Recursive Modelling of Volatility in the Presence of Extreme Observations

Hock Guan Ng[1] and Michael McAleer[2]

[1]Department of Accounting and Finance, University of Western Australia
[2]Department of Economics, University of Western Australia

Abstract This paper is concerned with recursive estimation, testing and forecasting of the volatility of daily returns in the Standard and Poor's 500 Composite Index (S&P 500) in the presence of extreme observations, or significant spikes in the volatility of daily returns. The empirical analysis increases the sample size up to 12000 observations recursively to examine the effects of extreme observations on: (i) the parameter estimates of the GARCH(1,1) process; (ii) their associated asymptotic and robust t-ratios; (iii) the second and fourth moment conditions for stationarity, consistency and asymptotic normality; and (iv) the forecast performance for periods with significant spikes in volatility and for periods of relative calm.

1 INTRODUCTION

The modelling of volatility has been a very active area of research in finance in recent years, and has been largely motivated by the importance of risk considerations in economic and financial markets. Estimates of volatility are used widely for a variety of reasons, including modelling the premium in forward and futures markets, portfolio selection, asset management, pricing primary and derivative assets, valuation of warrants and options, designing optimal hedging strategies for options and futures markets, evaluating risk spill-overs across markets, measuring announcement effects in event studies, and examining asymmetries and leverage effects.

Engle (1982) first captured the time-varying nature of volatility with the autoregressive conditional heteroscedasticity (ARCH(p)) model. The ARCH model was generalized to GARCH(p,q) by Bollerslev (1986), and this has proved to be the single most popular time-varying volatility model in practice. GARCH has two quite attractive features, namely the persistence of volatility, and mathematical and computational simplicity. Many theoretical results, including the statistical properties of the model and the asymptotic properties of several estimation methods, are now available, and these provide a solid foundation for applications of the model (see Li et al. (1999) for a survey, directed towards practitioners, of recent important theoretical results for GARCH models).

A common feature encountered in high frequency financial time series is the occurrence of extreme observations, or significant spikes in volatility, which can adversely affect the estimates and forecasts of volatility. This paper is concerned with recursive estimation, testing and forecasting of the volatility of daily returns in the Standard and Poor's 500 Composite Index (S&P 500) in the presence of extreme observations in the volatility of daily returns. The empirical analysis increases the sample size up to 12000 observations recursively to examine the effects of extreme observations on: (i) the parameter estimates of the GARCH(1,1) process; (ii) their associated asymptotic and robust t-ratios; (iii) the second and fourth moment conditions for stationarity, consistency and asymptotic normality; and (iv) the forecast performance for periods with significant spikes in volatility and for periods of relative calm.

Several interesting results emerge from the analysis, namely: expanding the sample sizes 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) process, the 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, so that increasing the sample size recursively beyond the extreme observation is unnecessary; the robust t-ratios are, in general, dramatically superior to the asymptotic t-ratios; and the second and fourth moment conditions are generally satisfied, but increasing the sample sizes recursively does not necessarily help to satisfy these conditions, or to improve the forecasting performance.

The plan of the paper is as follows. Section 2 presents the AR(1)-GARCH(1,1) model. Section

3 describes the data. The empirical results are analysed in Section 4. Some concluding remarks are given in Section 5.

2 THE AR(1)-GARCH(1,1) MODEL

The model to be estimated is AR(1)-GARCH(1,1), where the conditional mean (or log returns of the S&P 500 Index) has the structure given by

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \tag{1}$$

and the conditional variance of the unconditional shock ε_t is generated 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 for the GARCH process to exist are that $\omega>0$, $\alpha>0$, and $\beta \geq 0$.

Several statistical properties have been established for the GARCH(1,1) process in order to define the unconditional moments of ε_t. First, the second moment of ε_t exists if $\alpha + \beta < 1$, which ensures that the GARCH(1,1) process is strictly stationary and ergodic, and $E\varepsilon_t^2 < \infty$ (see Bollerslev (1986) and Ling and Li (1997)). Second, the 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 $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$.

For the GARCH(1,1) model, Nelson (1990) obtained the necessary and sufficient condition for strict stationarity and ergodicity as:

$$E(\ln(\alpha\eta_t^2 + \beta)) < 0. \tag{4}$$

A difficulty in applying the necessary and sufficient condition in (4) is that it needs to be simulated. Condition (4) allows $\alpha + \beta$ to be unity, or slightly greater than unity, 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 the condition for a finite fourth moment is $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$. The fourth moment condition is clearly the most stringent.

3 DATA

The daily closing values of the Standard and Poor's 500 Composite Index for the period 3 January 1950 to 31 March 1999 were extracted from the Datastream database, and the daily return was calculated as the ratio of the close-to-close change in the index to the previous trading day's close.

The long period selected includes many significant spikes in the volatility of daily returns, the largest of which occurred on 19 October 1987, and also many episodes of relative calm. Consequently, this data set offers an opportunity to study the effects of extreme observations on the estimation and forecasting of volatility over an extended period.

Various subsets of the data are used for estimation and forecasting. To evaluate the effects of extreme observations on estimation, 12000 observations from the period 3 January 1950 to 7 May 1997 are used. For the evaluation of forecasting performance, two separate "out-of-sample" periods consisting of 250 observations each are used: the first of these, from 8 May 1997 to 5 May 1998, includes some significant spikes in the volatility of daily returns, while the second, from 13 May 1996 to 7 May 1997, is a relatively calm period.

4 EMPIRICAL RESULTS

In order to evaluate the effects of increasing sample sizes and including extreme o bservations, the AR(1)-GARCH(1,1) model is estimated recursively. In the first set of results, the sample starts with 200 observations from 3 January 1950 to 21 November 1950. The sample is then expanded recursively through to the end of the sample until it reaches 12000 observations at 7 May 1997.

Figure 1 gives the estimated values of the ARCH parameter α as the sample size is increased recursively. The actual volatility of the daily returns is shown in the lower half of the figure to indicate where the volatility spikes occur. It is clear that the estimates of α are highly volatile when the sample size is below 2000. Significant spikes in the actual volatility correspond to huge variations in the estimates of α. When the sample size exceeds 4000, however, the variations in α estimates become less pronounced and the estimates begin a downward trend. The huge spike in volatility on 19 October 1987 caused a shift upwards in the trend, but this shift is relatively small compared with the variations for low sample sizes.

Figure 2 presents the asymptotic t-ratios and the robust t-ratios of Bollerslev and Wooldridge (1992). The robust t-ratios are designed to be insensitive to departures from normality, especially extreme observations. Both sets of t-ratios are somewhat erratic at sample sizes below

2000, but subsequently follow a relatively smooth upward trend. At sample sizes above 6000, the robust t-ratios change little, while the asymptotic t-ratios continue increasing. The effects of significant spikes in volatility on the two t-ratios are 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.

Estimates of the GARCH parameter β are given in Figure 3. This is virtually a mirror image of the estimates of α, with the β estimates moving in the opposite direction to the α estimates. There is also much variability in the β estimates at sample sizes below 4000 and subsequently a relatively smooth upward trend. The spikes in volatility also have larger impacts on the β estimates when the sample size is small.

Figure 4 shows the t-ratios for the β estimates. At sample sizes below 2000, there is substantial variability in the t-ratios, especially the asymptotic t-ratios. At large sample sizes, the asymptotic t-ratios have a significant upward trend while the robust t-ratios are much flatter.

The second moment condition for stationarity and consistency, $\alpha + \beta < 1$, and fourth moment condition for asymptotic normality, $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$, are given in Figures 5 and 6, respectively. These two graphs are almost identical in pattern. Again, spikes in the volatility of returns have large impacts on the conditions when the sample size is below 5000. When the sample size is very large, even a huge spike such as that of 19 October 1987 has a relatively small impact on the conditions. It is significant to note that both conditions are satisfied for all sample sizes in the forward recursions.

The second set of estimates is presented for backward recursions, with the end observation fixed at 7 May 1997. The sample begins with 200 observations from 23 July 1996 to 7 May 1997, and is then expanded recursively until it reaches 12000 observations at 3 January 1950.

Figure 7 presents the α estimates from these backward recursions. Spikes in volatility have huge impacts on the α estimates for small sample sizes. The most obvious feature of the swings is the huge shift in α estimates with the 19 October 1987 spike in volatility, after which the variations in the α estimates are much smaller in magnitude.

The graph of the t-ratios for the α estimates in Figure 8 shows that extreme observations have much greater impacts before the inclusion of the 19 October 1987 spike. Of note is the huge increase in the asymptotic t-ratios when this particular spike is included. In contrast, the impact of this extreme observation on the robust t-ratios is barely visible.

As in the forward recursions, the graph of the β estimates, as shown in Figure 9, is close to a mirror image of Figure 7, with some subtle differences. The impact of the 19 October 1987 observation is relatively smaller on the β estimates. At smaller sample sizes there are some changes in the β estimates that are much larger than the corresponding changes in the α estimates. This suggests that β estimates are more sensitive in smaller samples.

The sensitivity of β estimates in small samples is also reflected in Figure 10. Both the asymptotic and robust t-ratios show great 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.

In comparing the forward and backward recursions, the robust t-ratios for both α and β estimates show a slight upward trend at large sample sizes in the backward recursions, but are flatter in the forward recursions.

The second and fourth moment conditions for stationarity, consistency and asymptotic normality for the backward recursions are shown in Figures 11 and 12. Both the second and fourth moment conditions show high variations prior to the inclusion of the 19 October 1987 observation spike, and much smaller variations subsequently. While the second moment condition holds for all samples, the fourth moment condition is violated in some sample ranges. Some of these violations occur for relatively large samples (at around 8000 observations).

To evaluate the effects of increasing sample sizes and including extreme observations on the forecast performance of the GARCH(1,1) model, backward recursions are used. In the first set of forecasts, the forecast period is from 8 May 1997 to 5 May 1998, 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 for the period 8 May 1997 to 5 May 1998. The prediction errors from these 250 forecasts are then combined in the three measures of forecast performance, namely mean absolute prediction error (MAPE), mean absolute percentage prediction error (MAPPE), and root mean square prediction error (RMSPE). These measures are graphed in Figures 13 to 15.

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 conclusion that expanding the sample size for estimation by including an extreme observation does not necessarily improve the accuracy of predicting future extreme observations.

Figure 15 shows that RMSPE is also highly volatile for small sample sizes, but it has a decreasing trend as the sample size increases. Again, the inclusion of the 17 October 1987 observation spike does not lead to an improvement in forecast performance, but there does seem to be increased stability.

The second set of forecasts is for the period 13 May 1996 to 7 May 1997, which does not contain any large spikes in the volatility of returns. As before, the same backward recursion procedure and averaging of one-day ahead forecasts is used to obtain the forecast performance measures. Figures 16 and 17 show that MAPE and MAPPE are relatively stable and show mild U-shaped patterns between 200 and 2000 observations, with the minima occurring at about 1200 observations. The inclusion of the 17 October 1987 extreme observation spike shifts up both trends and smooths them, so that the inclusion of extreme observations in the estimation period does not necessarily help in prediction for a relatively calm period.

Figure 18 for RMSPE, however, shows a different picture, with RMSPE being rather volatile for small sample sizes before the 17 October 1987 observation spike is included. With the inclusion of this extreme observation, RMSPE stabilises and starts on a slight but clear downward trend.

5 CONCLUDING REMARKS

This paper has investigated the effects of increasing the sample sizes recursively, both with and without the inclusion of extreme observations, on the parameter estimates, t-tests and forecasts of the GARCH(1,1) model. The results indicate that the ARCH and GARCH parameter estimates, their asymptotic and robust t-ratios, the second and fourth moment regularity conditions, and various forecast performance measures, are all highly volatile for small sample sizes. However, when an extreme observation is included in the estimation period at sample sizes above 2000, all the sample estimates and their associated statistics 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 and fourth moment conditions for stationarity, consistency, and asymptotic normality are generally satisfied, but increasing the sample sizes recursively does not necessarily help to satisfy these conditions. Moreover, increasing the sample sizes recursively does not necessarily improve the forecasting performance.

6 ACKNOWLEDGEMENT

The second author is grateful for the financial support of the Australian Research Council.

7 REFERENCES

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

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

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

Li, W.K., S. Ling and M. McAleer, A survey of recent theoretical results for time series models with GARCH errors (submitted).

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

Nelson, D.B., Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370, 1989.

Figure 1: Estimation Using Forward Recursions

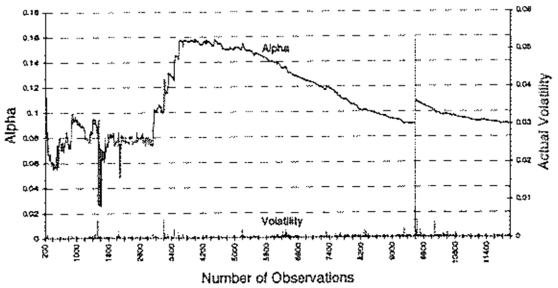

Figure 2: Estimation Using Forward Recursions

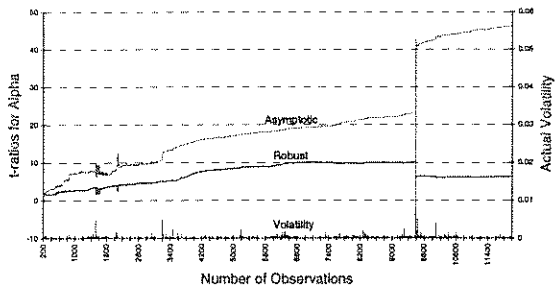

Figure 3: Estimation Using Forward Recursions

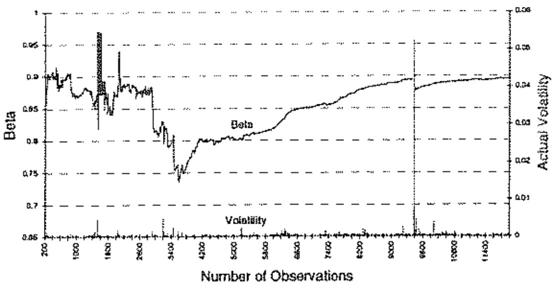

Figure 4: Estimation Using Forward Recursions

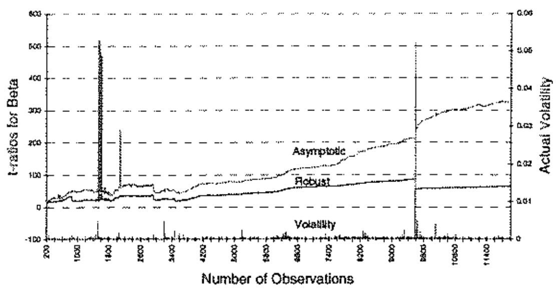

Figure 5: Estimation Using Forward Recursions

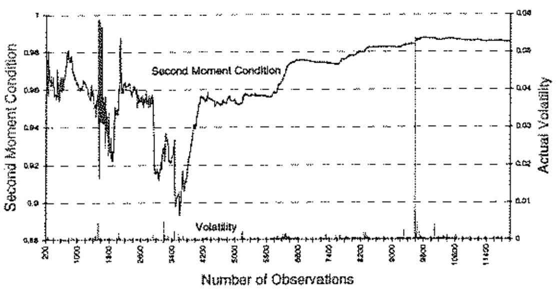

Figure 6: Estimation Using Forward Recursions

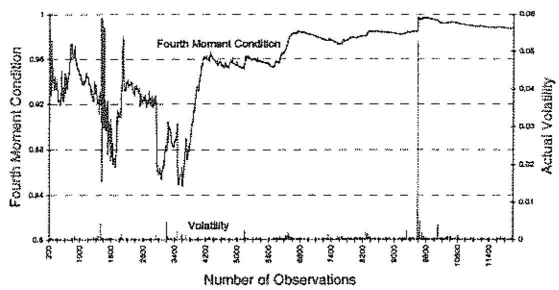

Figure 7: Estimation Using Backward Recursions

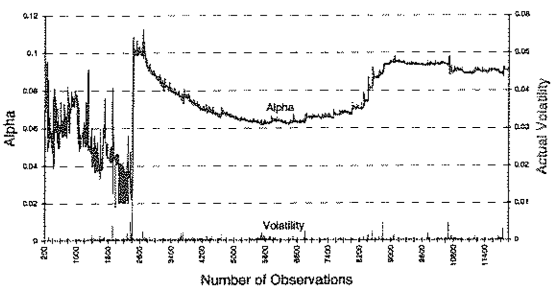

Figure 8: Estimation Using Backward Recursions

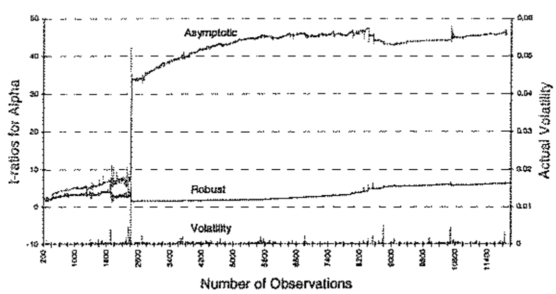

Figure 9: Estimation Using Backward Recursions

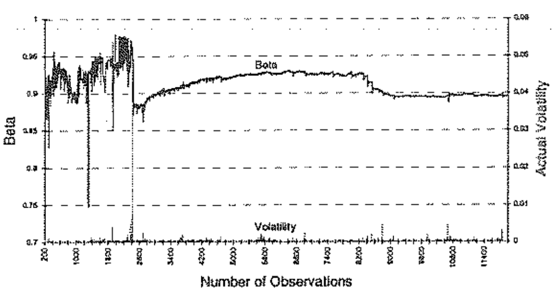

Figure 10: Estimation Using Backward Recursions

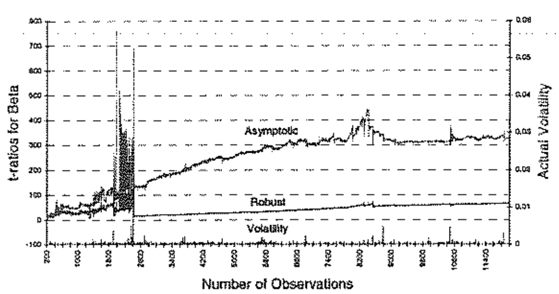

Figure 11: Estimation Using Backward Recursions

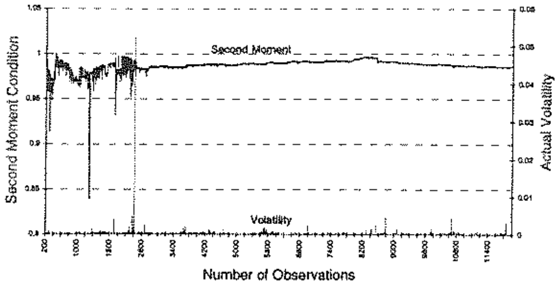

Figure 12: Estimation Using Backward Recursions

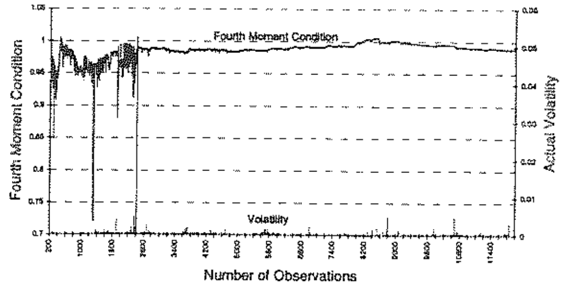

Figure 13: Forecast Performance Based on Backward Recursions for 8/5/97 to 5/5/98 (including October 1997)

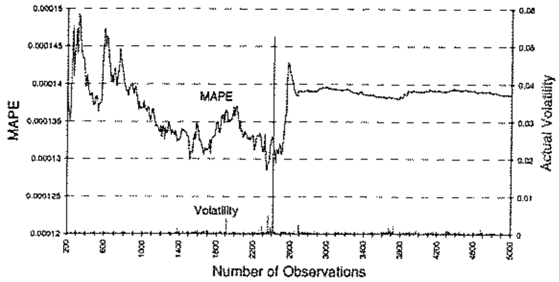

Figure 14: Forecast Performance Based on Backward Recursions for 8/5/97 to 5/5/98 (including October 1997)

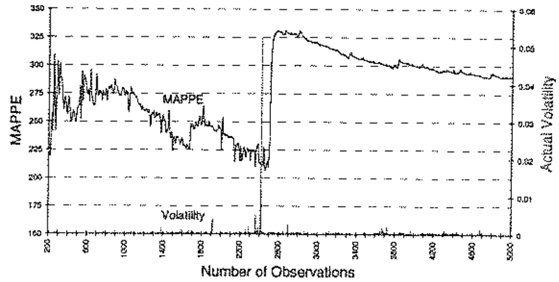

Figure 15: Forecast Performance Based on Backward Recursions for 8/5/97 to 5/5/98 (including October 1997)

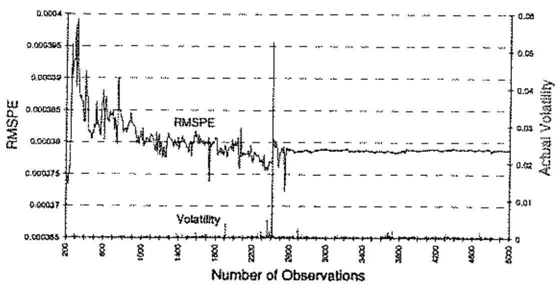

Figure 16: Forecast Performance Based on Backward Recursions for 13/5/96 to 7/5/97 (excluding October 1997)

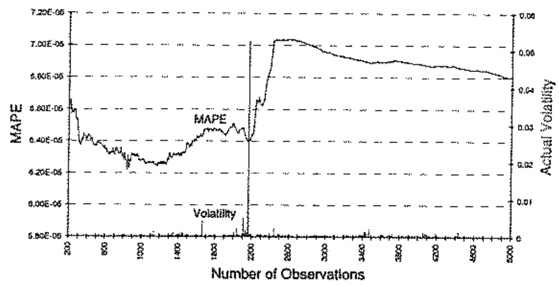

Figure 17: Forecast Performance Based on Backward Recursions for 13/5/96 to 7/5/97 (excluding October 1997)

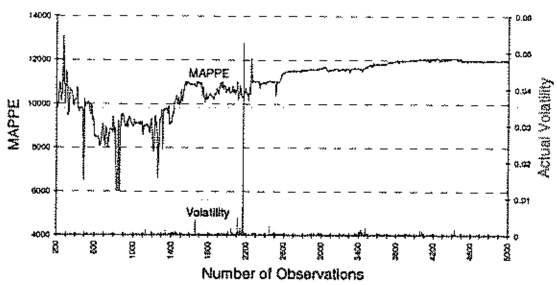

Figure 18: Forecast Performance Based on Backward Recursions for 13/5/96 to 7/5/97 (excluding October 1997)