of GARCH processes, *Journal of Econometrics*, 106, 109-117.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370.

Table 1: Comparison of forecast performance measures

Sample	Forecast Measure	Model		
		GARCH(1,1)	GJR(1,1)	RiskMetrics
S&P 500	**MAPE**	0.000139	0.000132	0.001950
8 May 1997 to 5 May 1998	**MAPPE**	289	251	3924
	RMSPE	0.000379	0.000375	0.001959
S&P 500	**MAPE**	0.000068	0.000068	0.001938
3 May 1996 to 7 May 1997	**MAPPE**	11879	11598	335336
	RMSPE	0.000113	0.000113	0.001941
Nikkei 225	**MAPE**	0.000318	0.000321	0.002181
8 May 1997 to 22 April 1998	**MAPPE**	467692	533934	2912579
	RMSPE	0.000587	0.000589	0.002215
Nikkei 225	**MAPE**	0.000172	0.000170	0.002179
23 May 1996 to 7 May 1997	**MAPPE**	40485	48686	834564
	RMSPE	0.000275	0.000268	0.002197

Note: Each model uses 5000 observations to obtain 250 one-day ahead forecasts covering the periods indicated. The prediction errors from these 250 forecasts are combined to form three measures of forecast performance, namely mean absolute prediction error (MAPE), mean absolute percentage prediction error (MAPPE), and root mean square prediction error (RMSPE).

Figure 1A: α Estimates of GARCH(1,1)
S&P 500

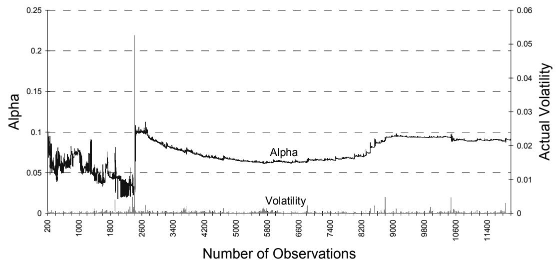

Figure 1B: α Estimates of GARCH(1,1)
Nikkei 225

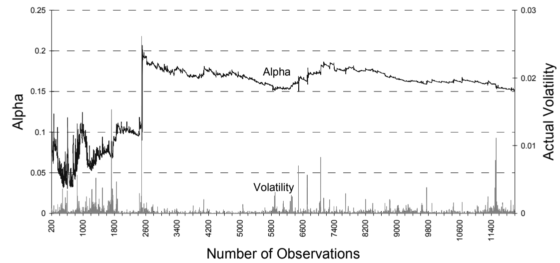

Figure 2A: α t-ratios of GARCH(1,1)
S&P 500

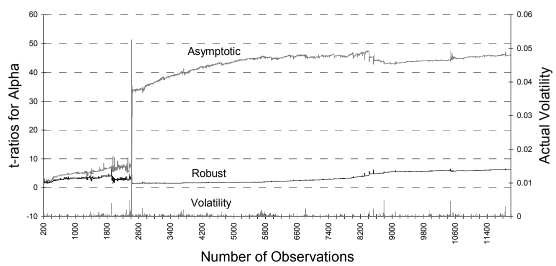

Figure 2B: α t-ratios of GARCH(1,1)
Nikkei 225

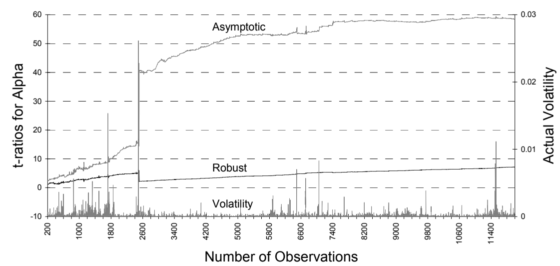

Figure 3A: β Estimates of GARCH(1,1)
S&P 500

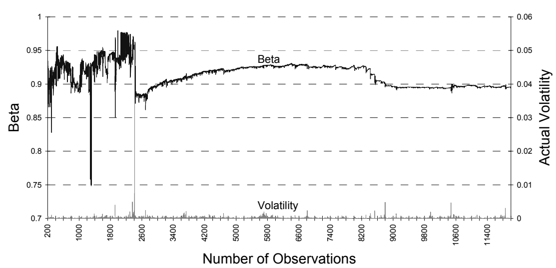

Figure 3B: β Estimates of GARCH(1,1)
Nikkei 225

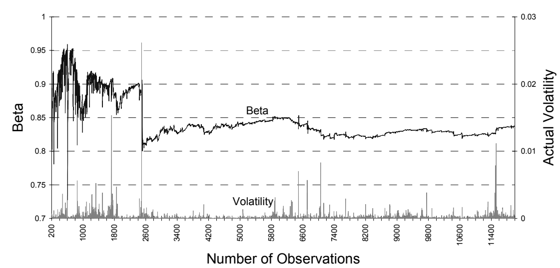

Figure 4A: β t-ratios of GARCH(1,1)
S&P 500

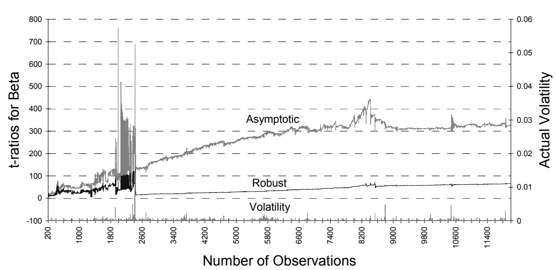

Figure 4B: β t-ratios of GARCH(1,1)
Nikkei 225

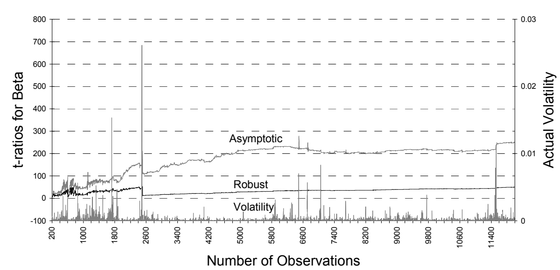

Figure 5A: Second Moments for GARCH(1,1)
S&P 500

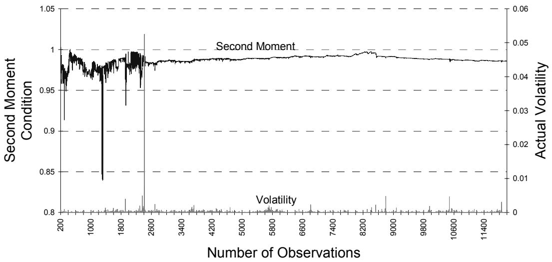

Figure 5B: Second Moments for GARCH(1,1)
Nikkei 225

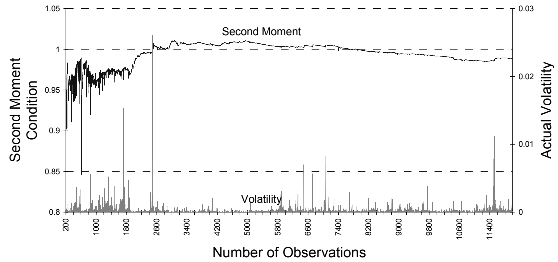

Figure 6A: Fourth Moments for GARCH(1,1)
S&P 500

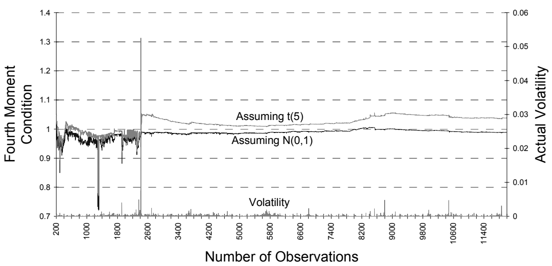

Figure 6B: Fourth Moments for GARCH(1,1)
Nikkei 225

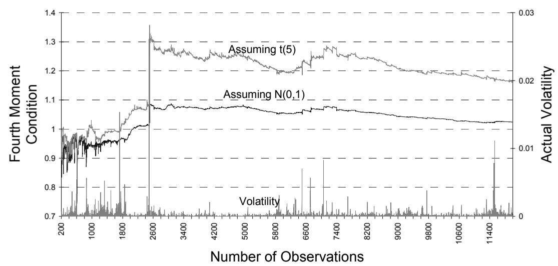

Figure 7A: α Estimates of GJR-GARCH(1,1)
S&P 500

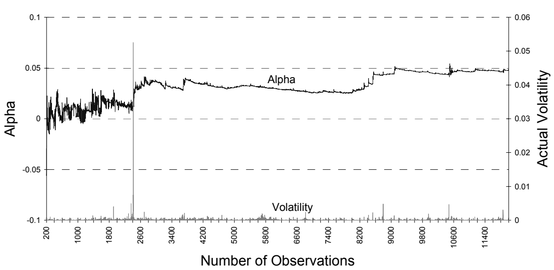

Figure 7B: α Estimates of GJR-GARCH(1,1)
Nikkei 225

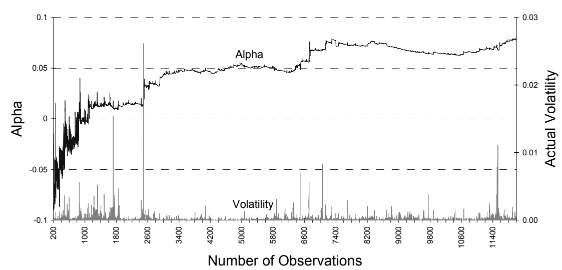

Figure 8A: α t-ratios of GJR-GARCH(1,1)
S&P 500

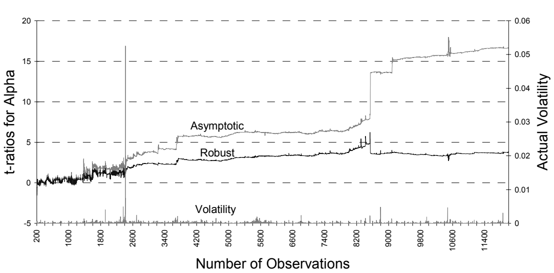

Figure 8B: α t-ratios of GJR-GARCH(1,1)
Nikkei 225

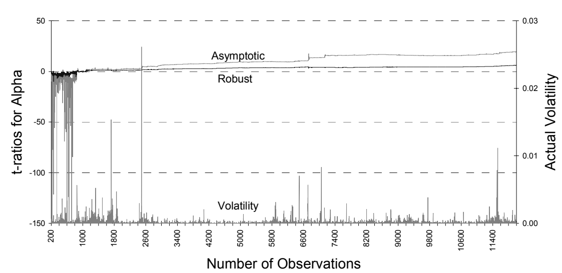

Figure 9A: β Estimates of GJR-GARCH(1,1)
S&P 500

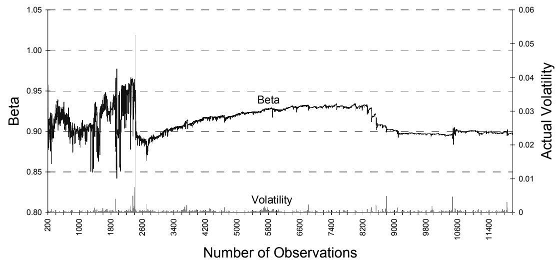

Figure 9B: β Estimates of GJR-GARCH(1,1)
Nikkei 225

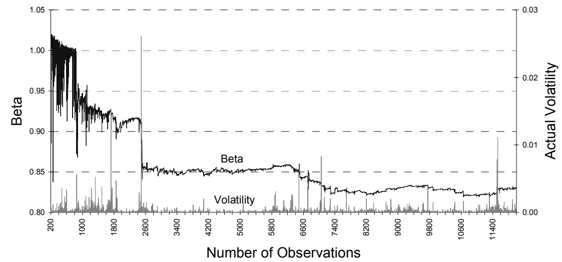

Figure 10A: β t-ratios of GJR-GARCH(1,1)
S&P 500

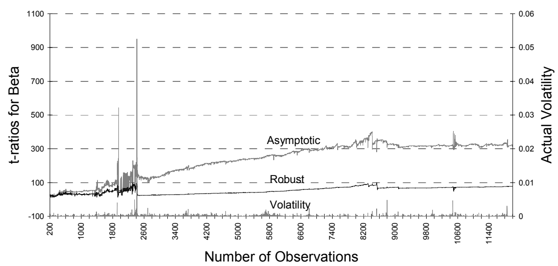

Figure 10B: β t-ratios of GJR-GARCH(1,1)
Nikkei 225

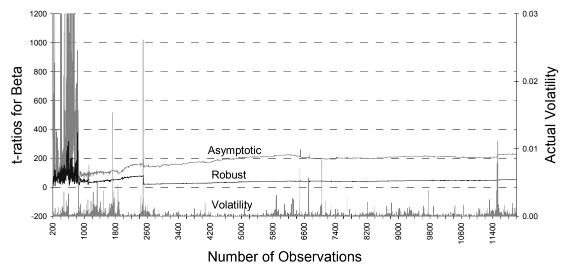

Figure 11A: γ Estimates of GJR-GARCH(1,1)
S&P 500

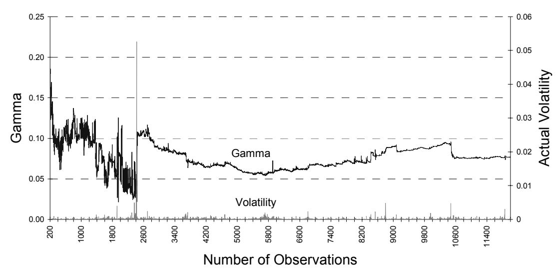

Figure 11B: γ Estimates of GJR-GARCH(1,1)
Nikkei 225

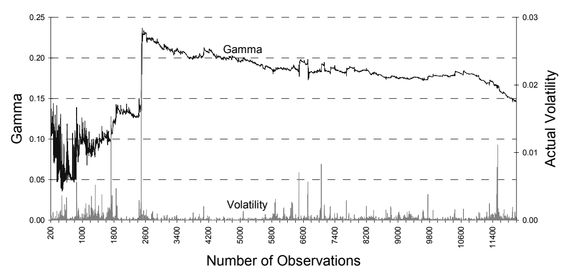

Figure 12A: γ t-ratios of GJR-GARCH(1,1)
S&P 500

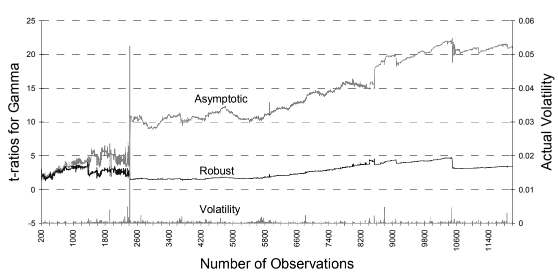

Figure 12B: γ t-ratios of GJR-GARCH(1,1)
Nikkei 225

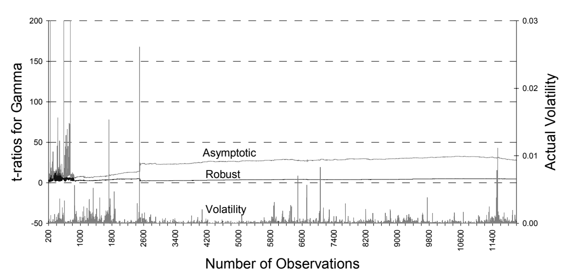

Figure 13A: Second Moments for GJR-GARCH(1,1)
S&P 500

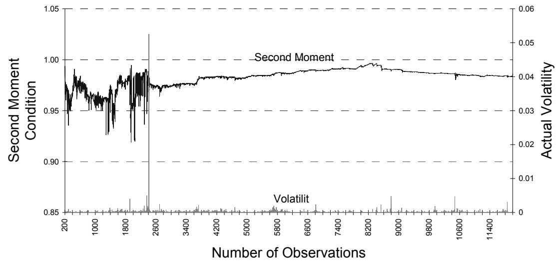

Figure 13B: Second Moments for GJR-GARCH(1,1)
Nikkei 225

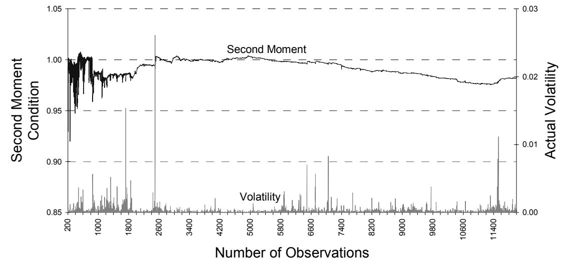

Figure 14A: Fourth Moments for GJR-GARCH(1,1)
S&P 500

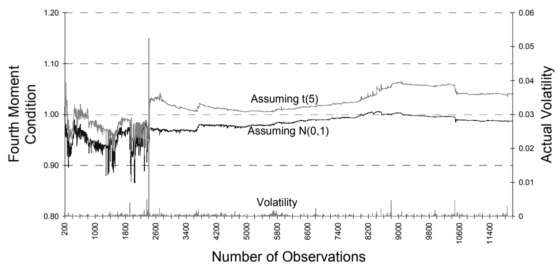

Figure 14B: Fourth Moments for GJR-GARCH(1,1)
Nikkei 225

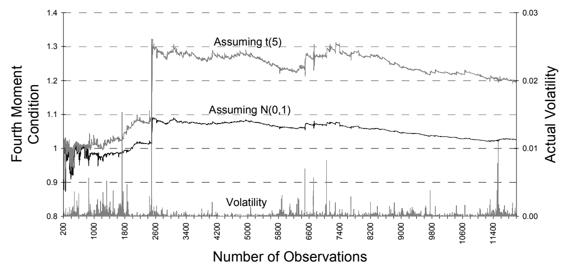

Figure 15A: GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

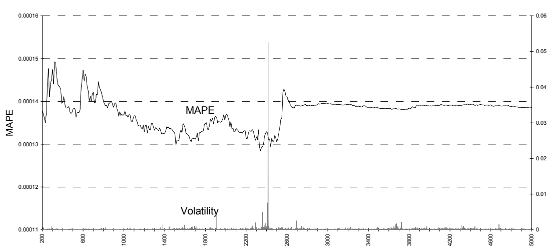

Figure 15B: GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

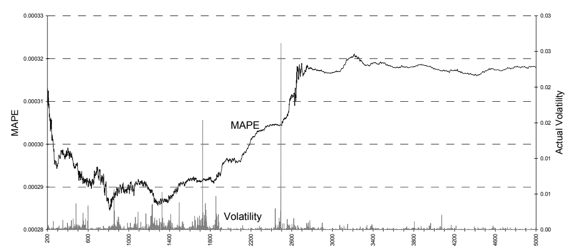

Figure 16A: GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

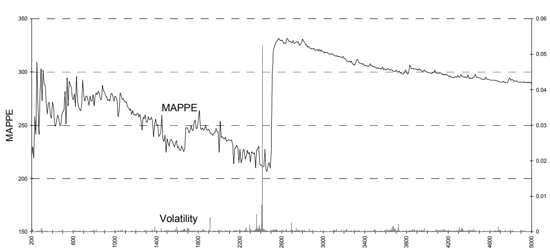

Figure 16B: GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

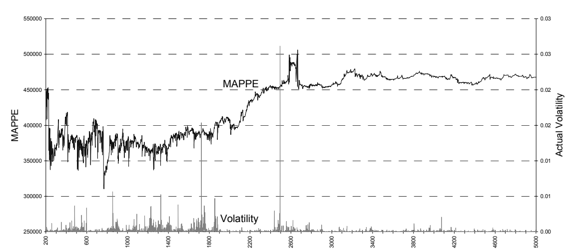

Figure 17A: GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

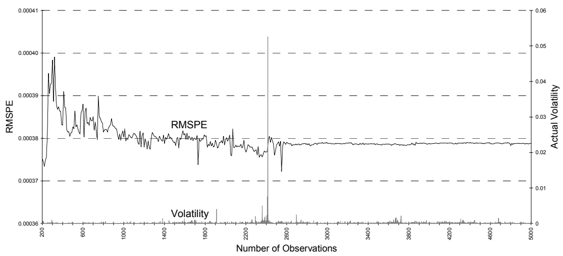

Figure 17B: GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

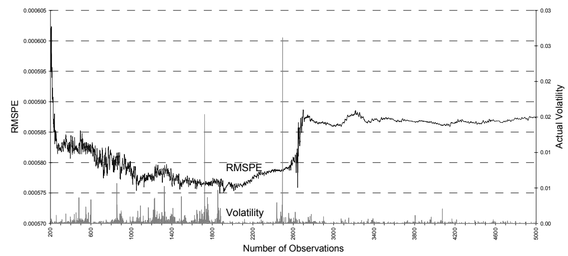

Figure 18A: GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

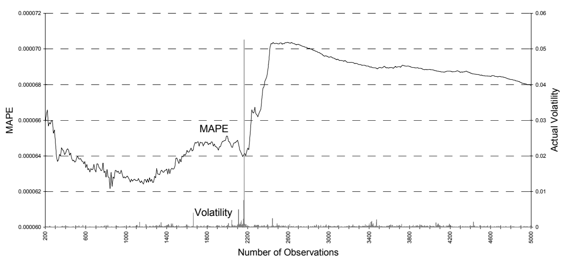

Figure 18B: GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225

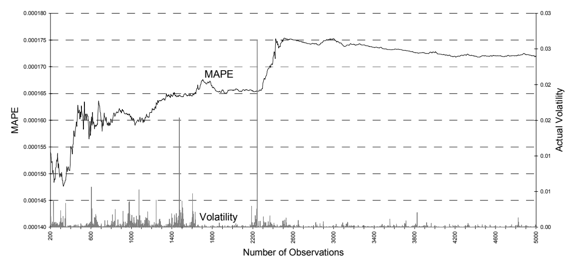

Figure 19A: GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

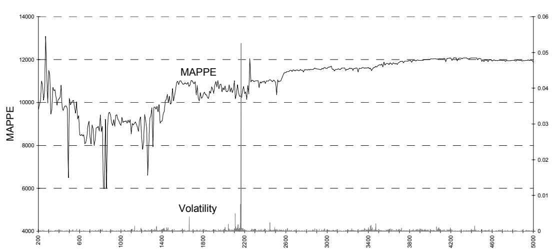

Figure 19B: GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225

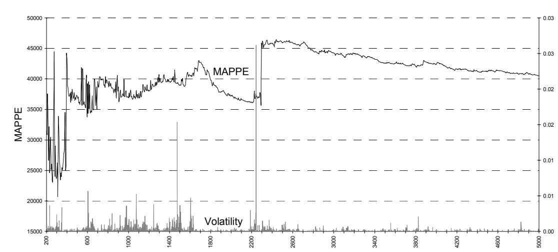

Figure 20A: GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

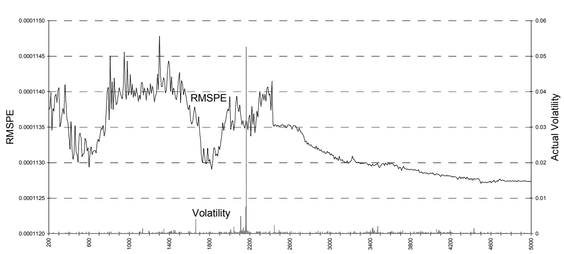

Figure 20B: GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225

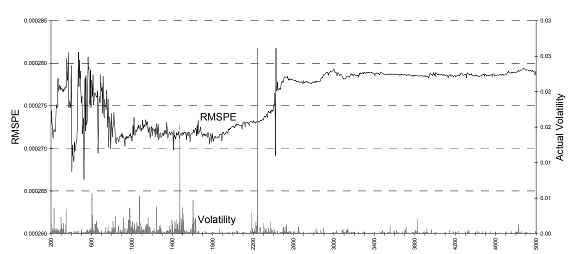

Figure 21A: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

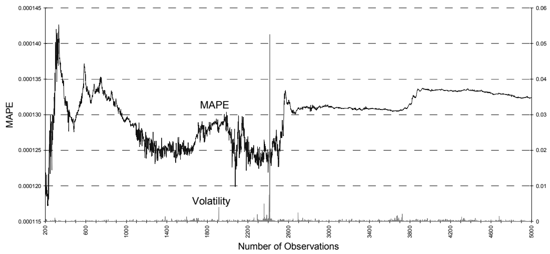

Figure 21B: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

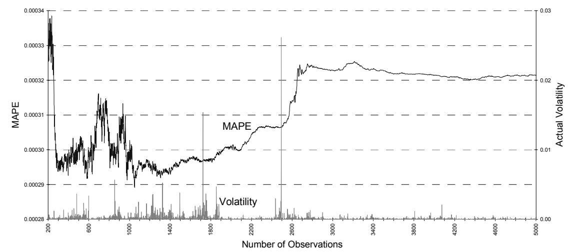

Figure 22A: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

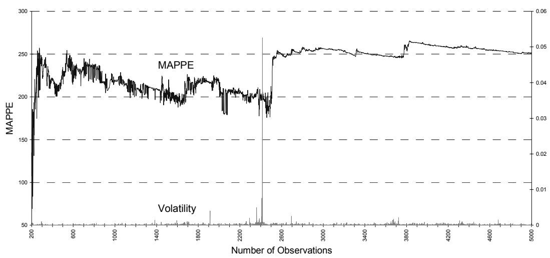

Figure 22B: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

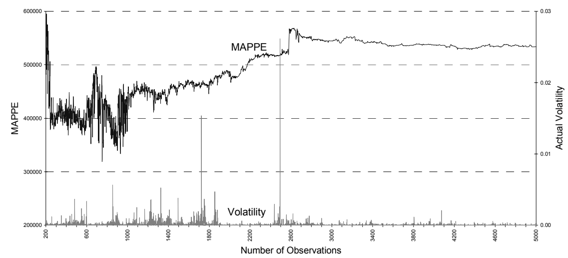

Figure 23A: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 5/5/98 (including October 1997)
S&P 500

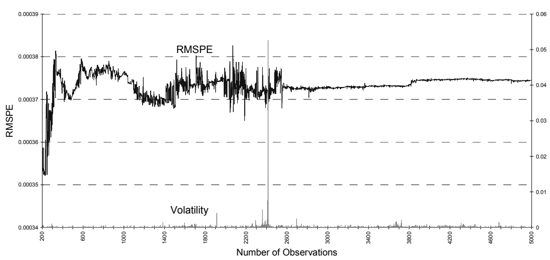

Figure 23B: GJR-GARCH(1,1) Forecast Performance
for 8/5/97 to 22/4/98 (including October 1997)
Nikkei 225

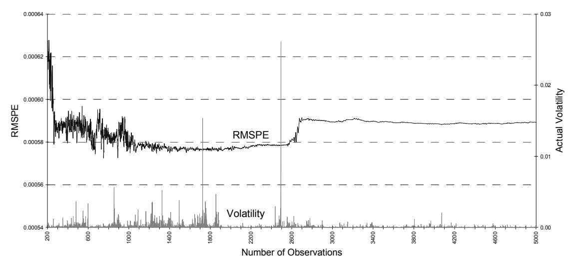

Figure 24A: GJR-GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

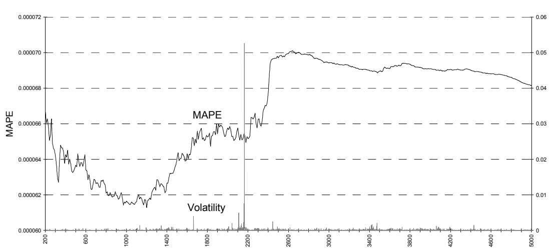

Figure 24B: GJR-GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225

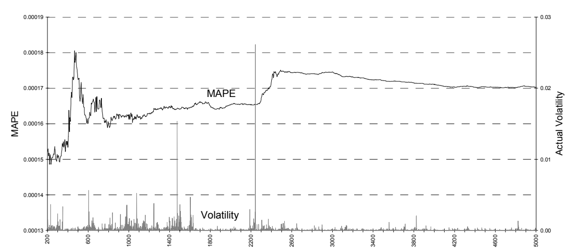

Figure 25A: GJR-GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

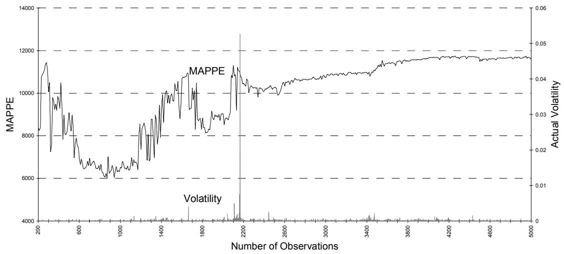

Figure 25B: GJR-GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225

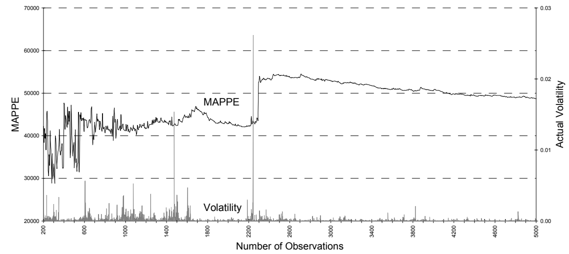

Figure 26A: GJR-GARCH(1,1) Forecast Performance
for 3/5/96 to 7/5/97 (excluding October 1997)
S&P 500

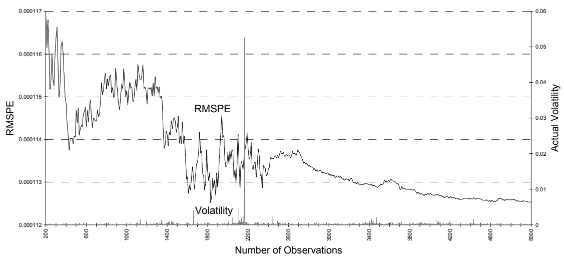

Figure 26B: GJR-GARCH(1,1) Forecast Performance
for 23/5/96 to 7/5/97 (excluding October 1997)
Nikkei 225