NEW YORK UNIVERSITY
STERN SCHOOL OF BUSINESS
FINANCE DEPARTMENT

Working Paper Series, 1994

Forecasting Volatility Using Historical Data

Stephen Figlewski

FD-94-32

FORECASTING VOLATILITY USING HISTORICAL DATA

by

Stephen Figlewski*

Draft of: May 2, 1994

* Professor of Finance and Yamaichi Faculty Fellow
New York University Stern School of Business
44 West 4th Street, Suite 9-160
New York, NY 10012-1126

Telephone: 212-998-0712
FAX: 212-995-4220
EMAIL: sfiglews@rnd.stern.nyu.edu

There have been so many valuable comments on this research on so many different occasions that at this point it is no longer feasible to mention by name even all of the seminars and conferences at which they were received. I therefore issue a generic thank you to all who have offered their helpful thoughts on this research over the years. You know who you are. I would especially like to thank Amar Gande, Edie Hotchkiss, and Sundar Polavaram for able and patient research assistance, and the Bank and Financial Analysts Association for funding this project.

FORECASTING LONG TERM VOLATILITY FROM HISTORICAL DATA

ABSTRACT

Applying modern option valuation theory requires the user to forecast the volatility of the underlying asset over the remaining life of the option, a formidable estimation problem for long maturity instruments. The standard statistical procedures using historical data are based on assumptions of stability, either constant variance, or constant parameters of the variance process, that are unlikely to hold over long periods. This paper examines the empirical performance of different historical variance estimators and of the GARCH(1,1) model for forecasting volatility in important financial markets over horizons up to five years. We find several surprising results: In general, historical volatility computed over many past periods provides the most accurate forecasts for both long and short horizons; root mean squared forecast errors are substantially lower for long term than for short term volatility forecasts; it is typically better to compute volatility around an assumed mean of zero than around the realized mean in the data sample, and the GARCH model tends to be less accurate and much harder to use than the simple historical volatility estimator for this application.

FORECASTING LONG TERM VOLATILITY FROM HISTORICAL DATA

Introduction

Option pricing theory has developed into a standard tool for designing, pricing, and hedging derivative securities of all types. The array of available and actively traded products has expanded enormously in recent years, as new classes of instruments have been created and traditional ones have been become more widely used.

All valuation models for options and instruments with any option component require at least one volatility parameter. For more elaborate models, the user may have to specify a set of parameters to define a time-varying stochastic volatility <u>process</u>. Since volatility is unobservable, this turns option valuation in the real world into a forecasting problem. A variety of methods for obtaining a volatility estimate are in common use.

Until fairly recently, explicit options (as opposed to embedded options, like a call provision in a long term corporate bond) have had maturities that were typically measured in weeks or months rather than years. Although volatility has proven to be notoriously difficult to predict accurately and it appears to change randomly over time, one generally assumed that treating it as a constant parameter over the short run was not too bad. However, the expansion in derivatives activity has also brought a marked lengthening in the horizons for which contracts may be written, first for over-the-counter derivatives such as puts and calls on foreign currencies, and then for exchange-traded instruments like LEAPS and FLEX contracts. Today, maturities of 5 to 10 years are not uncommon.

Valuation and risk management, or simply evaluation of credit risk, for such long term derivatives poses a major forecasting problem. How should one try to predict the volatility of, say, the Deutschemark/dollar exchange rate over the next ten years? Is one better off using a sophisticated approach that attempts to model the stochastic variation in volatility over time or a "rule-of-thumb" constant volatility procedure that is clearly over-simplified but may be more robust as the financial environment evolves over a long horizon? What is the probable magnitude of the forecast error for the best available prediction technique? The object of this paper is to explore these issues empirically for several important financial instruments. We will examine and contrast different volatility estimation procedures specifically from the perspective of their accuracy in producing out of sample forecasts.

In the next section, we consider the standard procedure for estimating a (constant) volatility parameter from historical data, and in Section 3 we discuss several tricky issues in implementing it in practice. Section 4 examines the forecasting performance of the standard historical volatility estimator in different markets as a function of the forecasting horizon and the number of past periods in the data sample. We also show that computing volatility around an assumed mean of zero rather than around the sample mean may increasse forecast accuracy.

Using historical volatility as the forecast of future volatility treats volatility as a constant parameter, even though a great deal of evidence suggests that it is not. In Section 5 we discuss formal models of time-varying volatility. If volatility is not constant, it would seem that a model explicitly allowing time-variation ought to produce more accurate predictions. But that is not necessarily the case. Using a GARCH model, for example, allows volatility to vary systematically over time, but now the GARCH parameters themselves must be constant and accurately estimable from past data. Otherwise it can turn out that even though volatility is not constant, using historical volatility produces more robust forecasts than more sophisticated, but more fragile, approaches. In fact, when we examine the performance of the GARCH(1,1) model in Section 6, we find considerable difficulty in attempting to forecast volatilities over long horizons with it, and no clear improvement in accuracy for the cases in which they could be used.

The final section summarizes our results.

2. Computing Historical Volatility

In theoretical option pricing models, the term "volatility" has a very clear and precise meaning, and academic financial economists immediately think of that interpretation when the volatility of security prices is discussed. Black and Scholes derived their option valuation equation under the assumption that stock returns, "log price relatives" to be precise, followed a logarithmic diffusion process in continuous time with constant drift and volatility parameters, as shown in equation (1).

$$\frac{dS}{S} = \mu\, dt + \sigma\, dz \tag{1}$$

where dS/S is the instantaneous proportional change in the price of the underlying asset, μ is the annual mean return, σ is the volatility, dt indicates an infinitesimal unit of time and dz represents Brownian motion, a Gaussian random variable with mean 0 dt, variance 1 dt, and independent increments over time.

Starting from an initial value S_0, the return over the period from 0 to T is given by

$$R = \ln(S_T / S_0)$$

and R has a Normal distribution, with

$$Mean = (\mu - \sigma^2/2)\, T$$
$$Standard\ deviation = \sigma\sqrt{T}$$

The logic of option pricing theory is that, under the assumptions of the model, if one knows the true volatility along with the other, observable, parameters, there exists a dynamic self-financing trading strategy that can be followed from the present until the expiration date that will exactly replicate the payoff on any given option. The volatility parameter needed to implement that strategy is the volatility that will be exhibited over the entire remaining lifetime of the option. That is, the parameter that must be forecasted is the standard deviation of the log price relatives for the underlying asset from now until expiration day, which may be a period of years for a long maturity contract. Generalizations of the basic Black-Scholes framework to allow for volatility that varies (nonstochastically) over time lead to a very similar result: the volatility parameter that goes into the model is the square root of the average annualized return variance over the option's lifetime.

When an asset's price follows the constant volatility lognormal diffusion model of equation (1), σ can be estimated easily from historical data. The difficulty arises because actual prices do not follow (1) exactly, so that price behavior may change over time and differ over intervals of different lengths. Moreover, the ways in which (1) fails in practice are not established and regular enough for an alternative model to have become widely accepted. It is common, therefore, to compute volatility using historical price data as if (1) were correct but to adjust the estimation methodology, or the volatility number it produces, in various ways to offset known or suspected problems. The resulting point estimate for σ then becomes the volatility input to the Black-Scholes model or another fixed volatility valuation equation. Even though true volatility may be believed to vary stochastically over time, Black-Scholes is familiar and easier to manipulate than any valuation models that adjust for random volatility formally.

The Standard Historical Volatility Estimate

Consider a set of historical prices for some underlying asset that follows the process defined in equation (1):

$$\{ S_0 , S_1 , \ldots , S_T \}$$

We begin by computing the log price relatives, i.e., the percentage price changes expressed as continuously compounded rates

$$R_t = \ln (S_t / S_{t-1}) , \text{ for t from 1 to T}$$

The estimate of the (constant) mean μ of the R_t is the simple average

$$\bar{R} = \frac{\sum R_t}{T} \tag{2}$$

The variance of the R_t is given by

$$v^2 = \frac{\sum (R_t - \bar{R})^2}{(T - 1)} \tag{3}$$

Annualizing the variance by multiplying by N, the number of price observations in a year and taking the square root yields the volatility,

$$\sigma = \sqrt{N v^2} \tag{4}$$

If the constant parameter diffusion model of (1) is correct, the above procedure gives the best estimate of the volatility that can be obtained from the available price data. This number then become the forecast for volatility going forward, over a time horizon of any length.

3. Problems with the Lognormal Diffusion Model

Unfortunately, prices for actual securities do not follow (1) in practice.

Time-Varying Volatility

One major problem is that volatility clearly changes over time. As an illustration consider Figure 1, which plots the volatility of the U.S. Treasury 20 year bond yield. Taking monthly data on the bond yield from January 1972 through July 1993, we used equations (2) through (4) to compute the realized volatility over the previous 36 months and plotted the resulting time series.

These estimates (which are the volatility forecasts that would have been made at each point in time, based on the available historical data) certainly do not appear to be just a constant parameter plus sampling noise. Indeed, in the case of Treasury yields, we have a good explanation for the sharp rise in volatility that occurred after 1979, when the Federal Reserve formally changed its operating policies to allow wider fluctuations in rates.

Serial Correlation in Returns

One virtue of empirical research using financial data is that it is often available in enormous quantity, in many cases down to the intraday level of individual transactions. But while this would permit calculations with extraordinary accuracy if (1) were exactly correct, the value of using all available data is severely limited by the fact that prices and returns for many securities appear to have some serial correlation and other distortions at both short and long intervals.

Apparent serial dependence may arise from several sources. Equation (1) is meant to describe the evolution of the equilibrium market price, but price data normally are produced only by transactions. Since the marketmaking process typically involves bid and offer quotes around the equilibrium, recorded transactions prices can show extremely high negative serial correlation, as they bounce back and forth between trades at the bid and the ask, while the equilibrium price is essentially unchanged.

Brown[1990] provides a striking example of the impact of the effect of the differencing interval on estimated volatility for Standard and Poor's 500 Index futures. Using closing price data for the month of October 1986, the annualized volatility of the December S&P future was calculated to be .158. Futures data are not available transaction by transaction, but they are recorded once a minute during the trading day. Using the 9185 minute by minute price observations, volatility for the same time period was calculated to be .372. When the sampling interval was lengthened to 1 hour, estimated volatility dropped to .324.

Two things are evident from these results. First, the choice of differencing interval can make a huge difference in the measured volatility. Second, the fact that prices do not obey equation (1) exactly at very short observation intervals means that the existence of vast amounts of intraday price data is probably not very useful in improving volatility forecasts.

Positive serial correlation is often found in reported daily closing prices for equities and other securities. This is generally thought to be due to the "nontrading effect." When transactions for less liquid securities lag behind movements in their equilibrium prices, the full impact of a large information event tends to get spread over two or more days' recorded closing prices. The resulting positive autocorrelation in returns will reduce estimated volatility.[1]

Sampling at longer intervals is an easy way to limit the effect of serial dependence a high frequencies, but it also means using fewer data points, which increases sampling error. The best choice of sampling frequency must depend on the statistical properties of the particular price series under consideration. One reasonable principle would be that if prices show no serial dependence at a given interval, there is no statistical reason to sample less frequently. In our empirical investigation of volatility forecasting procedures below, we use monthly observations.

One final point on this topic is that work by Fama and French [1988], Poterba and Summers [1988] and others has found evidence of significant negative autocorrelation over periods of several years in stock prices. While this will have little effect on volatility forecasting for most exchange traded equity options, it should affect valuation of equity warrants, as well as many newer derivative products, including LEAPS and FLEX options, and similar long maturity over-the-counter instruments.[2] Long run

---

[1] This is one reason that stock index futures prices often appear to have higher volatility than the underlying stock index: closing futures prices have virtually no measured serial correlation.

[2] LEAPS are exchange-traded options on individual stocks with maturities up to three years. FLEX options are stock index contracts traded at the Chicago Board Options Exchange, whose terms can be negotiated; maturities can be up to five years. The Chicago Board of Trade has also recently introduced a FLEX-type instrument based on T-bond futures and other long maturity option contracts

negative autocorrelation will not have much impact on the cost of option replication, since hedging costs are largely determined by short run price variability. But other risk measures, like the probability that an option which is initially deep out-of-the-money will end up in-the-money may be affected much more.

Nonnormal Returns Distributions

A third way in which actual securities returns differ from equation (1) is the well-documented problem of "fat tails." Equities and many other securities exhibit more large price changes than is consistent with the lognormal diffusion model. Some researchers have attempt to deal with the empirical returns distribution by fitting constant elasticity of variance models or other specifications that allow for this. See Macbeth and Merville [1980], for example. There are two problems with this approach. One is that except for special cases, the use of a more complex stochastic process for returns makes derivatives valuation substantially harder. But more importantly, it may not help solve the volatility forecasting problem at all, since the parameters of the alternative process must now be assumed to be stable and accurate estimates from past data, or another source are now required. There is no guarantee that the degree of tail fatness would prove to be easier to estimate, or more stable over time, than volatility is.

Perhaps the most important way in which this issue confronts actual participants in options markets is in deciding how to handle major events that are "unique" in some

---

are available at different exchanges.

sense. Dealing with the effect of an outlier like October 19, 1987 in estimating stock volatilities is a prime example of the difficulty. Figure 2 shows a rolling volatility for the S&P 500 Index, calculated at each point from daily closing prices over the previous 24 months. While there were clearly variations in the volatility from year to year during the 1960's and 1970's, the large jump after October 1987 is extreme. One day's price drop caused a huge increase in estimated volatility. This presented large problems for participants in the options markets in 1988, because after the Crash, the day to day variation in equity returns dropped quickly back to rather low levels, more consistent with a volatility of around 15 percent. In that circumstance, should market participants nevertheless have used a volatility close to the "historical" estimate of around 27 percent? Or should they have used 15 percent, essentially acting as if the Crash had never happened? What if they had needed a long term volatility forecast, in order to price a warrant with a maturity of several years? Whatever choice is made in such a case is bound to be arbitrary. A reflection of the arbitrariness of the decision is the "echo" effect of the Crash: exactly 24 months and one day after October 19, 1987, that data point drops out of the calculation and the historical volatility drops overnight to under 15 percent.

Noisy Estimates of the Mean

A different issue arises with respect to the estimate of the mean return. Since volatility is measured in terms of deviations from the mean return, an inaccurate estimate of the mean will reduce accuracy of the volatility calculation. Unfortunately, the sample average return, as shown in equation (2), is a very noisy estimate of the true parameter μ. With a diffusion process, sampling more frequently reduces the sampling error of the volatility estimate (as long as serial dependence does not appear), but the accuracy of the mean estimate depends only on the first price and the last price of the sample. This is easily seen by substituting for the R_t in (2):

$$\overline{R} = \frac{\sum R_t}{T} = \frac{\sum (\ln S_t - \ln S_{t-1})}{T} = \frac{\ln S_T - \ln S_0}{T} \qquad (5)$$

All of the prices observed between S_0 and S_T drop out of the calculation. Moreover, under equation (1), the standard deviation of ($\ln S_T - \ln S_0$) is $\sigma\sqrt{T}$, so the standard error of R as an estimate of μ is $\sigma/\sqrt{T}$. This only depends on the <u>length</u> of the sample period, and not on the number of prices observed during that period.

For example, suppose the volatility of the price process is 20 percent and we have 4 years of historical daily price data. The standard error of the sample average around the true mean is $20 / \sqrt{4} = 10$ percent. So, if the average annual return were, say, 15 percent in our 4 year sample (with more than 1000 data points), a 95 percent confidence region for the true mean would still range from -5 percent to +35 percent.

Equity option traders often estimate volatilities from 1 to 3 months of data. One month of prices for a stock with a volatility of .25 will yield a sample mean whose standard deviation around the true value is over 85 percent. In other words, roughly one third of the time, the trader's volatility estimate for a typical stock will be computed in terms of the deviations of its returns from a sample mean that is more than 85 percentage points above or below the correct value on an annualized basis!

Given that degree of imprecision, many researchers consider it more accurate simply to assume a value for the mean that is consistent with financial theory rather than trying to estimate the mean from the data. This amounts to a kind of Bayesian approach, based on the notion that the principles of finance allow us to place tighter bounds on an asset's true mean return than classical statistics does. For instance, we do not think the S&P 500 index should ever have an equilibrium ex ante mean return that is negative, regardless of the sample mean in a given set of data.

One viable approach with daily data is simply to impose a mean of 0. See Black [1976], for example. Another possibility is to use the risk free interest rate. Fortunately, the estimate of the volatility does not depend very heavily on the mean. Thus, while it is extremely difficult to obtain an accurate mean estimate from the data, the real problem as far as volatility calculation is concerned is to avoid using extreme sample mean returns that will periodically be produced from short data samples. A corollary of this principle is that if one is interested in volatility, using elaborate models for mean returns, e.g., allowing the risk premium to vary over time, is unlikely to be worth the effort in terms of any improvement in accuracy. Below, we will examine the effect of imposing a mean return of zero on empirical forecast accuracy in our volatility estimation.

Estimating volatility in practice

Given that actual securities prices do not come from a constant volatility lognormal diffusion process, computing historical volatility as shown in equations (2) - (4), is no longer theoretically optimal. But, while the problems we have just mentioned are well-known, option traders, and many academic researchers as well, typically ignore them and calculate historical volatility estimates by the most basic method.

The normal (though not necessarily optimal) way most traders deal with the fact that volatility changes stochastically over time is to use only recent observations in the calculation and discard data from the distant past. It then becomes necessary to decide how much past data to include in a historical sample. There is a tradeoff between trying to examine a large sample and trying to eliminate data that are so old as to be obsolete. One consideration in making this choice may be the length of the forecasting horizon. In trying to predict volatility over the next 3 months, it is plausible that one might prefer a short sample of more recent data, perhaps just the last 6 to 12 months, while to forecast volatility for the next 3 years, a longer historical sample might be called for. We examine these issues empirically in the next section.

4. The Forecasting Performance of Historical Volatility

The most common method of producing volatility forecasts from historical data is simply to select a sampling interval and the number of past prices to include in the calculation and then to apply equations (2) - (4), (making ad hoc adjustments when the procedure appears to be giving inappropriate answers). The thought that it may make sense to adjust the length of the historical sample for different forecasting horizons suggests that it would be worthwhile to examine the issue empirically.

Consider estimating volatility from k past prices in order to forecast the volatility that will be experienced over the next T periods. This might be called, simply, the (k,T) model. The volatility estimate from that procedure is given in equation (6)

$$\sigma = \left(\frac{\sum_{\tau=1}^{k} (R_{t-\tau} - \overline{R})^2}{k-1} \right)^{\frac{1}{2}} \tag{6}$$

We have used the (k,T) procedure to construct time series of volatility forecasts from monthly data for a large number of financial series, including interest rates, stock prices, and exchange rates. In this paper we report results for a selection of the most important series: the S&P 500 index, 3 month Treasury bill rates, 20 year Treasury bond yields, and the Deutschemark/dollar exchange rate. The length of the data samples varies, with the longest starting in 1947, while the exchange rate data only begin in 1970, after the era of floating rates. Table 1 provides details about the data series.

We want to analyze the accuracy of the (k,T) procedure, as a function of its parameters: the lengths of the historical sample, k, and the forecasting horizon, T. In the results reported below, we examine k and T values of 6, 12, 24, 36, 48, and 60 months. Forecast accuracy is measured by the root mean squared forecast error. For the results to be comparable across all different k and T values, the forecasts must cover exactly the same time periods. Thus, if t_{beg} and t_{end} represent the beginning and ending dates for a given data series, while t_{1st} and t_{last} are the first and last dates for which volatility forecasts are calculated, then we set

$$t_{1st} = t_{beg} + 59$$

and

$$t_{last} = t_{end} - 60$$

to allow up to five years of historical data prior to the first forecast period and five years to computing realized volatilities following the final forecast period.

Both historical and realized volatilities are computed around the an assumed value of zero for the mean returns. We present results later to show how much difference this makes to forecast accuracy. We have made no adjustment for October 1987, or any other unusual events. The distorting effect of the Crash is more limited here than in Figure 2, because we are using monthly data. A 20+ percent drop in stock prices in one day obviously produces a much larger annualized volatility than the same price change over a month.

The procedure therefore works as follows. Beginning at the data point t_{1st}, returns (i.e., log price relatives) over the previous k months are computed and the historical volatility is calculated around an assumed mean of 0. This will be the forecast as of date t_{1st} for volatility over all future horizons. Realized volatility is then computed over the next T periods, for all T values, and the forecast errors are recorded. The period is advanced one month, and the process repeated, with one data point dropped off the beginning of the sample and one new one added at the end, for the historical sample and each T period forecasting horizon. The procedure continues until forecasts and forecast errors at all horizons have been produced for all dates from t_{1st} to t_{last}. The root mean squared errors are then calculated for each (k,T) pair.

This procedure obviously results in potentially quite large autocorrelations in the forecast errors, because each month's forecast and realized volatility are computed from a data sample that only differs from that used in the previous month by two data points. We have made no attempt to adjust for this. Lack of independence in a time series does not change the estimate of the mean of the series of squared errors (which is still the sample MSE).[3] One way to think about this procedure is that we are looking at the forecasting performance that would have been experienced by a financial institution making markets in derivatives in every month over the entire sample period and consistently using the (k,T) approach to estimate future volatility.

Tables 2 - 5 contain the results for the four series we are examining here, and they are displayed visually in Figures 3 - 6. To illustrate them, consider Figure 3 which shows the forecasting accuracy of the (k,T) procedure in predicting the volatility of the S&P 500 index. The curve marked with dark boxes is the forecast accuracy of six month forecasts made from varying amounts of past data. The first point, for example, shows that using realized volatility over the previous six months as a forecast of what will be observed over the next six months gives a very inaccurate answer, with a root mean squared error of .0704. (Note that the average volatility of the S&P index was only .132

[3] The standard error on the MSE computed under the assumption of independence would be biased, but we do not calculate standard deviations for the root mean squared forecast errors.

over the whole sample, so the RMS forecast error is about half of the realized volatility.)

Volatility calculated from the prior 12 months gives a better estimate for the next 6, but the k value that produces the most accurate 6 month prediction is 60, the maximum considered, with RMSE of 060. One reason these results look so bad is that the 6 month volatility has a great deal of sampling noise, being constructed from only 6 observations. If we really needed a 6 month volatility, we should probably use daily or weekly data in the calculations. Forecasting at the two year horizon is more accurate, again reaching the minimum RMSE with 5 years of historical data. Five year forecasts turn out to be the most accurate, and in results not shown here, even longer historical samples were found to produce still better predictions.

The results for the other financial time series are broadly similar to what we have just seen for the S&P index, although 3 month T-bills show the somewhat anomalous result that the most accurate forecast of volatility over a 5 year horizon comes from historical volatility computed over only the previous 2 years of data.

In all cases, the predictability of average volatility seems to improve markedly for longer forecasting horizons. For example, the lowest RMSEs obtained for the S&P index were .0580 at 6 months, .0417 at two years, and as low as .0269 for five years. This was quite unexpected: We anticipated that the further in the future a forecast had to go, the less accurate it would become, but the opposite is clearly the case here. This suggests that volatility exhibits mean reversion over long horizons, so that (unlike a random walk) extreme levels that might occur in a short period tend to average out over time.

Another clear result, for all of these series (except T-bills), as well as others that are not shown here, is that the most accurate volatility estimates appear to come from the longest samples of past prices: the lowest RMSE is produced by the five year estimates. This is a much longer historical sample than is typically used by market participants, especially for a forecast horizon of two years and under.

To these two rather surprising results, we might add a third conclusion suggested by this analysis, which is that the predictability of volatility over the long term seems to be quite good (except, perhaps, for T-bills once again). The fact that the RMSE of a five year volatility forecast constructed from historical price data on the S&P index is as low as 2.7 percent, seems quite remarkable.[4]

The Effect of Estimating the Mean on Forecast Accuracy

As mentioned above, financial theory may be able to give us a better estimate of the true mean return than can be obtained from a limited amount of past returns data. In the results we have just discussed, volatility was computed around a mean assumed a priori to be zero. To examine the difference this makes in forecasting performance, we replicated the analysis of Tables 2 - 5, with volatility computed around the sample mean.

Table 6 shows results on the percent reduction in RMSE that constraining the mean produced, for a selection of historical sample and forecast horizon pairs. For example, when S&P 500 volatility is calculated from the previous 12 months of data and used to

[4] For the S&P 500, using 10 years of past data and forecasting 10 years ahead produced an RMSE of 2.3 percent over the 45 years spanned by our data.

forecast over the next 12 months, the RMSE is 6.8 percent lower when the mean is not estimated. Only in a handful of cases did constraining the mean lead to an increase in forecast RMSE. As one would anticipate, the effect is larger when shorter time periods are involved. These results indicate that even over quite long sample periods, more accurate forecasts may be obtained by computing volatility around zero rather than around the sample mean.

5. Forecasting Long Term Volatility with Models of the ARCH Family

Volatility needs to be forecasted because it changes over time. The procedures discussed in the previous sections are essentially ad hoc approaches that are based on a constant volatility framework. However, in recent years a number of related formal models for time-varying variance have been developed. In this section, we will discuss using these models to predict volatilities of asset returns.

Consider the following model for returns.

$$R_t = E[R_t] + \epsilon_t \qquad E[\epsilon_t] = 0 \quad Var[\epsilon_t] = \sigma^2_t \tag{7}$$

Although variants exist in which the mean in equation (7) is a function of the variance, we will restrict ourselves here to constant mean models and focus on the process followed by σ.

The simplest, of course, is the constant volatility model

for which the standard variance fitting procedure in equations (2) - (4) applied to all

$$\sigma^2_t = C \tag{8}$$

available historical data is the appropriate estimation strategy.

The first time-varying volatility model is the Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982). Variance in period t is modelled as a constant plus a distributed lag on the squared residual terms from previous periods. An ARCH(q) specification involves q lagged residual terms. Equation (9) shows an ARCH(3) model.

$$\sigma_t^2 = c + a_1 \epsilon^2_{t-1} + a_2 \epsilon^2_{t-2} + a_3 \epsilon^2_{t-3} \tag{9}$$

For stability, the sum of the a coefficients should be less than 1.0.

In principle, q may be any number, but generally only a few lags are used. Cases requiring variance effects that are expected to be of longer duration are better suited to the Generalized ARCH, or GARCH, framework developed by Bollerslev (1986). A GARCH model explains variance by two distributed lags, one on past squared residuals to capture high frequency effects, and the second on lagged values of the variance itself, to capture longer term influences.

The simplest, and most commonly used, member of the GARCH family is the GARCH(1,1) model shown in equation (10).

$$\sigma^2_t = C + a_1 \sigma^2_{t-1} + b_1 \epsilon^2_{t-1} \tag{10}$$

The GARCH(1,1) model embodies a very intuitive forecasting strategy: the variance expected at a given date is a combination of a long run variance and the variance expected for last period, adjusted to take into account the size of last period's observed shock.

Since the expected value of ϵ^2 is σ^2, the long run steady state value for the variance is given by

$$\sigma^2_{Long\,Run} = \frac{c}{1 - a_1 - b_1} \tag{11}$$

Here, long run stability requires $a_1 + b_1 < 1.0$.

The GARCH model has the virtue that it is quite simple but it captures the kind of time variation that seems plausible for variances. However, GARCH has two shortcomings. One is that it can be hard to fit, especially when more than one lag on each variable is involved. It also restricts the impact of a shock to be independent of its sign, whereas there is evidence of an asymmetric response for some markets, notably the stock market. Stock return volatility increases following a sharp price drop, but a price rise may even lead to lower volatility.

To deal with these problems, Nelson (1991) proposed Exponential GARCH (EGARCH), which models the log of variance, so that the right hand side of (10) can become negative without creating a problem. EGARCH also allows for an asymmetric reaction to positive and negative shocks. In this paper we will only present results for the GARCH specification.[5]

Problems with ARCH-family Models

These models have been widely examined and applied in economics and finance.[6] Nearly all of the work with them has focused on in-sample explanation of variance movements, rather than forecasting per se. The model is normally fitted by assuming a density function for the ϵ terms--the normal is by far the most common, for log price relatives--and deriving parameter estimates by maximum likelihood estimation.

Although financial economists automatically model security returns as lognormal, we also know that this model does not fit perfectly. The fact that "too many" large price changes are observed for a lognormal distribution is well known. The explanation for this fact is not agreed upon; one possibility is that the returns distribution appears to have fat tails because it really involves prices drawn from a distribution that is lognormal at every point in time, but with time-varying variance. If that is the source of the problem, an ARCH-type model may resolve it.

An important problem in implementing ARCH-family models is simply doing the estimation. These models seem to require quite a large number of observations before they behave well. Likelihood surfaces may be quite flat, making finding a maximum

---

[5] The out-of-sample performance of the EGARCH model for short term volatility forecasting is examined for five major financial assets by Cumby, Figlewski, and Hasbrouck (1993).

[6] See, Bollerslev (1992) for a review of their applications in finance.

difficult, or the maximum for a given sample may lie outside the theoretically acceptable region (negative coefficients or values greater than 1.0, implying long run instability of the system).

ARCH models in particular present the problem that one might like to allow a fairly long distributed lag on past shocks, but that would entail fitting a large number of parameters. Moreover, as more past squared residuals are added to the system, some of the estimated parameters are likely to become negative. Negative parameters can present great difficulties both for estimation and for forecasting, because a particularly large ϵ may drive the entire fitted variance negative. A GARCH formulation has the advantage that one fits only a small number of parameters, increasing the likelihood that they will all be well-behaved, but disturbances over all recent periods can enter into the calculation.

All ARCH-type models share three significant shortcomings as forecasting tools. First, they all seem to need a large number of data points for robust estimation. Second, they are subject to the general problem that the more complex any model is and the larger the number of parameters it involves, the better it will tend to fit a given data sample, and the quicker it will tend to fall apart out of sample. For any procedure to be useful in forecasting, it must be sufficiently stable over time that one can fit coefficient estimates on historical data and be reasonably confident that the model will continue to hold as time goes forward.

The third problem is that all three models essentially focus on variance one step ahead. They are not designed to produce variance forecasts for a long horizon. For example, consider the forecasts from a GARCH(1,1) model.

$$\begin{aligned} \sigma^2{}_t &= C + a_1 \sigma^2{}_{t-1} + b_1 \epsilon^2{}_{t-1} \\ E_t [\sigma^2{}_{t+1}] &= C + a_1 \sigma^2{}_t + b_1 E_t [\epsilon^2{}_{t+1}] \\ &= C + (a_1 + b_1) \sigma^2{}_t \qquad (12) \\ &\vdots \\ E_t [\sigma^2{}_{t+k}] &= C \sum_{s=0}^{k-1} (a_1 + b_1)^k + (a_1 + b_1)^k \sigma^2{}_t \end{aligned}$$

Because the forecast for variance in period t+1 involves the unknown value of the period t squared disturbance, we must substitute its expected value as of period t, which is simply the period t model variance. It is clear that once one is forecasting more than a few periods ahead, the forecasts can not incorporate any new information from the (unknown) future disturbances, and will simply converge to the long run variance at a rate that depends on the value of $(a_1 + b_1)$.

6. Forecasting Performance of the GARCH(1,1) Model

The discussion in the last section makes it clear that the GARCH formulation has several advantages over ARCH for our purposes. In order to evaluate the ability of GARCH to produce accurate out of sample long run volatility forecasts, we attempted to fit GARCH(1,1) models to the monthly data examined above. The first series we looked at was returns on the S&P 500 stock index.

A sequence of GARCH(1,1) models were to be fitted to a rolling sample of returns. As above, we wanted to explore the effect of changing the amount of past data on forecasting accuracy at various horizons. Because the estimation is time consuming, we reestimated the parameters only once per year, rather than every month. The smallest amount of past data we attempted to use was 5 years, i.e., 60 monthly observations.

For example, in the first experiment with the S&P 500 index we tried to fit a GARCH(1,1) on the monthly returns from January 1948 through December 1952. Those parameters would be used to construct out-of-sample GARCH forecasts for the first 6 months of 1953 (as shown in equation (12)). The monthly predicted variances would then be averaged, and the predicted average variance over the 6 month period turned into an annualized volatility, which could be compared to the realized "average" volatility over that period.[7] In a similar fashion, volatility forecasts for 12 month and 24 month horizons would be produced at the same time.

Once the forecast for January - June 1953 was constructed, we would advance the sample 1 month, by incorporating the squared residual from the realized return for January and forecasting the February - July volatility. After 12 such out-of-sample forecasts were produced, we would refit the GARCH model, adding the realized returns for 1953 into the sample and dropping the same number of observations from the beginning, to keep a window of fixed size.

---

[7] What we have called the "average" volatility is actually the square root of the average of the monthly variance forecasts. Because of Jensen's Inequality, this will not be equal to the average of the predicted monthly volatilities, i.e., the square roots of the variances. However, it is the correct way to construct the volatility input to a European option pricing model when variance changes (nonstochastically) over time.

This procedure turned out to be infeasible, because it was extremely difficult to fit the model on as few as 60 data points. Of the first 36 five-year periods, the estimation routines in GAUSS were unable to converge on parameter values in 30 of them. When we increased data in the estimation to a rolling 10-year sample, we still failed to achieve convergence in 10 of the periods. We finally settled on an updating procedure of allowing the initial observations to remain in the sample until it contained 15 years of data, after which we would begin dropping observations as with the fixed window procedure. In cases where updated parameter estimates could not be fitted, we simply continued using the old parameter values to produce forecasts. While we were still not able to estimate parameters for the first two 10 year periods (which were therefore dropped), with this procedure only 5 of the subsequent estimations failed.

The difficulty in fitting the GARCH(1,1) models even on long data samples was not unique to the S&P 500 index. In fact, the S&P index gave us the least amount of trouble of the four data series. It was impossible to use samples as short as five years for any of the series we examined. For the 20 year Treasury bond yield, a 10 year fixed window failed to converge 11 out of 30 times, but allowing the window to expand to 15 years as before (and dropping the first two periods) reduced the number of failures to 6 in 28. We were not able to fit the basic GARCH(1,1) model at all for the 3 month Treasury bill rate, even with 15 years of data, or for the Deutschemark exchange rate.

Table 7 shows the root mean squared forecast errors for the GARCH(1,1) models that we were able to fit for the S&P 500 index and 20 year Treasury yields, and compares them to the RMSEs of historical volatilities computed over the previous 5 and 10 years. Forecasts of S&P volatility performed relatively well, achieving comparable RMSEs to the historical volatilities at all three horizons, although no apparent superiority.

The results are different for the GARCH predictions of Treasury bond yield volatility. Here, the GARCH post-sample forecasts are distinctly less accurate than historical volatility, and they get substantially worse for the longest horizon. There is apparently not enough stability in the model (as fitted to monthly data) for it to perform well out-of-sample in this market.

Our conclusion from these results is that attempting to allow for predictable time-variation in asset volatilities with a GARCH specification poses very difficult estimation problems, and does not appear to produce any superiority in accuracy over the much easier procedure of simply computing the historical variance over a long sample of past data. One additional thing we see in these results is that, at least for the S&P 500 index and for 20 year Treasury bonds, the 10 year historical volatility is even better than the 5 year estimate.

7. Conclusions

Applying modern option valuation theory requires the user to forecast the volatility of the underlying asset over the remaining life of the option. This is a formidable estimation problem for long maturity instruments. The standard statistical procedures using historical data are based on assumptions of stability, either constant variance, or constant parameters of the variance process, that are unlikely to hold over long periods.

In this paper, we have examined the empirical performance of historical variance estimators different and of the GARCH(1,1) model for forecasting volatility in important financial markets over horizons up to five years. We have found several surprising results:

* In general, historical volatility computed over many past periods provides the most accurate forecasts for both long and short horizons.

* Root mean squared forecast errors are substantially lower for long term than for short term volatility forecasts.

* It typically increases forecast accuracy to compute volatility around an assumed mean of zero rather than around the realized mean in the data sample.

* The GARCH model tends to be less accurate and much harder to use than the simple historical volatility estimator for this application.

References

Black, Fischer. Fischer Black on Options 1, Xeroxed, May 17, 1976.

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroscedasticity." Journal of Econometrics 31, 1986, pp. 307-327.

Bollerslev, Tim. Ray T. Chou and Kenneth F. Kroner. "ARCH Modeling in finance." Journal of Econometrics 52, 1992, pp. 5-59.

Brown, Stephen. "Estimating Volatility." Financial Options: From Theory to Practice, Figlewski, et al, eds., Homewood Ill: Business One Irwin, 1990, pp. 516-537.

Cumby, Robert, Stephen Figlewski and Joel Hasbrouck. "Forecasting Volatilities and Correlations with EGARCH Models." Journal of Derivatives 1, Winter 1993, pp. 51-63.

Engle, Robert. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." Econometrica 50, 1982, pp. 987-1000.

Fama, Eugene and Kenneth French. "Permanent and Temporary Components of Stock Prices." Journal of Political Economy 96, April 1988, pp. 246-73.

Macbeth, James and Larry Merville. "Tests of the Black-Scholes and Cox Call Option Valuation Models." Journal of Finance 35, May 1980, 285-3000.

Nelson, Daniel. "Conditional Heteroscedasticity in Asset Returns: A New Approach." Econometrica 59, 1991, pp. 347-370.

Poterba, James and Lawrence Summers. "Mean Reversion in Stock Prices: Evidence and Implications." Journal of Financial Economics 22, Oct. 1988, pp. 27-59.

TABLE 1

Dates and Sources of Data Series

Series	Dates	Source
Standard and Poor's 500 Stock Index	1/47 - 12/92	CRSP
	1/93 - 12/93	Bloomberg
3 Month U.S. Treasury Bill Yield	1/47 - 12/92	Salomon Bros.
	1/93 - 11/93	Bloomberg
20 Year U.S. Treasury Bond Yield	1/50 - 12/92	Salomon Bros.
	1/93 - 7/93	Bloomberg
Deutschemark Exchange Rate (DM per $)	1/70 - 10/92	International Financial Statistics (IMF)

TABLE 2

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE

S&P 500 STOCK INDEX

Overlapping Period: Jan'52 to Dec'88

Volatitity about 0

PAST OBS	FORECASTING HORIZON					
	6 MONTHS	12 MONTHS	24 MONTHS	36 MONTHS	48 MONTHS	60 MONTHS
6 MONTHS	0.0704	0.0637	0.0646	0.0644	0.0625	0.0599
12 MONTHS	0.0640	0.0586	0.0591	0.0572	0.0539	0.0516
24 MONTHS	0.0640	0.0583	0.0553	0.0504	0.0453	0.0435
36 MONTHS	0.0634	0.0563	0.0500	0.0435	0.0386	0.0363
48 MONTHS	0.0607	0.0524	0.0447	0.0382	0.0330	0.0299
60 MONTHS	0.0580	0.0493	0.0417	0.0349	0.0291	0.0269
AVG REALIZE	0.132	0.135	0.139	0.142	0.144	0.145

TABLE 3

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE

Treasury Bill- 3 Months

Overlapping Period
Jan'52 to Dec'87

Volatility about zero

PAST OBS	FORECASTING HORIZON					
	6 MONTHS	12 MONTHS	24 MONTHS	36 MONTHS	48 MONTHS	60 MONTHS
6 MONTHS	0.2616	0.2696	0.2710	0.2721	0.2656	0.2625
12 MONTHS	0.2691	0.2664	0.2629	0.2627	0.2542	0.2532
24 MONTHS	0.2655	0.2594	0.2572	0.2516	0.2455	0.2468
36 MONTHS	0.2739	0.2660	0.2573	0.2514	0.2471	0.2490
48 MONTHS	0.2831	0.2710	0.2574	0.2523	0.2494	0.2520
60 MONTHS	0.2861	0.2714	0.2581	0.2542	0.2523	0.2506
AVG REALIZE	0.293	0.308	0.320	0.324	0.325	0.327

TABLE 4

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE

Overlapping Period:
Jan'55 to Jul'88

T-BOND 20 YR.
Volatitity about 0

PAST OBS	FORECASTING HORIZON					
	6 MONTHS	12 MONTHS	24 MONTHS	36 MONTHS	48 MONTHS	60 MONTHS
6 MONTHS	0.0596	0.0585	0.0578	0.0584	0.0596	0.0608
12 MONTHS	0.0574	0.0540	0.0530	0.0533	0.0544	0.0556
24 MONTHS	0.0556	0.0514	0.0494	0.0498	0.0509	0.0507
36 MONTHS	0.0550	0.0505	0.0483	0.0486	0.0482	0.0471
48 MONTHS	0.0560	0.0512	0.0487	0.0475	0.0461	0.0452
60 MONTHS	0.0572	0.0524	0.0482	0.0461	0.0448	0.0433
AVG REALIZE	0.107	0.111	0.114	0.116	0.117	0.118

TABLE 5

Overlapping Period
Jan 75 to Oct 87

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE

DM

Volatility about Zero

PAST OBS	FORECASTING HORIZON					
	6 MONTHS	12 MONTHS	24 MONTHS	36 MONTHS	48 MONTHS	60 MONTHS
6 MONTHS	0.0613	0.0515	0.0509	0.0502	0.0498	0.0484
12 MONTHS	0.0520	0.0425	0.0425	0.0409	0.0404	0.0377
24 MONTHS	0.0492	0.0419	0.0402	0.0377	0.0341	0.0304
36 MONTHS	0.0504	0.0430	0.0398	0.0337	0.0286	0.0240
48 MONTHS	0.0508	0.0420	0.0346	0.0272	0.0214	0.0169
60 MONTHS	0.0476	0.0359	0.0275	0.0195	0.0143	0.0109
AVG REALIZE	0.111	0.114	0.117	0.119	0.121	0.122

TABLE 6

Percent Reduction in Forecast RMSE from Computing Volatility around Zero rather than the Sample Mean

Past Obs	Forecast Horizon	S&P 500	3 Month T-Bills	20 Year T-bonds	Deutsche-mark
6	6	-13.3	-2.8	-6.0	-12.6
12	12	-6.8	-3.9	-3.9	-4.3
36	12	-5.2	-3.3	-1.8	-2.5
36	36	-3.5	-2.9	1.5	-1.5
60	12	-5.4	-1.7	-1.1	-2.2
60	36	-3.9	-1.2	0.9	-0.5
60	60	-3.2	-0.7	0.0	-4.4

TABLE 7

Out-of-Sample Root Mean Squared Errors GARCH(1,1) Forecasts versus Historical Volatility

Standard and Poor's 500 Stock Index

Months forecasted: Jan 1960 - Dec 1992
Observations: 386
Successful estimations: 27
Failed estimations: 5

Horizon	GARCH RMSE	5 Year Historical RMSE	10 Year Historical RMSE
6 months	.0656	.0658	.0640
12 months	.0530	.0530	.0534
24 months	.0461	.0485	.0461

Yield to Maturity on 20-Year U.S. Treasury Bonds

Months forecasted: Jan 1963 - Dec 1990
Observations: 288
Successful estimations: 22
Failed estimations: 6

Horizon	GARCH RMSE	5 Year Historical RMSE	10 Year Historical RMSE
6 months	.0597	.0616	.0606
12 months	.0596	.0547	.0521
24 months	.0731	.0480	.0442

Note: GARCH models were fitted on a minimum of 120 months of historical data. Models were reestimated every 12 months and new data points were added to the estimation sample. Data points were dropped from the beginning of the sample only when the sample size exceeded 180 (15 years).

FIGURE 1

YIELD VOLATILITY FOR LAST 36 MONTHS
20 YEAR US TREASURY BOND

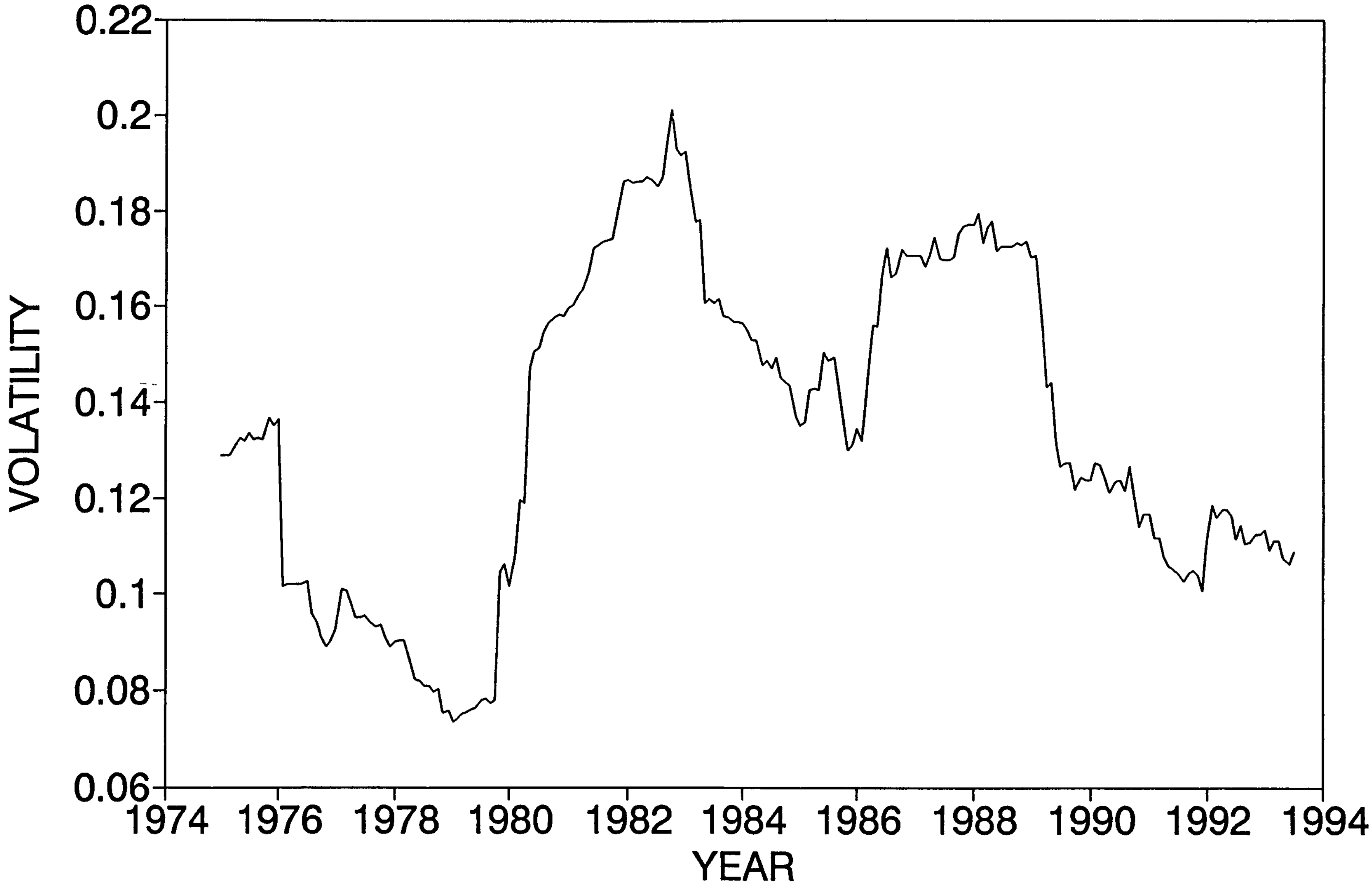
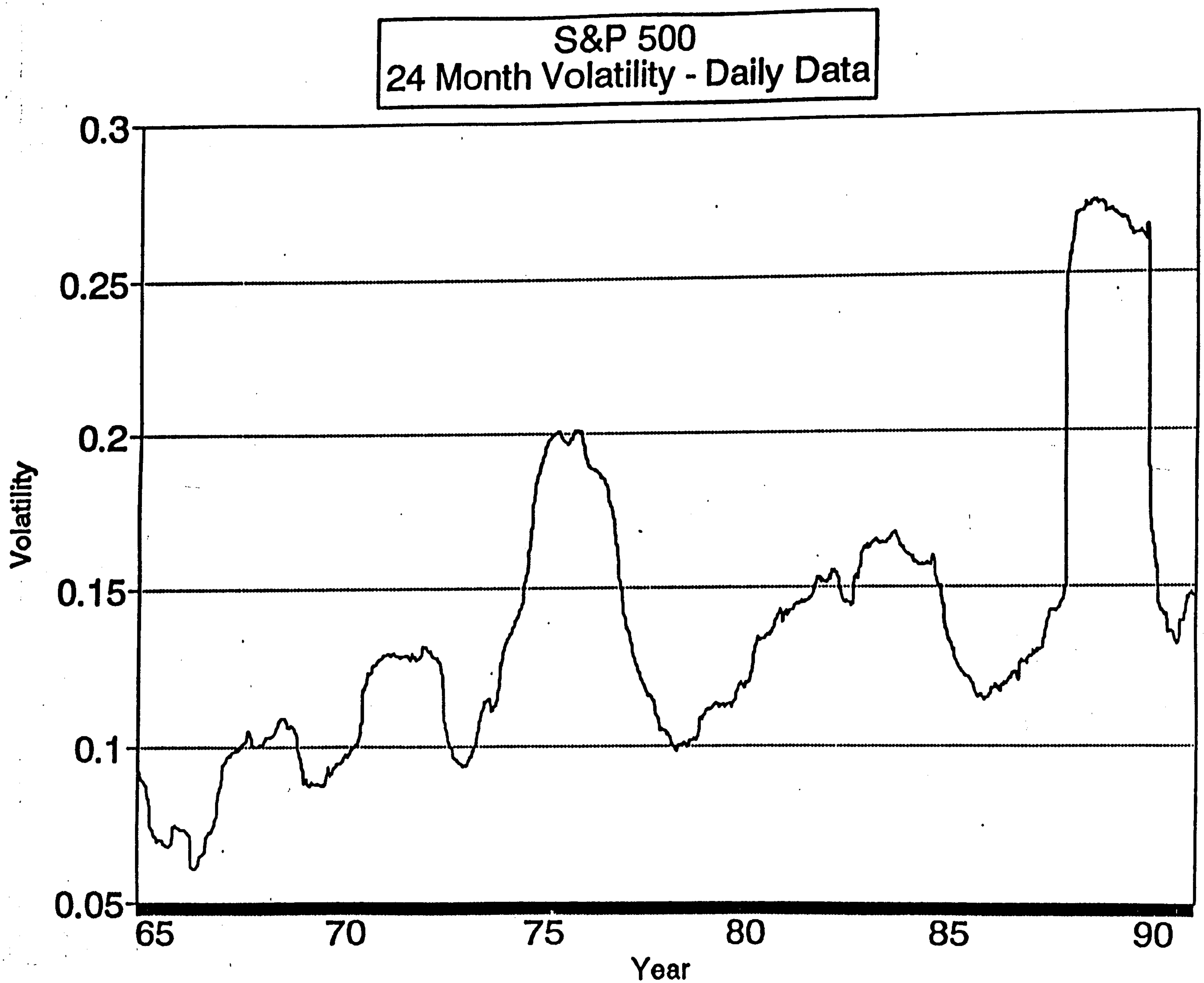


FIGURE 2



S&P 500
24 Month Volatility - Daily Data
Volatility
0.3
0.25
0.2
0.15
0.1
0.05
65
70
75
80
85
90
Year

FIGURE 3

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE
S&P 500 STOCK INDEX

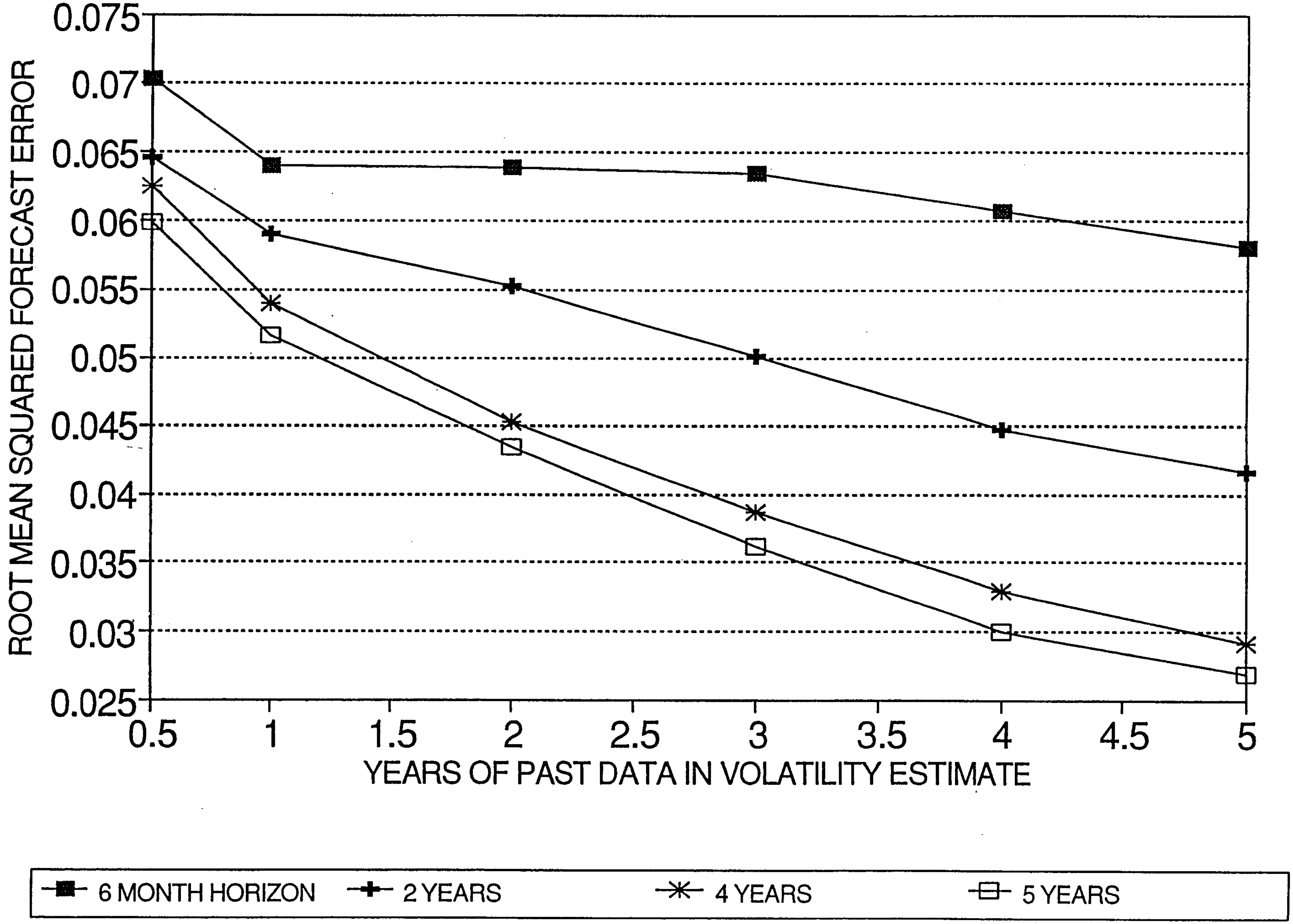


FIGURE 4

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE

Treasury Bill- 3 Months

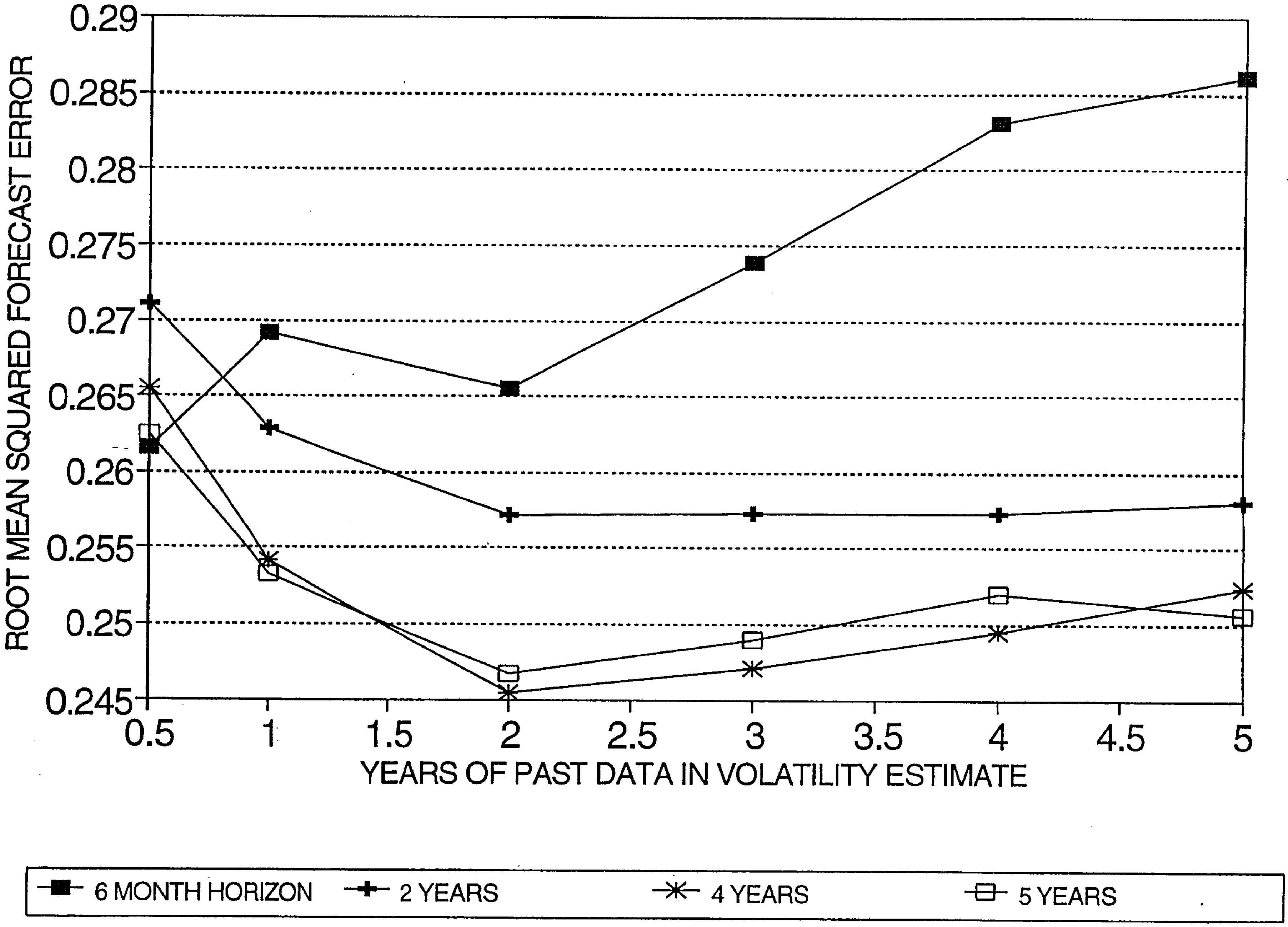


FIGURE 5

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE
T-BOND 20 YR.

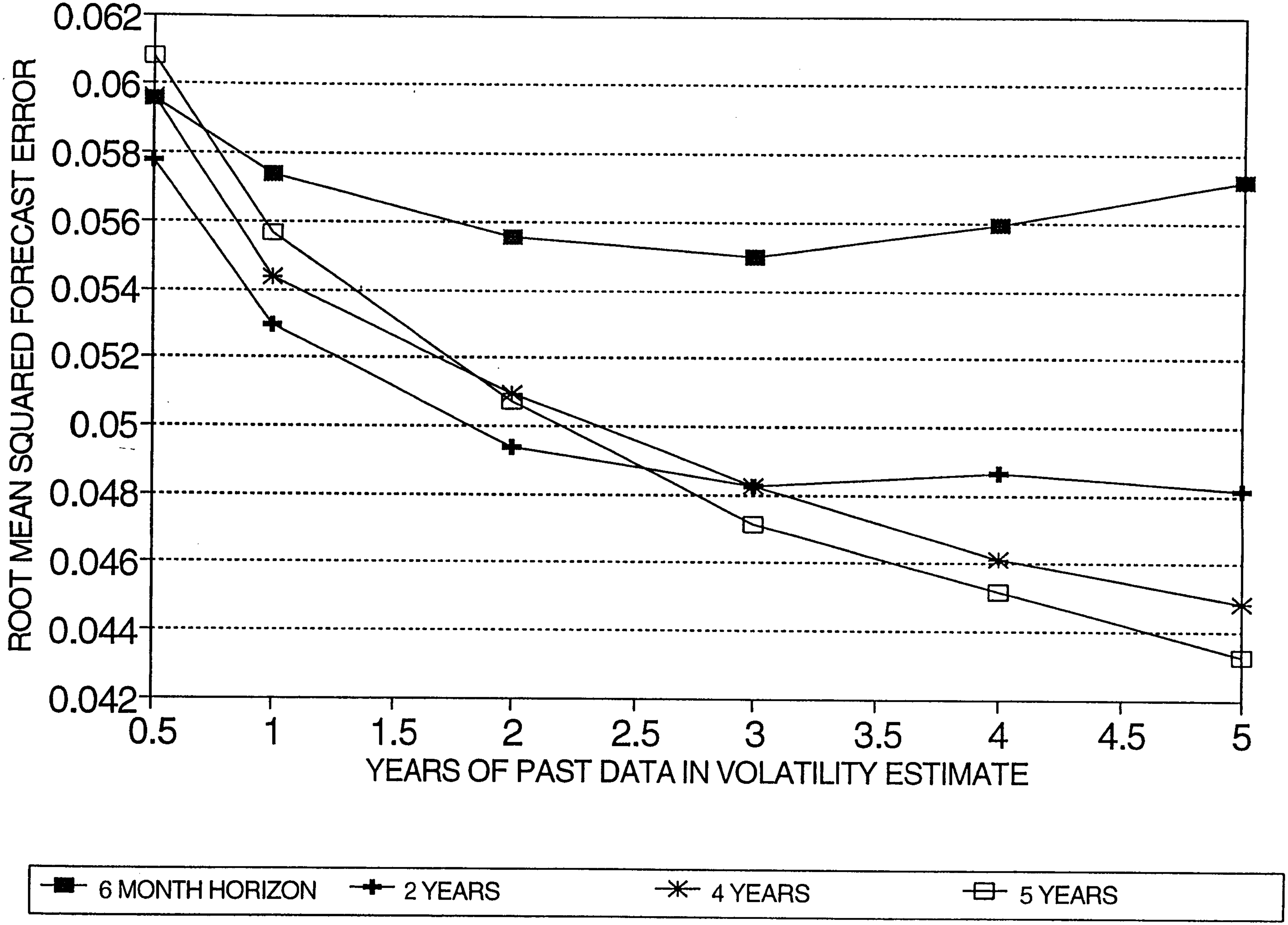


FIGURE 6

PERFORMANCE OF HISTORICAL VOLATILITY ESTIMATE
DM