FORECASTING VOLATILITY IN THE S&P 500 INDEX – AN EMPIRICAL TEST OF OPTION MARKET EFFICIENCY[1]

Summary

While tests of the efficiency of underlying asset markets have been the focus of a very considerable body of empirical research, it is only recently that attention has turned to examining the behavior and efficiency of asset volatility markets. Views on asset volatility are typically expressed in options markets, with the S&P 500 Index options being amongst the most liquid: around 100,000 contracts trade daily, with around 4,000,000 contracts in open interest. Hull and White (1987) show that, when volatility is constant, the Black-Scholes implied volatility of an at-the-money option approximately equals the expected future volatility over the life of the option. Consequently, assuming options markets are efficient, the option implied volatility is likely to be an unbiased predictor of future market volatility. To the extent that options markets are efficient, implied volatility should be a superior predictor of future volatility than any forecasting model based on widely known theory and using historical volatility data.

In this analysis we are concerned with the issue of whether market forecasts of volatility, as expressed in the Black-Scholes implied volatilities of at-the-money European options on the S&P500 Index, are superior to those produced by a new forecasting model in the GARCH framework which incorporates long-memory effects. The ARFIMA-GARCH model, which uses high frequency data comprising 5-minute returns, makes volatility the subject process of interest, to which innovations are introduced via a volatility-of-volatility (kurtosis) process. Despite performing robustly in- and out-of-sample, an encompassing regression indicates that the model is unable to add to the information already contained in market forecasts. However, unlike model forecasts, implied volatility forecasts show evidence of a consistent and substantial bias. Furthermore, the model is able to correctly predict the direction of volatility approximately 62% of the time whereas market forecasts have very poor direction prediction

[1] The author wishes to acknowledge the contribution to the research in this chapter due to the many helpful comments and suggestions from Professor James Davidson of the University of Exeter. The analysis was conducted using Professor Davidson's Time Series Modelling 4.05.

ability. This suggests that either option markets may be inefficient, or that the option pricing model is mis-specified. To examine this hypothesis, an empirical test is carried out in which at-the-money straddles are bought or sold (and delta-hedged) depending on whether the model forecasts exceed or fall below implied volatility forecasts. This simple strategy generates an annual compound return of 18.64% over a four year out-of-sample period, during which the annual return on the S&P index itself was -7.24%. Our findings suggest that, over the period of analysis, investors required an additional risk premium of 88 basis points of incremental return for each unit of volatility risk.

Introduction

There have been a number of studies examining the relationship between implied and future realized volatility, with very mixed findings with regard to the central question of market efficiency. Camina and Figlewski (1993) find that the implied volatility of S&P 100 index options contain no information about future volatility, suggesting that the options market may be inefficient. However, Christensen and Prabhala (1998) point out that the use of overlapping periods to estimate historical volatility introduces serious autocorrelation bias because the volatility estimates for two consecutive periods share almost all of the same information. Their findings are consistent with the interpretation provided by Figlewski (1997) that investors may be making optional forecasts of future volatility, but that these forecasts are not reflected in option prices due to market frictions. Fleming (1998) corrects for serial dependence and goes on to find that the implied volatility from S&P100 index options produces superior forecasts of future realized volatility compared with forecasts made using historical volatility.

Taylor and Xu (1997) examine a parallel question in the foreign exchange markets, comparing the forecasting ability of implied and historical volatility for the Deutschemark/Dollar exchange rate. They find that when daily observations are used to construct the data series implied volatility forecasts are superior. However, when intra-day 5-minute returns are used, historical volatility outperforms implied volatility in forecasting realized volatility. In a more extensive study, Pong, Shakelton, Taylor and Xu (2003) compare the performance of implied volatility forecast against a number of short- and long-memory ARMA and GARCH models for Pound, D-Mark and Yen exchange rates against the US-Dollar. Their tests indicate that model forecasts have incremental information not found in implied volatilities, for forecast horizons of up to one week. By contrast, implied volatilities are found to incorporate most of the relevant historical information when the horizon extends to one month or longer. On the other hand, Blair, Poon and Taylor (2001) claim almost all the useful predictive information is in option prices when forecasting S&P 100 index volatility one or more days into the future. Corrado and Miller (2003) confirm those findings and similarly conclude that implied volatility indices dominate historical index volatility in providing forecasts of future volatility in the S&P 100 and Nasdaq 100 indices.

In general, the use of high-frequency data has led to a dramatic improvement in both measuring and forecasting volatility. Anderson et al (2001) use a fractionally integrated autoregressive process to capture the long memory effects in volatility and forecast realized or integrated volatility, defined as the summation of high-frequency squared returns. Taylor (2000) compares the implied volatilities for options on the S&P 100 Index derived from short and long memory (Fractionally Integrated Exponential GARCH) models and finds that they differ by more than 1% and consequently that the economic consequences of a long memory assumption are important. Martens and Zein (2002) find that high-frequency forecasts do have incremental information over that contained in implied volatilities for futures and options on futures for the S&P 500 Index.

Harvey and Whaley (1992) take a different approach to assessing option market efficiency. They test the profitability of a trading rule based on forecasting implied volatility, concluding in favor of option market efficiency after allowing for transaction costs. Noh and Engle (1994), by contrast, find that a simple strategy involving trading straddles on the S&P 500 Index based on GARCH model volatility forecasts can make significant profits after transaction costs.

The purpose of this study is to examine the question of option market efficiency for the S&P 500 Index from two perspectives. The first objective is to discover whether forecasting models using high-frequency data and incorporating both long- and short-term memory effects are capable of outperforming implied volatility in forecasting realized volatility. Secondly, an analysis is made of the potential for generating abnormal profits using a simple straddle trading strategy, based on model-driven volatility forecasts, taking into account both transaction and (delta) hedging costs.

Data and Sampling Procedure

Tick by tick data for the S&P 500 cash index from February 1983 to December 2003 was collected from multiple data feeds via a commercial data vendor and subjected to scrubbing methodologies to detect and repair bad ticks. Estimates of the frequency of bad ticks vary. Dacorogna et al. (1995) estimate that error rates on forex quote data are between 0.11% and 0.81%. Lundin et al. (1999) describe the use of filters in preprocessing forex, stock index, and implied forward interest rate returns whereby 2%–3% of all data points were identified as false outliers.

Following cleaning, the data were then aggregated at five minute intervals and used to compute approximately 5,000 observations of daily realized variance as per Andersen, Bollerslev, Diebold and Ebens, 2000. The realized variance data series was Winsorized at six standard deviations to remove outliers and then subjected to log transformation to provide the base time series used for modeling purposes. While the log variance series is approximately Normally distributed, it does exhibit positive skewness (0.23) and excess kurtosis (3.29) when compared to the theoretical distribution. These small but statistically significant departures from Normality are detected by the standard Chi-squared, Anderson-Darling and Kolmagorov-Smirnov tests, all of which reject the null hypothesis of Normality at the 5% level.

In common with parallel research, we find evidence of long memory in the series of daily realized volatility observations, as evidenced by the slowly declining pattern of autocorrelations with significant autocorrelations out as far as 100 days, as illustrated in Figure 1.

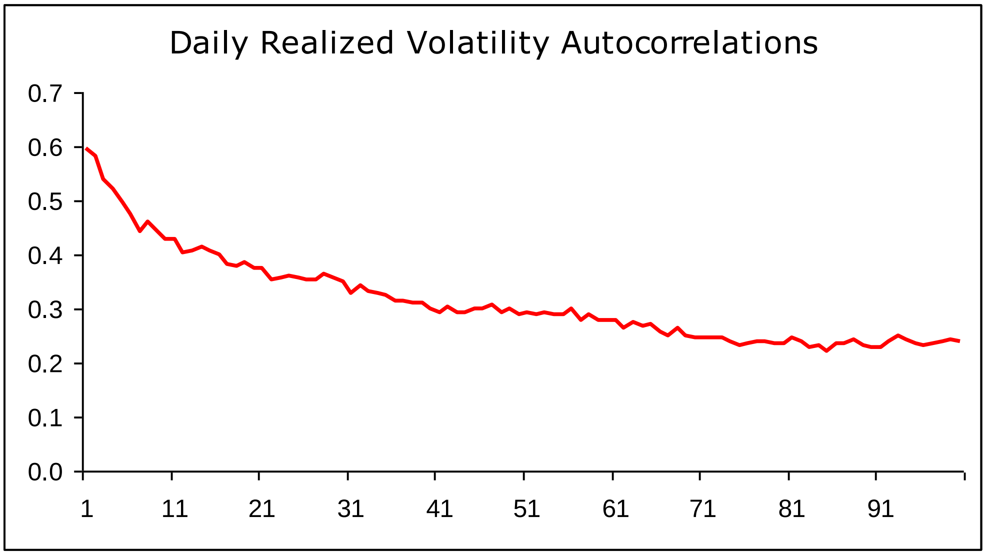

Figure 1 Autocorrelations in S&P500 Daily Realized Volatility

A further examination of the data in Figure 2 shows evidence of a regime shift in the variance process.

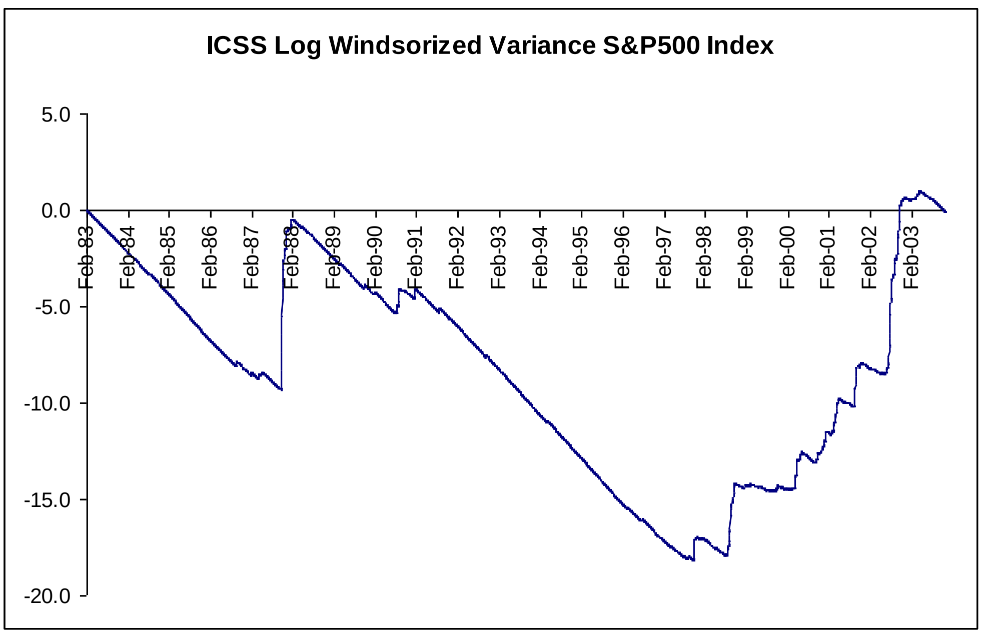

Figure 2 ICSS Test Statistic for Log Windsorized Variance of the S&P 500 Index

The Iterative Cumulative Sum of Squares test statistic suggested by Aggarwal et al (2001) exceeds the 95% confidence limit, indicating a regime shift in the volatility process, in October 1987 and again in 1998 (see Figure 2). While the former is undoubtedly the result of the market crash in Oct 1987, the reason for the latter may be less intuitive. One likely explanation is the evaporation of liquidity and flight to quality resulting from the problems at the Long Term Capital Management hedge fund. Accordingly the cause of the 1987 crash was arguably exogenous, whereas the LTCM event was undoubtedly endogenous. For this reason the series was truncated and the data prior to 1988 excluded, while all of the data from 1988, including the data from 1998 onwards was used for model estimation. The truncated series is close to Normally distributed, having a skewness of 0.17 and kurtosis of 2.74 (see figures following).

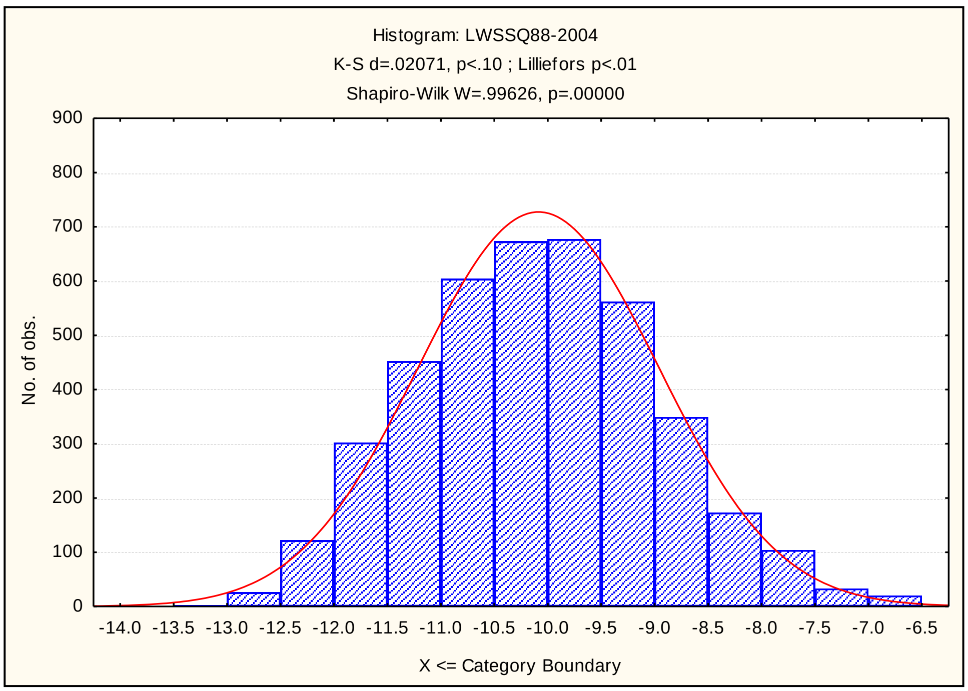

Figure 3 Log Realized Volatility Distribution

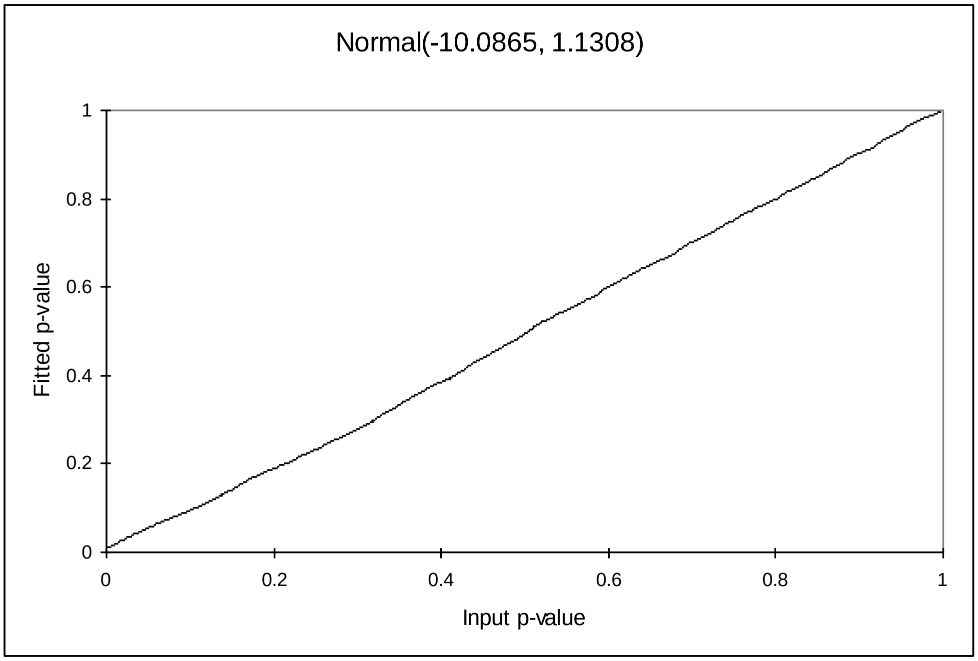

Figure 4 P=P Plot of Log Realized Volatility Distribution

The data series was partitioned to provide a sample used for model estimation, comprising 3,033 daily observations from 4-Jan-1988 to 31-Dec-1999. The out-of-sample series, comprising 1,004 daily log realized volatility observations from 3-Jan-2000 to 31-Dec-2003, was used for model testing as described below.

A further data source was used to provide daily closing prices and implied volatilities for options on the S&P 500 Index from January 2000 to December 2003 and option expiration dates were identified for each month in the out-of-sample period. The S&P 500 Index options are European and the option prices (and implied volatilities) reflected the estimated dividend yield on the index. For each month, a trade date was selected approximately 20 trading days prior to expiration and the average of the Black-Scholes implied volatilities of the at-the-money call and put options used as a proxy of the market's forecast of index volatility over the (approximately) three week period to expiration. This process yielded ex-ante market forecasts of S&P 500 Index volatility for 48 non-overlapping periods between January 2000 and December 2003. Implied volatility estimates are derived from the Black Scholes option pricing formula:

$$C = SN(d_1) - Xe^{-rt}N(d_2) \tag{1.1}$$

$$P = Xe^{-rt}N(-d_2) - SN(-d_1) \tag{1.2}$$

$$d_1 = \frac{\ln(S/X) + (r + \sigma/2)t}{\sigma\sqrt{t}} \qquad d_2 = d_1 - \sigma\sqrt{t} \tag{1.3}$$

Where S is the index price, assumed to follow a log-normal distribution with constant volatility σ , X is the option strike price, t is the time to maturity and r is the risk-free interest rate, with N() denoting the Gaussian density. The implied volatility is that which equates the observed market price of a given option with the model price for a call option C, or put option P.

The forecasting model, initially estimated to the end of 1999, was used to produce ex-ante forecasts of actual index volatility over each expiration period for comparison with implied volatility forecasts. The model was updated with data up to each successive expiration date for the current period before producing ex-ante volatility forecasts to the end of the next expiration period. The use of high-frequency volatility estimates for non-overlapping periods overcomes the principal objections to the less stringent methodology used in earlier studies.

Forecasting Model

Model Form

We adopt a form of model which reflects some of the more important empirical findings about volatility behaviour, specifically mean reversion, a tendency to exhibit slow decay in serial autocorrelations (i.e. long memory) and a responsiveness to random "shocks" which cause the process to jump (temporarily) to higher levels (GARCH behaviour).

Let Y_t for t = 1, . . . T, denote the time series to be modeled. The model estimated is an ARFIMA-GARCH model of the form:

$$(1-L)^d \Phi(L) Y_t = \gamma_0 + \theta(L) u_t \tag{1.4}$$

In which $u_t = h_t^{1/2} e_t$ where error terms $e_t \sim$ iid N(0,1) and h_t is defined by:

$$h_t = \kappa + \alpha u_{t-1}^2 + \beta h_{t-1} \tag{1.5}$$

$\phi(L)$ and $\theta(L)$ are finite-order lag polynomials, such that, for example,

$$\theta(L) = 1 - \theta_1 L - \ldots - \theta_p L^P \tag{1.6}$$

The fractional difference operator term takes the following form:

$$(1-L)^d = \sum_{j=0}^{\infty} \pi_j L^j, \quad \pi_j = \frac{j-1-d}{j} \pi_{j-1}, \quad \pi_0 = 1 \tag{1.7}$$

The model describes volatility as a long memory process, i.e. having long term serial autocorrelation, interrupted by periodic shocks from the kurtosis process, The model is analogous to a standard GARCH model, except that, firstly, it is volatility process itself rather than the returns process which is the subject Y_t of the model and, secondly, Y_t is modeled as a long-memory ARFIMA process. Shocks to the volatility process are fed through via the variance of volatility ("kurtosis") process h_t, which itself evolves as an ARMA-in-squares process with Gaussian error terms, just as in a standard GARCH model (except that it is now the conditional variance of volatility that is the transmission mechanism, rather than the conditional volatility itself).

Post Estimation Tests

Standard tests for serial dependence are the provided by the Q tests for levels (Box-Pierce 1970) and squares (McLeod-Li 1983) of the data. The latter is a test for nonlinear dependence in a serially uncorrelated (white noise) series. The default portmanteau test statistic is the Box-Pierce (1970) formula for autocorrelations up to lag k:

$$Q(k) = n \sum_{s=1}^{k} \rho_s^2$$

which is asymptotically chi-squared with (k–p–q) degrees of freedom when applied to the residuals of an ARMA(p, q) model.

One-step ex-post forecasts are obtained by fitting the model using data up to time T, and then computing the usual fitted equation and residuals for periods T + 1 to T + F, such that all right- hand side variables are treated as known. The main purpose of this option is to test model stability. Two test statistics are computed,

$$\text{Forecast Test I} = \frac{\sum_{t=T+1}^{T+F} \hat{u}_t^2}{Var_T(\hat{u}_t)} \qquad (1.9)$$

where the denominator is the usual residual variance from the sample period, and

$$\text{Forecast Test II} = \frac{F^{-1}\sum_{t=T+1}^{T+F} \hat{u}_t^2 - T^{-1}\sum_{t=1}^{T} \hat{u}_t^2}{\sqrt{F^{-1}Var_F(\hat{u}_t^2) + T^{-1}Var_T(\hat{u}_t^2)}} \qquad (1.10)$$

Test I is an asymptotically valid version of Chow's prediction test, distributed as $Chi^2(F)$ under the null hypothesis of model stability, also assuming the disturbances are Gaussian. Test II is the usual difference-of-means test on the residual and forecast error variances. It is asymptotically N(0,1) under the stability hypothesis, assuming 4th moments exist, where 'asymptotic' is interpreted as min(F,T) $\rightarrow\infty$. That is, these tests are appropriate to 'small' and 'large' forecast periods, respectively.

Results

In-Sample Model Estimation

Parameter estimates for the initial sample (to 30-Dec-1999) are shown in Table 1 below. All are significant at the 5% confidence level and confirm the presence of important long memory effects in the volatility process. The size of the GARCH AR1 coefficient, β, indicates that shocks to the volatility process introduced through the fourth moment dissipate quickly, having a half life of only 1.58 days ($0.64536^{1.58} = 0.5$). Note that the MA 1 parameter θ_1 is the only significant term in the polynomial lag function $\Theta(L)$ in equation 1.4. We note also that the sum of the GARCH parameters α and β exceeds 1, which indicates process non-stationarity.

	Intercept γ_0	ARFIMA d	MA1	GARCH Intercept κ	GARCH AR1 β	GARCH MA1 α
Estimate	-9.98129	0.46343	0.33878	0.73917	0.64538	0.58729
S.E.	0.4044	0.0356	0.0491	0.0257	0.2328	0.2518
Prob.	0	0	0	0	0.005	0.019

Table 1 ARFIMA-GARCH Model Parameter Estimates

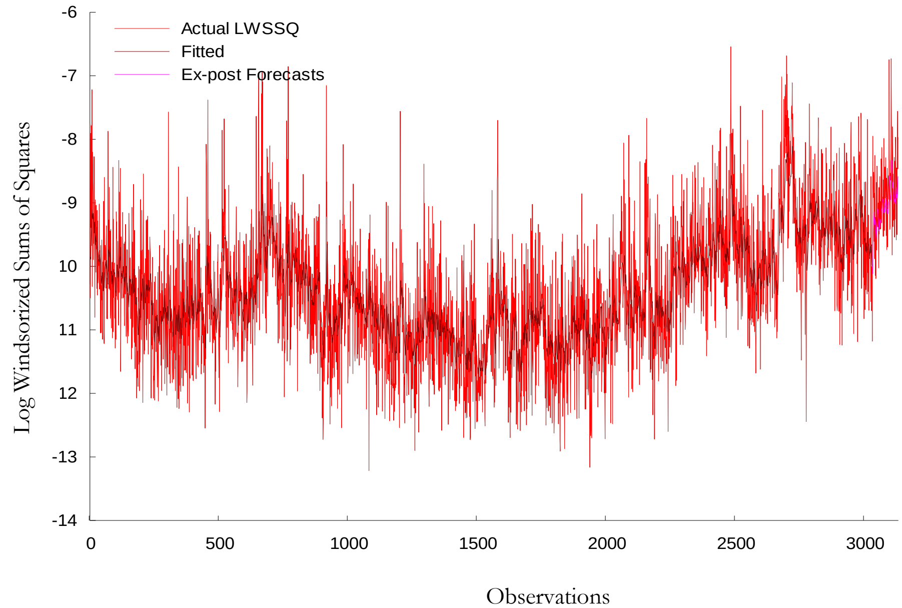

Figure 5 Actual and Forecast Log Volatility

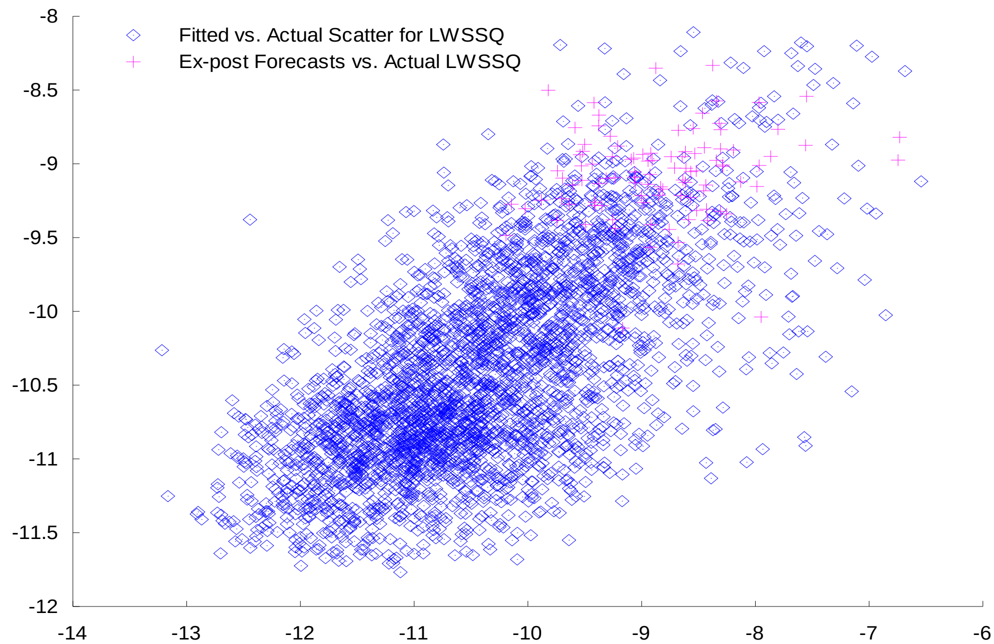

Figure 6 Actual vs. Forecast Log Volatility

There is some evidence of non-Normality in the error process as indicated by the small residual skewness and kurtosis values and confirmed by the significant value of the Jaque-Bara test statistic in Table 2 (probabilities are shown in brackets). As the distribution and Q-Q plots confirms, this is primarily due to excess kurtosis ("fat tails") in the error process (see Figures 8 and 9 following). However, Box-Pierce tests at 40-lags indicate no significant patterning in the residual autocorrelations, nor any indications of residual GARCH effects and this is confirmed by Lagrange Multiplier and Conditional Moment tests (see Table 3).

R Squared	Residual SD	Residual Skewness	Residual Kurtosis	Jarque-Bera Stat.	Box-Pierce(40)	Box-Pierce (Sq)(40)
0.4075	0.7984	0.3386	3.4547	81.317	47.1468	30.4714
				{0}	{0.203}	{0.861}

Table 2 Residuals Test Statistics

Lagrange Multiplier and Conditional Moment Tests
Autocorrelation (LM): ChiSq(40) = 45.5831{0.251}
Autocorrelation (CM): ChiSq(40) = 47.2716{0.199}
Neglected ARCH (LM): ChiSq(40) = 8.3297{0.999}
Neglected ARCH (CM): ChiSq(40) = 36.1712{0.643}

Table 3 Lagrange Multiplier and Conditional Moment Tests

A battery of diagnostic tests confirm that the process is correctly identified as a fractionally integrated process rather than a stationary, or first-difference stationary process (see Table 4).

Tests of I(0)/I(1) (Parzen kernel with bandwidth 4)
Lo's RS test = 7.87784{<0.005}
KPSS test = 12.9966{<0.01}
Phillips-Perron test = -32.2676{<0.01}
Robinson's d = 0.428306

Table 4 Fractional Integration Tests

Stability tests conducted on a sample of 100 ex-post forecasts suggest no instability in the model specification (see Table 5).

Model Stability Tests
Forecast Test 1: ChiSq(100) = 101.967 {0.426}
Forecast Test 2: N(0,1) = 0.1417 {0.887}

Table 5 Stability Tests

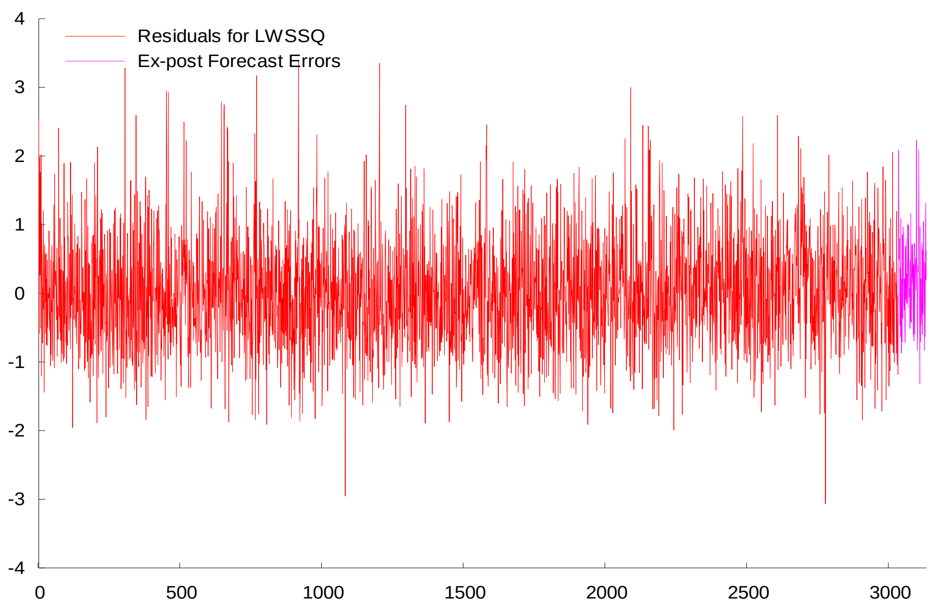

Figure 7 Residuals Plot

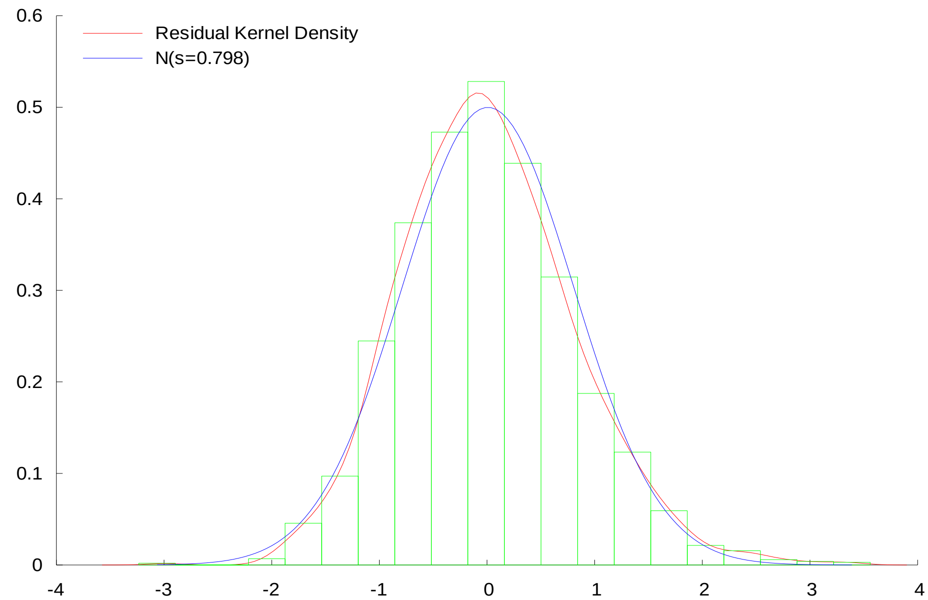

Figure 8 Residuals Distribution

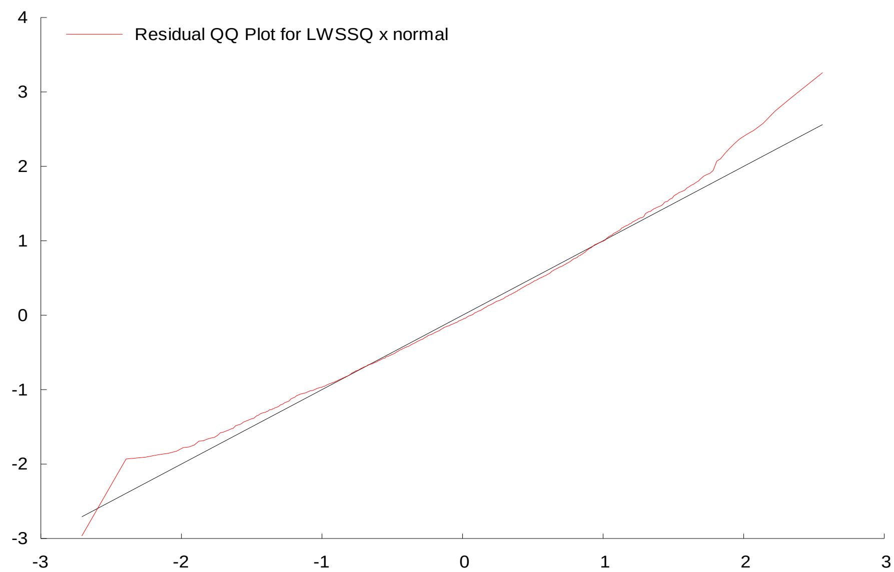

Figure 91 Residuals Q-Q Plot

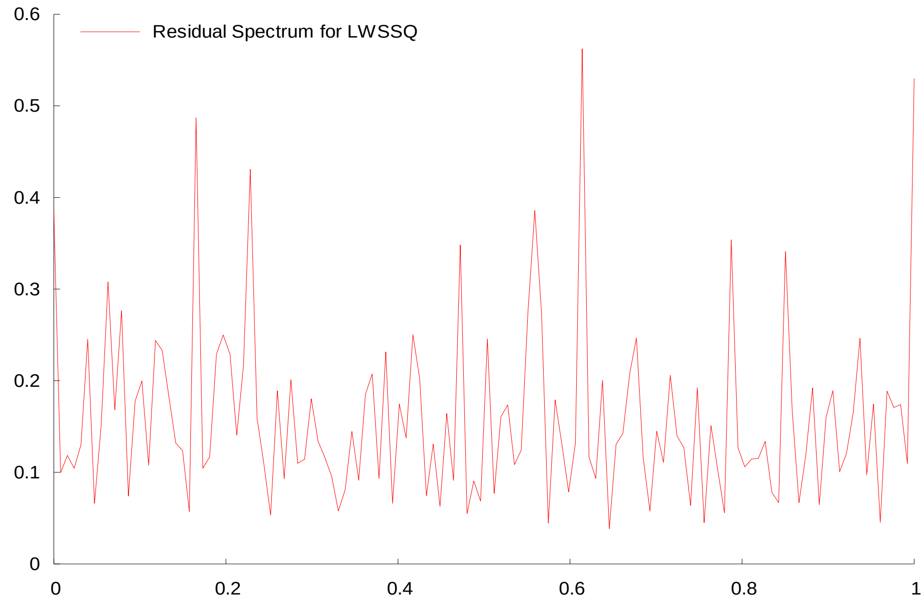

Figure 10 Residual Spectrum

Out-of-Sample Recursive Estimation and Volatility Forecasting

In the second stage of the analysis the model is recursively re-estimated over the entire out-of-sample period from January 2000 to December 2003 and ex-ante daily log volatility forecasts produced to the end of each expiration period. Daily forecasts are aggregated to produce ex-ante forecasts for volatility over each entire expiration cycle and these are compared to market forecasts of realized volatility-to-expiration, as expressed in the implied volatilities of call and put options on the S&P 500 Index. In this analysis only the closest to the money strikes are used to estimate implied volatility, as the presence of a substantial volatility skew in the out-of-the-money put options would otherwise inflate the market estimate of volatility.

Results of the recursive estimation procedure are summarized in the figures in the Appendix to this paper. All of the model parameter estimates show signs of drift over the out-of-sample period, as the figure for the ARFIMA-d parameter estimates illustrates (see Figure 11 below). Note that the 95% confidence bands around the recursive ARFIMA-d parameter estimates include values in excess of 0.5, which would indicate non-stationarity in the volatility process. However, portmanteau test statistic estimates show stability over the entire period and are statistically insignificant for both error and error-squared autocorrelations. Recursive estimate of residual skewness and kurtosis are likewise very stable.

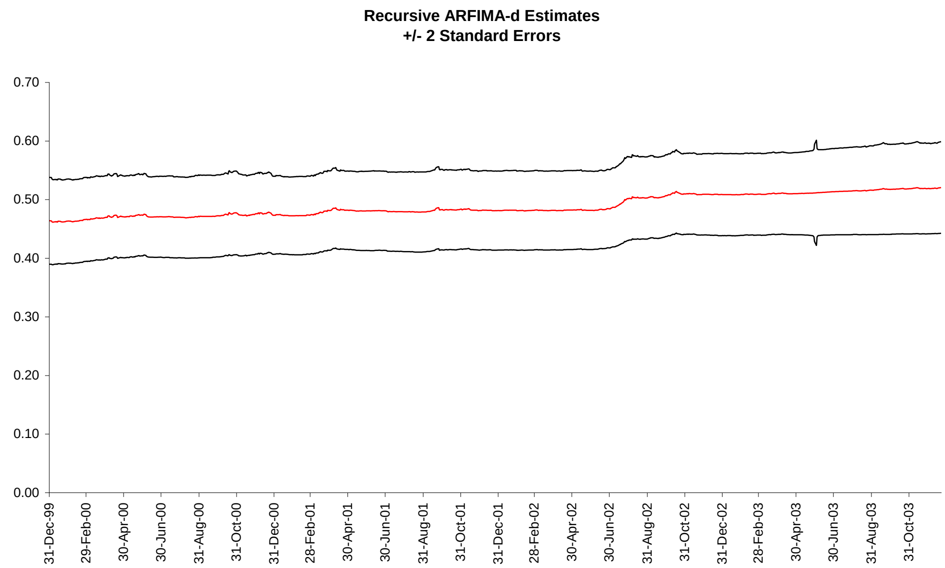

Figure 11 Recursive Estimates of ARFIMA-d Parameter

Turning now to an examination of the forecasts produced by the model we find that, as shown in Table 6, the correlation between realized and implied volatility is significantly higher than that between realized and forecast volatility, while the standard deviation of implied volatility estimates more closely matches that of realized volatility than does the standard deviation of forecast estimates.

	Realized Volatility	Implied Volatility	Forecast Volatility
N	48	48	48
Mean	16.31%	22.41%	16.07%
SD	5.16%	5.33%	3.43%
Skewness	1.06	1.27	0.87
Kurtosis	1.31	1.88	0.23
Range	23.43%	24.78%	13.37%
Min	8.58%	14.62%	11.40%
Max	32.02%	39.40%	24.78%
MAPE		43.2%	17.8%
MSE		23.4%	6.7%
Corr.		78.6%	68.4%
Sign		40.4%	61.7%
# Hits		19	29
Prob		87.85%	3.95%
Theil's U		1.46	0.82

Table 6 Forecast Test Statistics

In other words, implied volatility appears to do a better job of anticipating the variation in future volatility (or kurtosis). This explains why the results of regression analysis tend to favor implied volatility over forecast volatility as a (linear) predictor of future volatility. An encompassing regression of the form

$$RV_{t,T} = \alpha + \beta_1 IV_{t,T} + \beta_2 FV_{t,T} \tag{1.11}$$

where

$RV_{t,T}$ is the realized volatility for period (t,T),

$IV_{t,T}$ is the implied volatility forecast for period (t,T), and

$FV_{t,T}$ is the model forecast volatility for period (t,T),

indicates that there is no additional predictive power to be derived from the addition of volatility forecasts based on historical data, once implied volatility is included (see Table 7).

	Coefficients	*SE*	*t Stat*	*P-value*
Intercept	-0.00776	0.0228	-0.34076	0.734871
Implied	0.758165	0.1806	4.198554	0.000125
Forecast	0.005751	0.28	0.020539	0.983704

Table 7 Encompassing Regression

Separate regression models in which implied volatility and model forecasts are treated as the sole independent variables yield parameter estimates as shown in Table 8 below.

Indep. Variable	**Alpha**	**Beta**	**R-Sq**
Implied Volatility	-0.0076	0.7614	0.6181
S.E.	0.0203	0.0882	
Prob.	0.7117	0	
Forecast Volatility	-0.0021	1.0278	0.4685
S.E.	0.0265	0.1614	
Prob.	0.9376	0	

Table 8 Parameter Estimates for Separate Regression Models

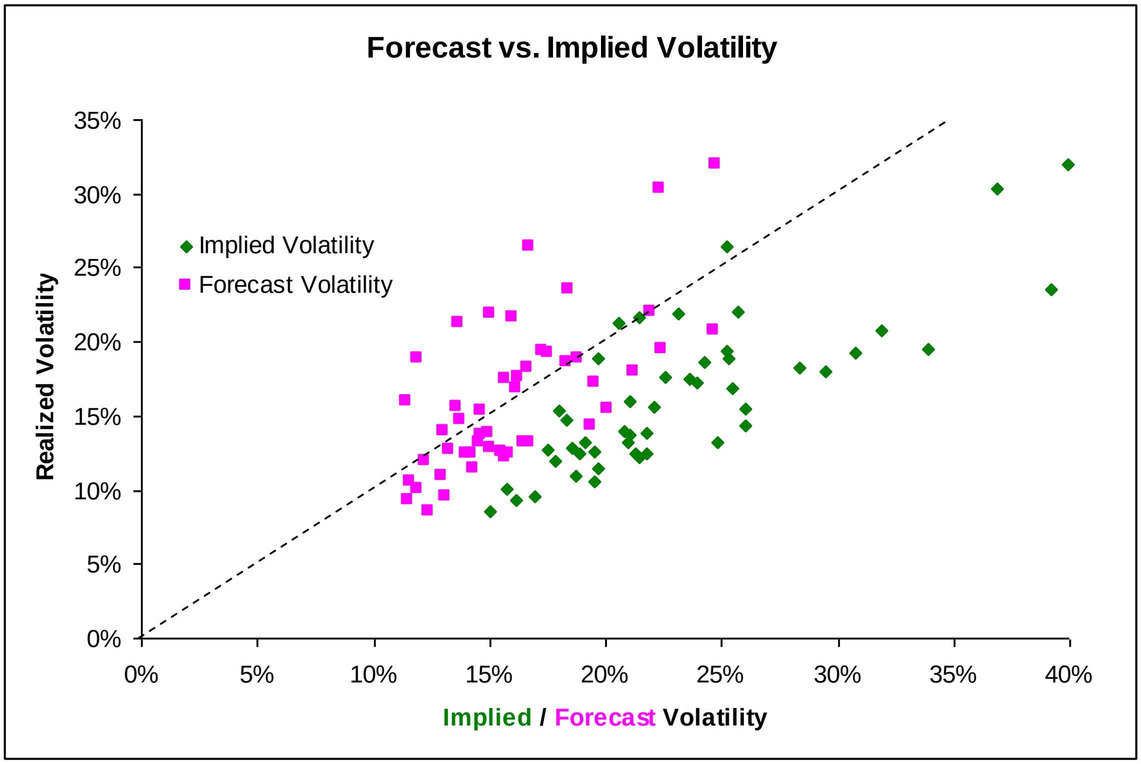

Figure 12 Forecast vs. Implied Volatility

In common with earlier studies, we find that implied volatility estimates, which have greater explanatory power (higher R-squared), yield biased forecasts of future realized volatility (the F-test of the joint null hypothesis that alpha = 0 and beta = 1 is rejected at the 5% level). By contrast the ARFIMA-GARCH model, although having lower explanatory power, provides volatility forecasts that are found to be unbiased: the F-test is unable to reject the joint null hypothesis that alpha = 0 and beta = 1.

Returning to the analysis shown in Table 6, a comparison of the actual (realized), implied and forecast volatility reveals a consistent pattern of bias. Implied volatility projections tend to over-estimate future realized volatility, by an average of 610 basis points over the out-of-sample period, although there is considerable variation in the level of the bias over time, ranging from a minimum of -1.22% (in June 2002) to a maximum of 15.61% (in March 2001) and with an annual standard deviation of 12.25%. Model forecasts tend to underestimate realized volatility, but only by an average of 24 basis points and with a somewhat smaller range of -9.69% (in June 2002) to 4.79% (in April 2001). Further, the mean absolute percentage error (MAPE) of the volatility forecasts is less than half that of implied volatility estimates, while the mean square error (MSE) is less than a third.

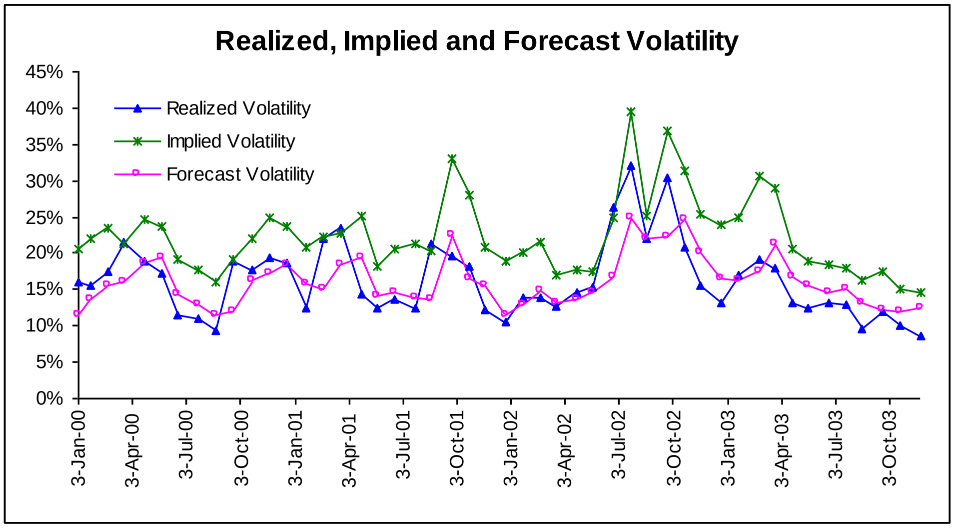

Figure 13 Out of Sample Realized, Implied and Forecast Volatility

Exploring the relative predictive abilities more deeply, we find that (perhaps unsurprisingly, given the tendency towards upward bias) implied volatility provides very poor predictive ability vis-à-vis the direction of future realized volatility. Of the 48 out-of-sample periods, implied volatility correctly anticipated the change in direction in realized volatility in only 19 cases (40.43%), whereas volatility forecast correctly anticipated the change of direction in 29 periods (61.70%), which is significant at the 4% level. The predictive superiority of the model is confirmed by the Theil's-U test statistic, which at 0.82 demonstrates a substantial improvement over the naïve predictor. Implied volatility, by contrast, performs much worse than a naïve predictor throughout the out-of-sample period (Theil's-U= 1.46).

There are at least two possible explanations of the superiority of model forecasts in terms of bias/direction prediction ability vis-à-vis market forecasts expressed in Black-Scholes implied volatility. The first is that option markets operate inefficiently, one conjecture being that the model's superior sign prediction capability derives from its ability to capture important long-memory effects. The latter may not feature appropriately in the ad-hoc day-to-day volatility estimates of market traders, as it must surely be very difficult for a trader in the pit to assign appropriate weightings for historical volatility in the asset as far back two years or more, even supposing the data is available to him. If correct, this finding would have important implications for option market efficiency and the potential for making abnormal returns from volatility trading. Firstly, the ability to predict the general level of volatility is considered more important than the ability to explain its variation: an average overpricing of index options by 610 volatility basis points is very likely to facilitate the generation of abnormal returns. Secondly, the ability to predict the direction of a market, rather than the magnitude of the change, is in and of itself likely to be valuable. Market timing strategies that achieve a "hit" rate of 56% - 57% are generally considered by market practitioners to afford a sufficient edge to be worth trading in most markets. Here the edge in terms of market timing ability is substantially greater.

Volatility Trading Strategy

The potential for making abnormal profits from volatility trading is tested by means of the following simple procedure. We gathered daily data for near-to-the money options on the S&P 500 Index for each day in the out-of-sample period, including both the price and the option delta at market close. We then simulated the purchase or sale of at-the-money straddles, depending on whether the forecast volatility to expiration exceeded or was exceeded by the average implied volatility of the two option contracts. The resulting long or short straddle position was held to maturity and delta-hedged at the end of each trading day (using the net of the closing deltas of the two option contracts). It can be assumed that delta hedging is accomplished by trading SPYDrs, contracts that trade at the American Stock Exchange at 1/10 the size of the S&P 500 Index. Hedge contracts are bought (sold) at the reported end-of-day offer (bid) price. We factor in trading costs at a rate of $1 per round turn for each option contract and at $0.005 for each SPYder. These costs are somewhat higher than would be incurred by a sizeable fund trading on the available ECN's (electronic exchanges). No allowance is made for market impact, which however should be negligible in these highly liquid instruments up to notional capital amounts of $500M or so.

The results of this rather elementary trading strategy are presented in the table and figure following:

		Winner	Losers
# Periods	48	30	18
		62.5%	37.5%
Avg, Return	1.56%	4.10%	-2.68%
Largest		17.52%	-8.35%
St. dev.	4.78%	4.07%	2.13%
Compound	98.10%		
Ann. Compound	18.64%		
Ann. Volatility	16.55%	14.08%	7.39%
Sharpe	0.88		
Beta	0.00		
Alpha	22.64%		
Correl.	0.01		
Max Drawdown	-19.26%		
# Periods	7	(18 Mar 02 - 21 Oct 02)	
Recovery Time	10		

Table 9 Strategy Performance Analysis

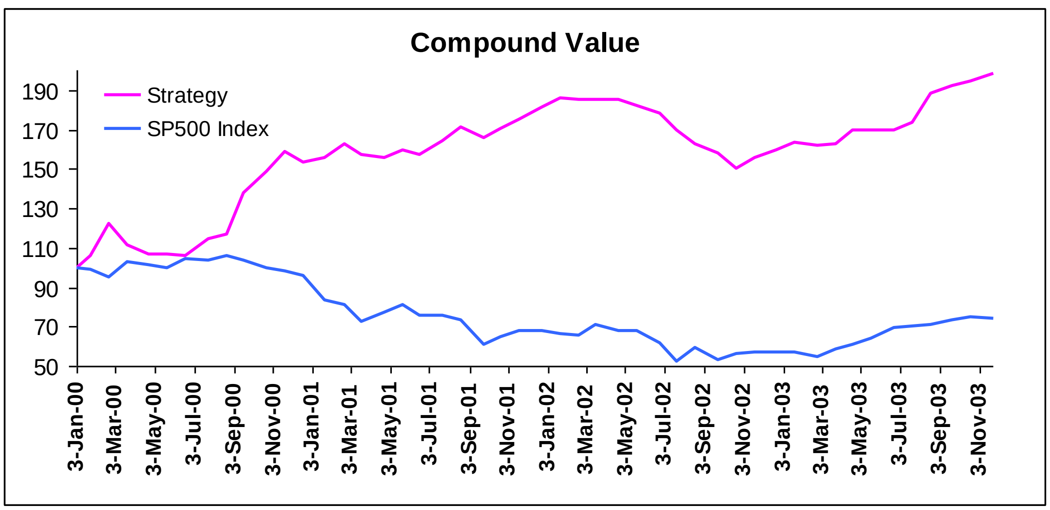

Figure 14 Compound Returns for Volatility Trading Strategy

Of the 48 option expiration periods, 30 (62.5%) turned out to be profitable for the strategy, a proportion which closely mirrors that of the percentage direction prediction accuracy (see Table 6). The average return per option expiration period was 1.56%, which equates to a net annual compound return of 18.64% over the 4-year test period, during which the annual return on the S&P index itself was -7.24%. The strategy's risk-adjusted rate of return (Sharpe ratio)[2] of 0.88 is higher than the long-term average of the majority of Tremont hedge fund indices and, unlike for many investment strategies, returns are uncorrelated with the underlying market. With an annual alpha[3] of 22.64%, this simple strategy appears to offer clear empirical evidence of the potential for generating abnormal returns using volatility forecasts based on high-frequency observations of historical volatility alone.

This is not to say, however, that the strategy is an investable proposition. The assumption that the portfolio is to be delta-hedged only at market close is probably unrealistic, as it ignores the sizeable delta-swings that often arise from intra-day market movements. Again, there is a lengthy period of seven months from March to October 2002 in which the strategy suffers a drawdown in excess of 19%. Many investors would regard as unacceptable an investment proposition which could result in losses on that scale over half the trading year. An examination of the evolution of realized volatility in the preceding Figure 23 reveals the nature of the problem: over the period from March to October 02 the model, while often correctly predicting the direction of volatility movements, significantly under-forecasts future realized volatility. This is the result of a weakness previously identified in the model – its underestimation of the kurtosis of the returns process and hence of the magnitude of volatility spikes. This could in principle be corrected by fine tuning the model – either by using a non-Gaussian distribution (such as, for instance a Student-t distribution) to model the error process, or by including an explicit kurtosis term in the model itself.

[2] Defined as $S = (R_s – R_a) / \sigma$ where Rs is the average return for the strategy, R_f is the risk-free rate, and σ is the standard deviation of strategy returns.

[3] Defined as the excess return over the underlying S&P 500 Index.

Conclusion

Echoing the findings of parallel empirical research, this study points to the conclusion that historical realized volatility adds little to the explanatory power of implied volatility forecasts. However, one perplexing feature of implied volatility forecasts is their persistent upwards bias. As a result, forecasting models using high-frequency historical data may have an edge over implied volatility forecasts in predicting the *direction* of future realized volatility. The ability to time the market by correctly predicting its direction approximately 62% of the time appears to offer the potential to generate abnormal returns by a simple strategy of buying and selling at-the-money straddles and delta-hedging the resulting positions on a daily basis through to expiration, even after allowing for realistic transaction and hedging costs.

References

Aggarwal, R., Inclan, C., Leal, R., 1999, Volatility in Emerging Stock Markets, Journal

of Financial and Quantitative Analysis, Forthcoming.

Andersen, T.G., 2000, Some reflections on analysis of high-frequency data, *Journal of Business & Economic Statistics* 18, 146-153.

Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (1999), "The Distribution of Exchange Rate Volatility," Wharton Financial Institutions Center Working Paper 99-08 and NBER Working Paper 6961.

Andersen T.B., Bollerslev, Diebold and Labys, 2001. Modeling and forecasting realized volatility. Working paper Northwestern University

Bakshi G., Cao C., and Chen Z., 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003-2049.

Breidt, J.F., Crato, N., de Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Economics 83, 325–348.

Canian, L. and Figlewski, S., 1993, The information content of implied volatility, *Review of Financial Studies* 6, 659-681.

Christensen, B.J., and Prabhala, N.R., 1998, The relationship between implied and realized volatility, *Journal of Financial Economics* 50, 125-150.

Dumas, B., Fleming, J., and Whaley, R.E., 1998, Implied Volatility Functions: Empirical Tests, *The Journal of Finance* 6, 2059-2107.

Fleming, J., 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317-345.

Geweke, J., Porter-Hudak, S., 1983, The Estimation and Application of Long Memory

Time Series Models, Journal of Time Series Analysis, 4, 15-39.

Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–29.

Harvey, C.R., Whaley, R.E., 1992, Market volatility prediction and the efficiency of

the S&P 100 index option market. Journal of Financial Economics 31, 43-73.

Heston, S.L., 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327-343.

Hull, J., White, A., 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281-300.

Martens, M. and Zein, J., 2002, Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility, forthcoming.

Nandi, S., 1998, How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence from pricing and hedging in the S&P 500 index options market, *Journal of Banking and Finance* 22, 589-610.

Noh, J., Engel, R.F., Kane, A., 1994. Forecasting volatility and option prices of the S&P 500 index. Journal of Derivatives 2(1), 17-30.

Parkinson, M., 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61-68.

Poteshman, A.M., 2000, Forecasting future volatility from option prices, working paper, University of Illinois at Urbana-Champaign.

Pong, S, Shackleton, M. B., Taylor, S. J. and Xu, X., 2003, Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models, forthcoming, Journal of Banking and Finance.

Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188.

Taylor, S.J., and Xu X., 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* 4, 317-340.

Yang, D., and Zhang Q., 2000, Drift independent volatility estimation based on high, low, open and close prices, *Journal of Business* 73, 477-491.

Appendix

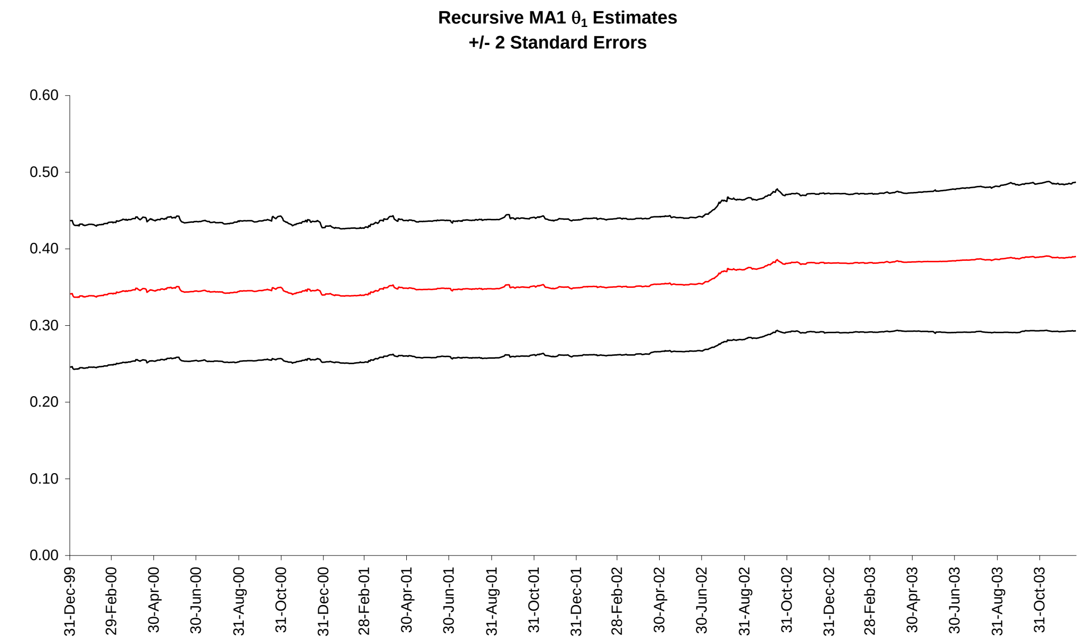

Figure 15 ARFIMA MA(1) Recursive Estimates

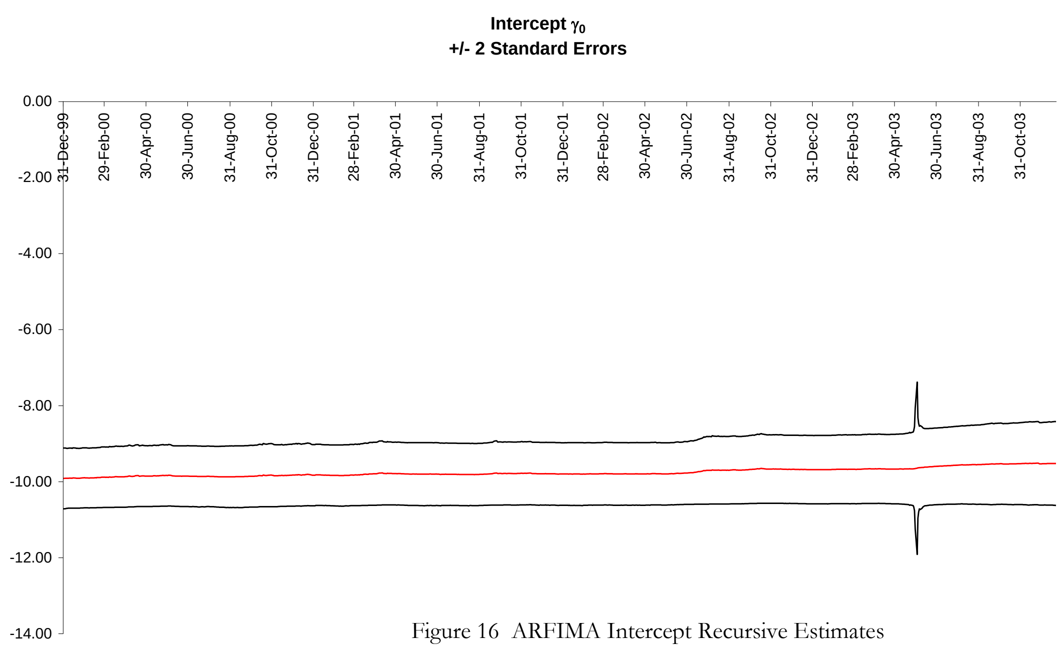

Figure 16 ARFIMA Intercept Recursive Estimates

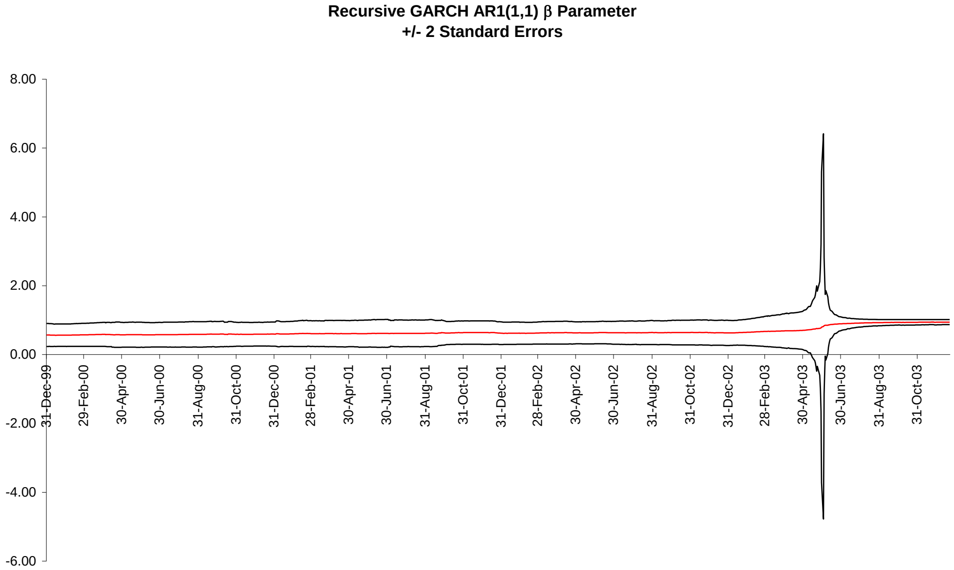

Figure 17 GARCH AR1(1,1) Recursive Estimates

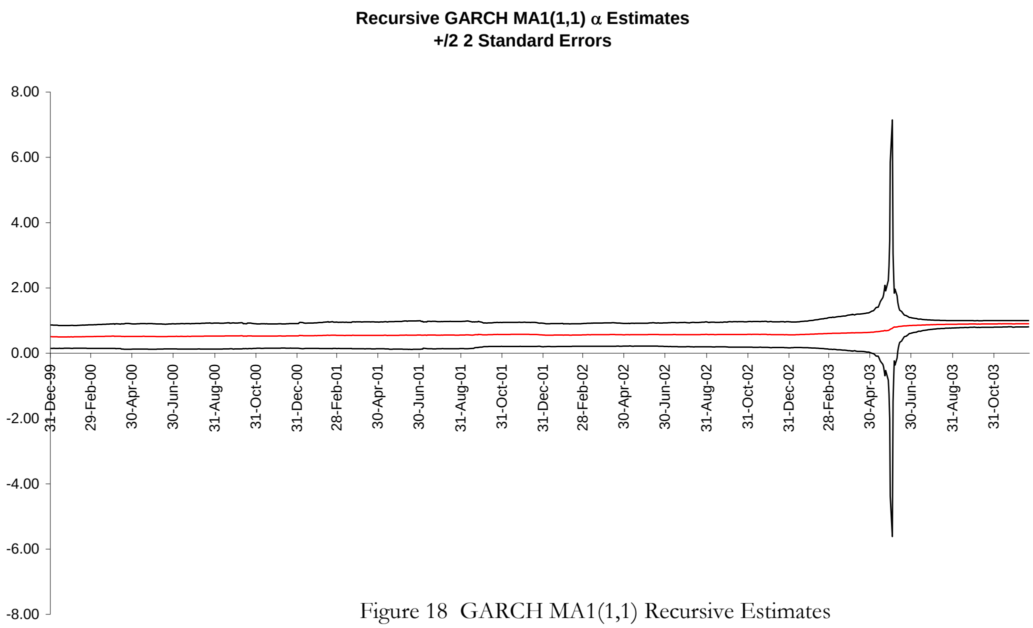

Figure 18 GARCH MA1(1,1) Recursive Estimates

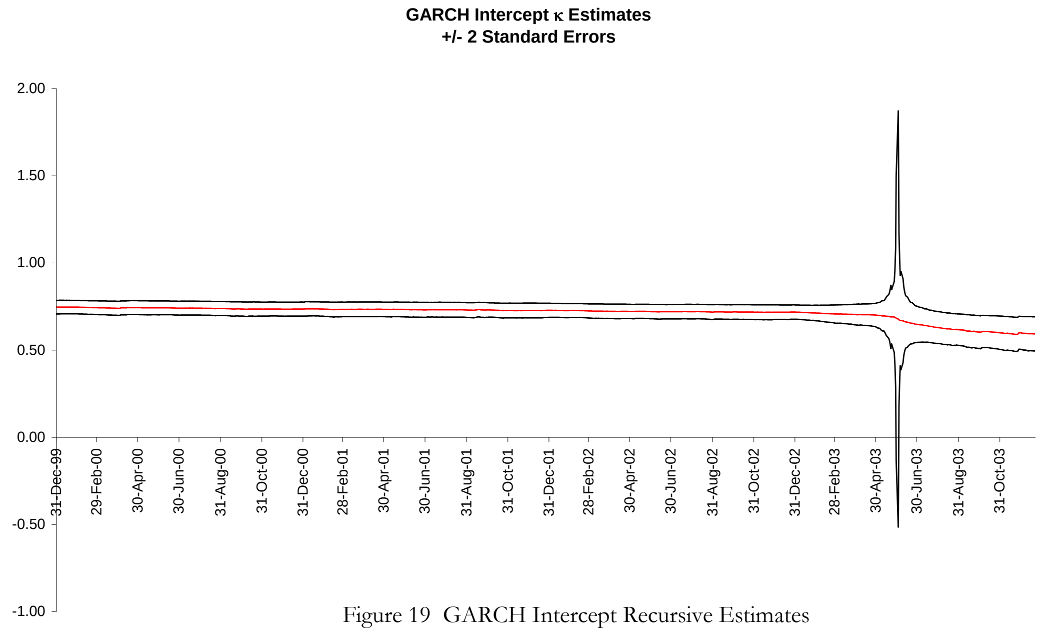

Figure 19 GARCH Intercept Recursive Estimates

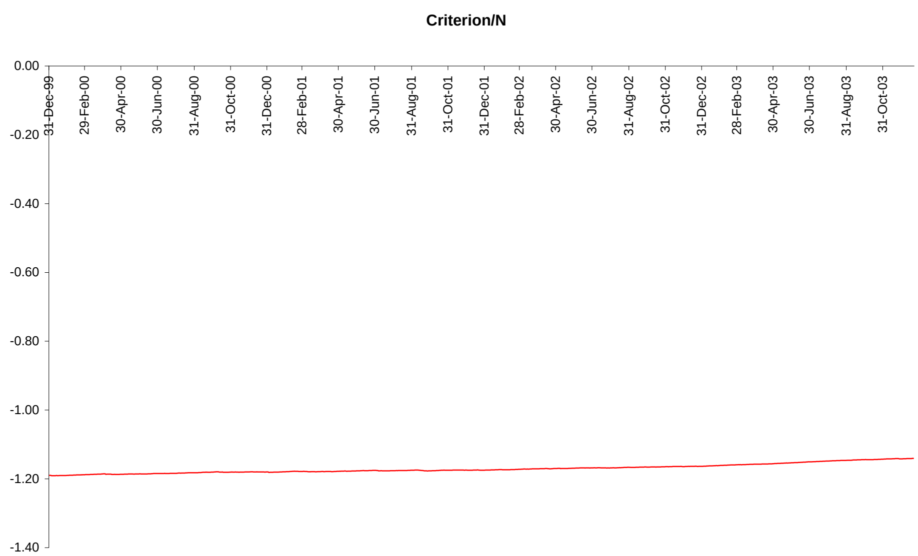

Figure 20 Criterion / N recursive Estimates

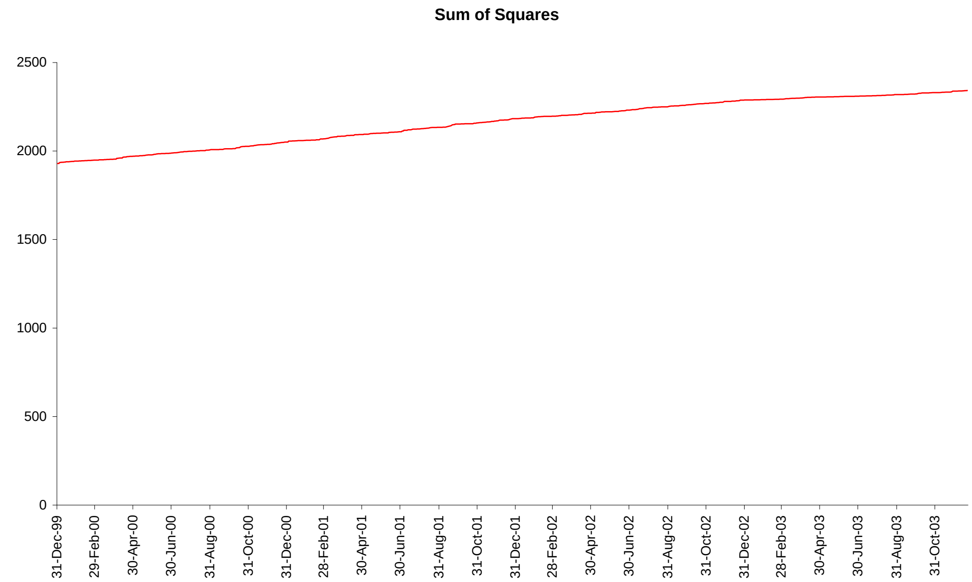

Figure 21 Sum of Squares Recursive Estimates

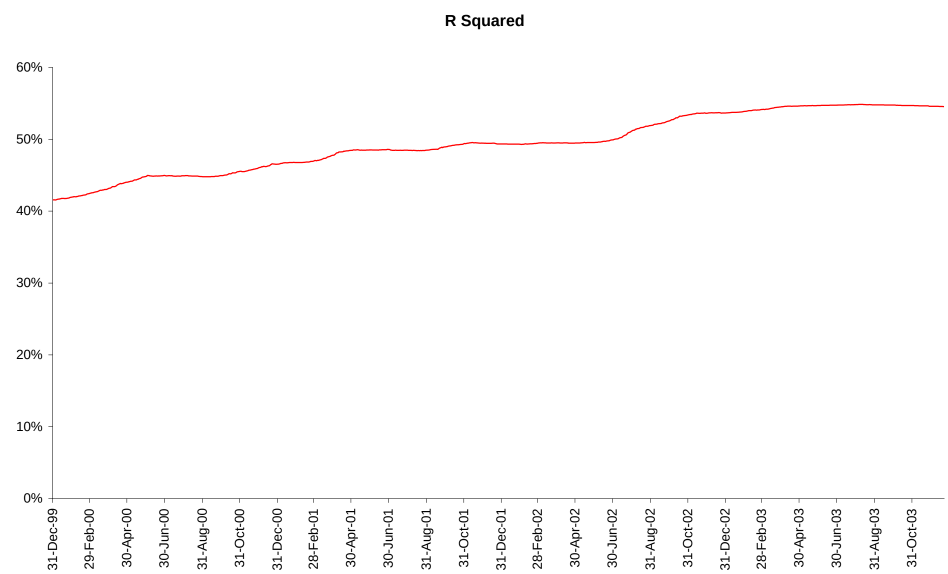

Figure 22 R-Squared Recursive Estimates

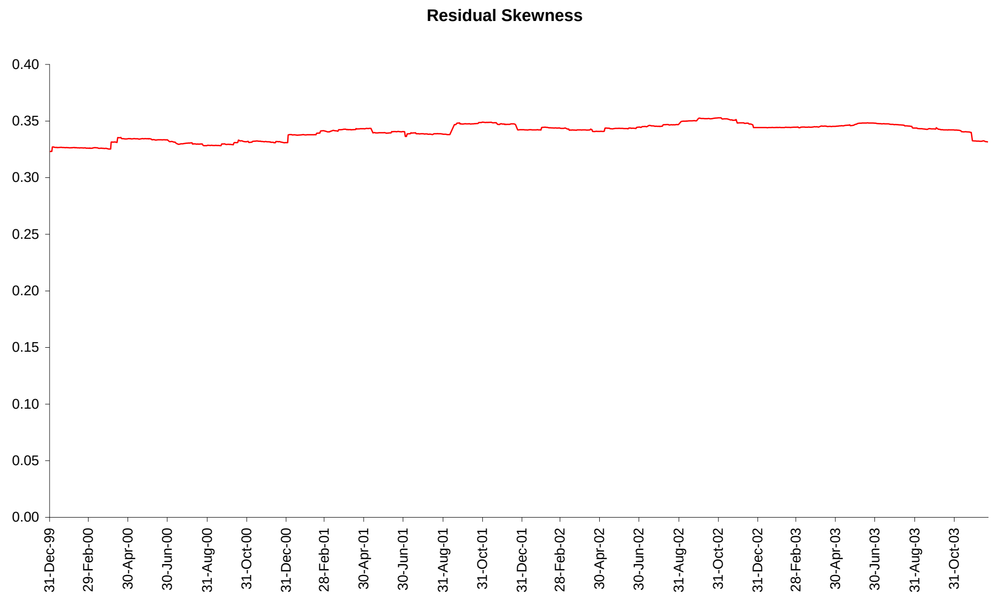

Figure 23 Residual Skewness Recursive Estimates

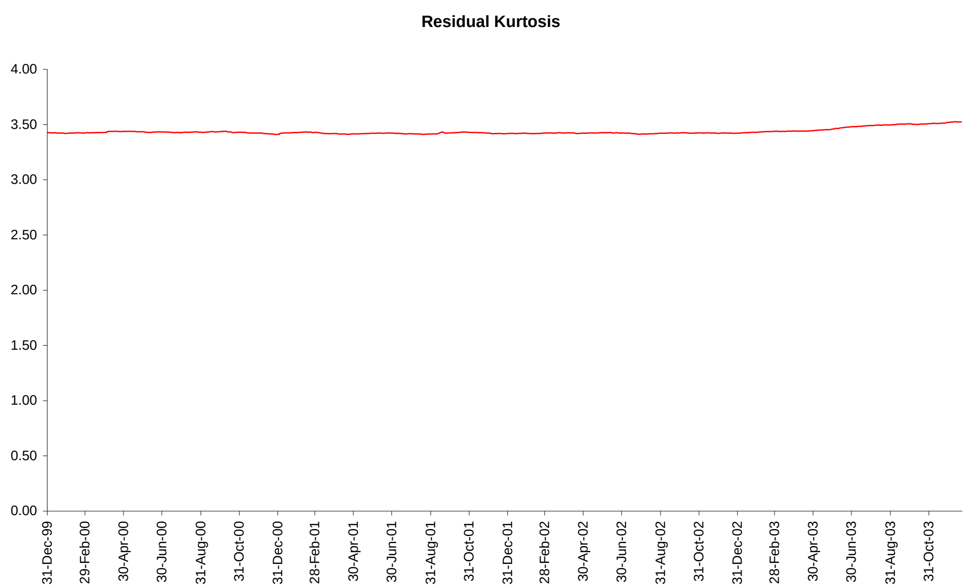

Figure 234 Residual Kurtosis Recursive Estimates

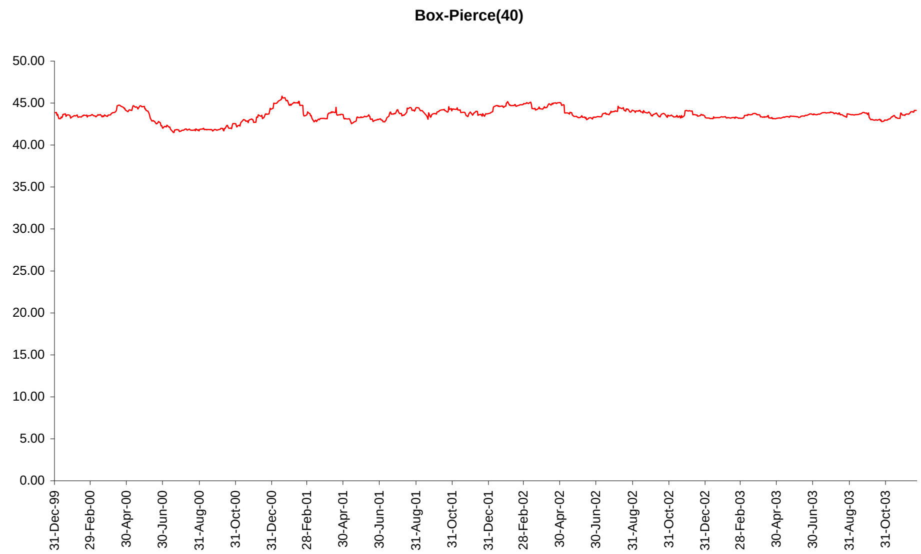

Figure 25 Box-Pierce(40) Recursive Estimates

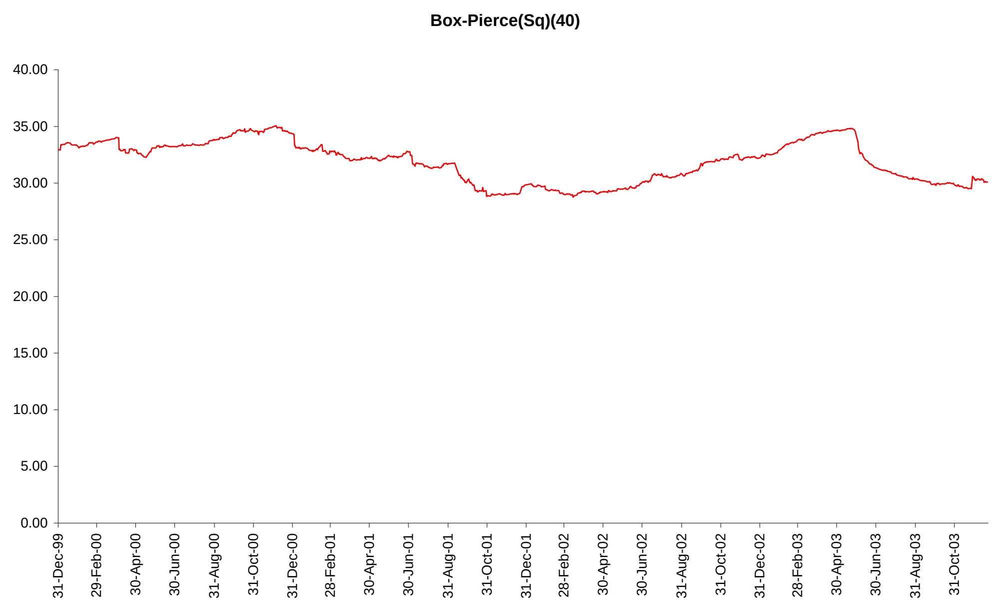

Figure 26 Box-Pierce (Sq)(40) Recursive Estimates