FORECASTING VOLATILITY

by

Stephen Figlewski

FINAL DRAFT

April 24, 2004

New York University Stern School of Business
44 West 4th Street, Suite 9-160
New York, NY 10012-1126

212-998-0712 Voice
212-995-4220 FAX

sfiglews@stern.nyu.edu

PREFACE

This monograph puts together results from several lines of research that I have pursued over a period of years, on the general topic of volatility forecasting for option pricing applications. It is not meant to be a complete survey of the extensive literature on the subject, nor is it a definitive set of prescriptions on how to get the best volatility forecast. While at the outset, I had hoped to find the Best Method to obtain a volatility input for use in pricing options, as the reader will quickly determine, it seems that I have been more successful in uncovering the flaws and difficulties in the methods that are widely used than I have been in determining a single optimal strategy myself.

Since I am not revealing the optimal approach to volatility forecasting, the major value of this work, if any, is more to share with the reader a variety of observations and thoughts about volatility prediction, that I have arrived at after investigating the problem from a number of different angles. Two major themes emerge, both having to do with the connection, or perhaps more correctly, the possibility of a disconnection between theory and practice in dealing with volatility prediction and its role in option valuation. Two general classes of theories are involved.

First, there is the statistical theory involved in modeling price behavior in financial markets. In Chapter I we bring out the distinction between a physical process and an economic process in terms of the stability of their internal structure and the prospects for making accurate predictions about them. We argue that simply applying the theoretical estimation methodology appropriate for physical processes to the economic process of price behavior in a financial market can lead one to build models that are too complex and hold inappropriately high expectations about the potential accuracy of volatility forecasts from those models.

The second area where conflict between theory and practice arises is in the use of implied volatility from option market prices. The conflict comes from the disparity between the trading strategies arbitrage-based derivatives valuation models assume investors follow and what actual market participants do. In theory, the implied volatility is the market's well-informed prediction of future volatility. In practice, however, the arbitrage trading that is supposed to force option prices into conformance with the market's volatility expectations may be very hard to execute. It will also be less profitable and entail more risk than simple market making that maximizes order flow and earns profits from the bid-ask spread. The latter, however, does little to enforce theoretical pricing in the face of the forces of supply and demand in the market.

In both cases, I try to point out important implications for estimating volatility that tend to be overlooked by those following the more traditional lines of thought. I hope the reader will find some of these insights to be of value.

In the long course of this research, there have been many people who helped in many ways.

First, I would like to thank a long series of able and patient research assistants who are responsible for much of the empirical work discussed below. They include Linda Canina, N.K. Chidambaran, Amar Gande, Clifton Green, Jeffrey Heisler, Edith Hotchkiss, and Sundar Polavaram. Other benefactors have helped by providing data. Thanks to Ajay Dravid, Arthur Ferri, Mark Flannery, Richard Levich, and Bill May. Over the years, helpful comments and suggestions have come from so many sources and at so many different presentations of portions of this work, that it is no longer feasible to mention 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.

One person who gave me valuable comments on the final manuscript, however, needs to be mentioned. Special thanks to Rob Engle whose gentle but firm defense of GARCH methods led me to reconsider whether I had given them a fair enough examination in the first version, and to do the additional research that persuaded me I had not. Finally, I am grateful to the Bank and Financial Analysts Association for the original funding that got this project started.

Chapter I.
INTRODUCTION

Volatility has become a topic of enormous importance to almost anyone who is involved in the financial markets, even as a spectator. To many among the general public, the term is simply synonymous with risk: high volatility is thought of as a symptom of market disruption. To them, volatility means that securities are not being priced fairly and the capital market is not functioning as well as it should. But for those who deal with derivative securities, understanding volatility, forecasting it accurately, and managing the exposure of their investment portfolios to its effects are crucial.

Modern option pricing theory, beginning with Black and Scholes [1973], accords volatility a central role in determining the fair value for an option, or any derivative instrument with option features. While the returns volatility of the underlying asset is only one of five parameters in the basic Black-Scholes (BS) option pricing formula, its importance is magnified by the fact that it is the only one that is not directly observable. Stock price, strike price, time to option expiration, and the interest rate are all known or can be easily obtained from the market, but volatility must be forecasted. Although the realized volatility over recent periods can easily be computed from historical data, an option's theoretical value today depends on the volatility that will be experienced in the future, over the option=s entire remaining lifetime. Simply projecting observed past volatility into the future is a common way to make a forecast, but it is only one of several common methods, and need not be the most accurate. Moreover, there are numerous variations in exactly how historical price data are used in predicting volatility. Volatility forecasting is vital for derivatives trading, but it remains very much an art rather than a science, particularly among derivatives traders.

From the beginning, volatility prediction has posed significant problems for those interested in applying derivatives valuation models, but the difficulty has become greater in recent years as the maturities of available instruments have lengthened dramatically. In the 1970s, most options trading was in equity options with maturities of only a few months. While it was recognized that a security's return volatility could be expected to change over time, as long as this only occurs gradually, it should be possible to get a reasonably good short term forecast by simply assuming that volatility over the near future will remain about the same as what was realized in the recent past. That assumption becomes less tenable the longer the maturity of the option that is being priced.

Today there is active trading in derivatives of all kinds with maturities that may be 10 years or more. How should one go about calculating the appropriate volatility parameter to value a 10 year cap contract on the Deutschemark / dollar exchange rate? However one decides to make such a forecast, it is bound to be subject to considerable error. How much uncertainty is there around the best possible prediction for a time span like that? These are some of the issues we will focus on in this monograph.

In the next chapters we will discuss and evaluate the major procedures for forecasting volatility, always with an eye toward prediction rather than modeling and explaining volatility behavior. Moreover, we will be most concerned with forecast accuracy, not with theoretical or econometric elegance, since elegance often comes at the expense of robustness in out-of-sample forecasting.

The remainder of this introduction will consider the fundamental question of what volatility actually is and why people need to forecast it. One of the major difficulties in resolving the arguments about whether derivatives trading increases the volatility of the market is that the term is understood in different ways by different people. Restricting our attention to professional derivatives traders and securities firms who use mathematical option pricing models, and to the academics who build them, one might expect fairly close agreement about how to define volatility, at least as far as how it is used in the models. It turns out, however, that even among those whose object is to obtain a parameter estimate to put into a standard theoretical valuation model, there are wide disparities in what they do, and in what they ought to do, for their particular purposes. These disparities arise from differences in how volatility affects their trading strategies, and in how they understand the fundamental mechanism of security valuation in a financial market. Many of the issues we will discuss are not particularly well recognized even by the professionals involved, who for the most part think they are all doing basically the same thing.

I.1. What is Volatility?

Empirical and theoretical research on security prices since the 1950s has largely supported the "efficient markets" or "random walk" model. Actually, the term random walk has a precise mathematical meaning that is not a fully accurate description of how security prices should and do move over time. But the expression was used in some of the earliest research on the topic, and being more colorful than the more precise "martingale" or "supermartingale," it has stuck in popular usage.

In an efficient market, asset price movements can be described by an equation like (I.1).

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \mu_t + \varepsilon_t \; ; \qquad E[\varepsilon_t] = 0 \; , \; Var[\varepsilon_t] = \sigma_t^2 \qquad \textbf{(I.1)}$$

The return at time t, r_t, is the percentage change in the asset price S over the period from t-1 to t. This is equal to μ_t, a nonrandom mean return for period t, plus a zero mean random disturbance ε_t that is independent of all past and future ε's. It is the lack of serial correlation in the random ε's that is the defining characteristic of efficient market pricing: past price movements give no information about the sign of the random component of return in period t.

Along with lack of serial correlation, a mathematical random walk adds further restrictions to equation (I.1). If S follows a (geometric, or proportional) random walk, the expected value of the return is zero and the variance of the random component is constant over time. Thus, μ_t would have to be zero and the variance of the ε's would be the same for all dates. Neither of these describes actual security price behavior. What theory and empirical evidence seem to rule out is the possibility of using information from past returns to predict the random component of future returns. By contrast, we expect normal assets to pay nonzero expected returns at rates that may vary (nonstochastically) over time, and there is no contradiction between efficient pricing and returns variance that changes over time.

In deriving the option pricing formula, Black and Scholes needed to model stock price movements over very short intervals of time, so that they could consider a trading strategy of continuously rebalancing a portfolio consisting of an option and its underlying stock. The formulation they adopted is the logical extension of the random walk model to continuous time.

$$\frac{dS}{S} = \mu \, dt + \sigma \, dz \qquad \textbf{(I.2)}$$

It is the limiting process of (I.1) as the time interval goes to zero, holding constant the mean and variance of returns per year. The result is the lognormal diffusion model shown in equation (I.2). where dS is the asset price change over an infinitesimal time interval dt, μ is the mean return at an annual rate, dz is a time independent random disturbance with mean 0 and variance $1 \cdot dt$ (a stochastic process known as Brownian motion), and σ is the volatility, i.e., the standard deviation of the annual return.

By the nature of the limiting process, this model produces (continuously compounded) returns that follow a normal distribution and asset prices that have a lognormal distribution (that is, the logarithm of S has a normal distribution). This implies that the cumulative return over a finite holding period of length T has

$$\text{Expected Value} = \mu\ T$$

$$\text{Variance} = \sigma^2\ T \qquad \textbf{(I.3)}$$

$$\text{Standard Deviation} = \sigma\ \sqrt{T}$$

An important feature of this asset price process is that with a constant volatility σ, the standard deviation of total return over a holding period increases with the square root of the length of the period. This is a consequence of the time independence of the random dz component of return.

This model, and subsequent extensions of it, has become the standard way to model asset price behavior, both for derivatives pricing and in financial applications generally. Thus, options traders will set price quotes by putting a forecast of σ into the Black-Scholes model. A firm issuing a 5 year warrant may price it using the same volatility figure, obtained from some securities firm=s options trading desk, which may have been computed from the observed daily returns on the underlying stock over the past few months.

A bank trying to gauge the potential credit exposure on a long term loan may also attempt to estimate the probability distribution for the future value of the underlying assets that have been financed with the loan, by modeling the evolution of asset value over time with an equation like (I.2). The mean and standard deviation of the distribution at each future date will then be evaluated using (I.3). A securities exchange setting margin requirements on its members may model the potential mark to market variation in their account equity using (I.2) and treat the risk exposure as going up with the square root of time as in (I.3). And a firm may compute its own risk exposure and capital needs to carry a given asset position by looking at a "Value at Risk" calculation (essentially a way to estimate σ) and applying the square root rule for an assumed risk management period.

Even though the user's needs are quite different in each of these examples, in a world where asset returns obey (I.2) they all should use the same volatility parameter in making their calculations. When the instantaneous mean and volatility are constant over time and dz is a true Brownian motion, (I.3) gives the exact parameters of the normal distribution for returns over any holding period.

However, given that no model is ever a perfect description of the world, it is no surprise to learn that equations (I.2) and (I.3) are not either. Obviously, if σ for some security were really a constant parameter like, say, the melting point of lead is, there would be no need for forecasting; one could simply look it up once in a reference book and not have to bother about it again. The very fact that it is necessary to forecast volatility rather than getting it out of a book means that we expect it to be time-varying, and in a stochastic manner. That is, we do not believe (I.2) and (I.3) are strictly correct. In fact, the empirical evidence shows that the behavior of asset returns in the real world differs substantially from (I.2) and (I.3) in a number of ways. The following are some of the more important:

Time variation in the returns distribution: As we have discussed, volatility changes randomly over time. Optimal forecasting must take this into account.

Transactions prices are not equilibrium prices: The model is meant to describe the evolution of the "true" or equilibrium price, but price data normally come from recorded transactions. The bid/ask spread, less-than-continuous transactions frequency, different trading hours in different markets, and other aspects of market "microstructure" can introduce noise, and other statistical problems like nonsynchronous price data across related markets, into the computation.

Serial correlation: Price movements in actual securities markets are not perfectly uncorrelated over time, especially not at very short intervals. Positive (negative) correlation between consecutive price changes lowers (raises) measured volatility relative to the true value that should be used for σ. Negative serial correlation may come from bid-ask Abounce,@ as sequential trades alternate between the bid and the ask prices in the market, while positive serial correlation can be caused by the price effects of breaking large market orders into a sequence of smaller sized pieces for easier execution, and also from infrequent trading of some of the component stocks in an index.

Non-lognormality: Observed price changes deviate consistently from lognormality. There are more very large changes and (consequently) more very small ones than a lognormal distribution calls for. The commonly used term for this is "fat tails:" There is more weight in the tails of the actual returns distribution than in a lognormal distribution with the same variance. In some markets, the lognormal diffusion model fails because the price can "jump" occasionally from one level to another without trading at the prices in between, as in the case of a formal devaluation of an exchange rate or a discrete change in a managed interest rate like the prime rate.

Mean reversion in the volatility: Stochastic volatility models generally build in mean reversion for the volatility parameter, to reflect the observation that periods of extremely high or low volatility tend to be followed by a reversion toward a more moderate long term level.

Mean reversion in the price level: In many cases, particularly for interest rates, the value of the underlying is expected to move toward a long run mean level over time. This will not affect the theoretical option value in most models, because the drift term μ does not affect how the option is priced relative to the underlying asset, even if μ is not constant over time. But the probability distribution for the value of the underlying at the end of a finite holding period will depend on mean reversion, because the standard deviation will no longer rise with the square root of the time period. Instead, over a long horizon, the probability distribution for a stable mean-reverting process converges to a fixed steady state or Aergodic@ distribution, and extending the holding period will not increase the variance at all.

These differences from the process described in (I.2) and (I.3) give rise to an empirical question of what the optimal forecasting strategy is in practice. Depending on what the user is trying to do with the volatility forecast, different ones of these problems may be more or less important. Thus, the best volatility estimation method may differ for different uses.

First consider what an option trader needs to know. The Black-Scholes model and all similar derivatives valuation formulas are based on an arbitrage strategy that involves hedging the option against the underlying asset and continuously adjusting the position as the price changes and as time elapses. To make this a risk free trade in theory (it can not be made riskless in practice), it must be followed over the option's entire lifetime. Thus, the volatility input to a theoretical option pricing formula must be the volatility that is expected instant by instant from the present through the option expiration day. Under (I.2) and (I.3) with constant σ, that is simply $\sigma\sqrt{T}$, where T is the time to option maturity.

With time-varying but nonrandom volatility, the constant σ becomes σ_t, a parameter that changes over time. Even so, the BS and similar models can be adjusted to give the correct option value simply by setting the volatility parameter in the formula, σ^*_t, equal to the square root of the average variance over the option's life.

$$\sigma^*_t = \left(\frac{1}{T-t} \int_t^T \sigma^2_s \, ds \right)^{1/2} \qquad \textbf{(I.4)}$$

Allowing volatility to be stochastic as well as time-varying adds an important new degree of complexity. There are now two sources of risk, price risk and volatility risk, and both should affect the theoretical option value. Black-Scholes pricing no longer holds; only in the special case in which traders are indifferent to volatility risk can the formula be partially salvaged. In that case, the theoretical option value becomes the expected value of the BS price, with the expectation taken over the probability distribution for the average volatility over the option's life. (See Hull and White [1987].) This is both considerably more complicated than simply using the BS model with a single volatility input, and also incorrect except under the unwarranted assumption that traders do not require any compensation for bearing volatility risk (meaning also that they would see no reason to hedge against volatility risk exposure).

This discussion brings out one of the internal contradictions that are pervasive in applying theoretical derivatives pricing models in practice. Volatility is known to be time-varying and stochastic, so a variety of methods to forecast it and to manage volatility risk are in use. Nevertheless, options are generally priced simply by computing a point forecast for the unknown volatility and putting it into a constant volatility option pricing model like Black-Scholes.

In any case, it is clear that in using an option pricing model to compute the fair value and also to set the hedge ratio, the required volatility parameter is the expected volatility over the whole life of the option, whether that is one week or ten years. This is because the derivation of the model is based on the (theoretical) possibility of following an arbitrage strategy that will lock in any discrepancy between the option's market price and the model value as a risk free extra profit. For the arbitrage to be riskless, it must be followed continuously over the whole life of the option.

Computation of the volatility input is greatly affected by the time variation in σ_t and also by measurement problems due to market microstructure noise and serial correlation. The optimal forecasting strategy for theoretical option valuation should be robust against these problems. Mean reversion in prices can also cause an estimation problem, as Lo and Wang [1995] demonstrate, though it is probably fairly small in most cases. Non-lognormal price jumps can have a substantial influence on option valuation for out-of-the-money options close to expiration, but not much for longer maturity contracts.

Although derivatives theory is clear on the principle that the volatility parameter to put into the formula must apply to the instrument's whole lifetime, real world market makers are often quite surprised by that notion, especially for long-lived contracts. For many options markets, a market maker holds an inventory that turns over constantly, with expected holding periods often measured in hours or days, rather than months and years. A market maker who buys a 3 year warrant in the morning and expects to sell it in the afternoon is typically astonished at the academic's claim that to price it properly, he should be trying to predict the volatility of the underlying asset over a three year horizon.

Instead, the market maker=s two concerns are first, how big a price move might there be over the expected holding period (which would determine how hard it might be to hedge the instrument with the underlying asset), and second, how will the market be pricing the option when the time comes to sell it (which will depend on the market's volatility prediction at that time). Thus he is interested in the instantaneous volatility σ_t rather than average volatility over the option=s life, σ^*_t as defined in (I.4), and especially in the implied volatility that will be embedded in the option=s market price when he sells it. Implied volatility refers to the value of the volatility parameter that would set the theoretical option value equal to its current market price. Implied volatility is widely interpreted as "the market's" volatility forecast, and is used in pricing options both by practitioners and by academics. We will discuss the predictive ability of implied volatility in detail in Chapter III.

The point of this discussion is that while both an option theorist and a market maker may compute a value for σ from historical price data, their understanding of exactly what it is they are trying to forecast may be quite different. Their choice of methods may also be quite different. For example, because of their short expected holding periods, practitioners tend to use quite limited samples in computing historical volatility, like closing prices from the last 30 to 60 days, regardless of the maturity of the derivative instrument they are pricing.

Now let us briefly consider the type of problem for which a volatility forecast is needed but options are not involved. The bank trying to gauge long term credit risk exposure might use (I.3) to project the future probability distribution of the value of the assets supporting the repayment of its loan. Because there is no need to trade continuously, the bank's concern is not how day to day price variability will behave over a long period. Rather, the bank wants to know how far asset value might fall during the life of the loan and what the probability is that it will be less than the outstanding principal amount at maturity. Thus non-lognormality in the form of occasional price jumps is nearly irrelevant, but mean reversion in the price level is crucial. While option valuation does not depend on the behavior of the mean, the terminal probability distribution does very much. Specifically, the lognormal diffusion process of (I.2) with constant mean and volatility drifts arbitrarily far from its starting point over the long run, while a process with mean reversion approaches a long run probability distribution that does not change as the maturity is extended further. Thus the volatility forecasting method that is optimal for the bank's purpose may be quite different from what an option trader would use.

Finally, a securities exchange setting margin requirements needs to focus on the short run, and is especially concerned about large market moves. Here, the issues of serial correlation and especially of large non-lognormal price jumps become critical, while time variation and mean reversion are not likely to have much impact over the very short run. And the firm doing value-at-risk calculations may find itself in a tricky situation, since the daily price data that typically go into the calculation inherently produce very short term forecasts while it may be much more appropriate for the firm to focus on a longer horizon.[1] Simply applying the square root rule to

[1] For example, J.P.Morgan's RiskMetrics system provides volatility and correlation estimates for a large number of financial rates and prices based on a weighted average of the last 25 trading days, with the bulk of the weight going on the most recent observations. See Morgan Guaranty Trust [1995].

the one-day volatility obtained in a value-at-risk calculation may give highly erroneous results if there is serial correlation or fat tails in the daily data.

The final answer to the question of what volatility is, therefore, must be that in a theoretical world as described by (I.2) and (I.3) the meaning of volatility and the best technique to estimate it from past data are straightforward, but in the real world, where prices do not behave exactly the way they are modeled, different methods may be appropriate for different uses. Having highlighted the potential for different approaches to forecasting volatility, our focus in this monograph will be on derivatives pricing.

I.2. Forecasting out-of-sample: The moons of Jupiter vs the whale

The object of this monograph is to explore some of the major ways of forecasting volatility. The previous section discussed a variety of circumstances in which such a forecast is needed and pointed to reasons that it might be appropriate to use different procedures for different cases. Before looking at the methods in depth, however, it is worthwhile thinking briefly about the conceptual framework underlying the entire exercise. As we have seen in the discussion about the model embodied in equation (I.2), two analysts may use the same set of tools to work on very similar problems and yet come to quite different results without realizing that their fundamental approaches are not the same.

The essential element of the problem that concerns us is prediction: making use of data that is available in the present and using it to forecast "out-of-sample." Those schooled in classical statistics and estimation theory tend think about this general problem in a particular way, which greatly influences the kinds of models and procedures that they use and what they expect to be able to obtain from such models. I will argue that the classical statistics view of the world does not accurately represent the nature of the underlying structure of a financial market. It tends to lead statisticians to build models that are too complex, to expect too much out of their models, and to test them in inappropriate ways.

The following anecdote gives a good illustration of the difference between a classical statistics approach to forecasting and the conceptual approach that I believe is most fruitful in practice.

Some years ago, I was attending a conference on options, and heard a well-known academic present the theory of a new pricing model, along with a small amount of empirical evidence that appeared to be consistent with the model, though not definitive empirical confirmation of it. The model had been derived in the standard manner, by determining the option price which ruled out the possibility of risk free excess returns from an arbitrage strategy, a very clever one in this case.

The problem was that the strategy itself was so complex that one could be confident no actual investor in the market was following it. I asked him why we should expect prices observed in the market to behave according to the postulated pricing relationship when there were no actual market participants who understood the model or used it to do the trades that would push prices to the correct values. The answer was that he thought of an equilibrium pricing model for a financial market the way an astronomer thinks of the physical laws of motion that apply to a celestial body. Although the moons of Jupiter do not understand why they behave in a particular way, an outside observer who knows the laws of motion they follow can make very accurate predictions about where they will be thousands of years in the future. By the same token, he believed, if he could uncover the Alaws of motion@ of derivatives valuation in this market, it would not matter if no actual traders understood them.

This is essentially the way classical statistics models an estimation problem. There is assumed to be some fixed but unknown underlying structure, or "data generating process," and the statistician has a set of observations produced by that process from which estimates of its parameters will be deduced. One aspect of this conceptual framework is that estimation and forecasting are very similar to each other. Standard "goodness of fit" statistics that tell how closely a model fits the data that were used in estimating it, like a regression R^2 or a standard error of regression, give a good guide to the accuracy one might expect when the model is used to forecast out-of-sample. The only difference is a, typically slight, increase in the anticipated standard error due to the fact that the prediction is made with estimates of the model parameters and not their true values. A second feature of this framework is that one might hope to get arbitrarily good parameter estimates if one has a large enough data sample.

I believe the classical statistics framework fundamentally misrepresents the nature of a financial market and leads those who adopt it to expect much better forecasting performance than can be achieved in practice. Consider a different and more earthly estimation and prediction problem from that of the moons of Jupiter. Suppose we wanted to predict the movements of a whale, based on observing it over a period of time. Being a large animal with a lot of momentum, the movements of a whale must be fairly predictable over the short run simply by extrapolation. Yet we do not think of a whale as following a fixed and immutable pattern the way the moons of Jupiter do, at least not one that we could ever hope to understand completely.

As a complex living organism, a whale's behavior must remain at least partially unpredictable no matter how much past data we may have. In this case, we are not looking at a fixed structure with constant but unknown parameters, but rather at a system that evolves over time, and perhaps alters its behavior rapidly on occasion. Because its evolution is partly stochastic, no amount of past data will allow us to know the exact structure of the system now or in the future. Observation of past behavior allows us to develop an approximation to what the structure was (on average) during the period that produced the data sample. Prediction is possible only because the system usually evolves slowly and therefore our accumulated information from observing it only decays slowly.

In this case, there may be an enormous difference between how well a model fits in-sample and how well it can forecast out-of-sample, and classical goodness of fit statistics may give little guidance about the latter. Also, having a large data sample for estimation does not guarantee that accurate parameter values can be computed. Indeed, given that the data is generated by a structure that changes in unknown ways over time, expanding the estimation data set by adding observations from the distant past can easily make the estimates of the current state of the system worse rather than better.

Finally, given that the structure does not remain constant, there is a great premium on models and estimation procedures that are robust against small changes. The more detailed and elaborate a model is, the better the fit one is generally able to obtain in-sample, but the faster the model tends to go off track when it is taken out-of-sample. Thus, an oversimplified but robust forecasting approach that captures the major features of the system may give significantly more accurate prediction, particularly for longer horizons, than a more ambitious model which tries to capture its fine structure that may change relatively faster over time.

As should be obvious, I believe a financial market is much more like a whale than like the moons of Jupiter. In particular, forecasting is a very different operation from in-sample estimation. We shall see evidence of considerable whale-like behavior in the problem of forecasting volatility in the empirical results presented below.

I.3. Overview of the Study

Following this introduction, our major results are presented in two main chapters. Chapter II examines procedures for estimating volatility from historical price data, while Chapter III looks at the forecasting performance of implied volatility calculated from observed option prices. The final chapter offers our conclusions and suggestions for further research.

Chapter II begins with the standard logarithmic diffusion model for asset prices that underlies the Black-Scholes option pricing model, shown above as equation (I.2). The timeinvariant volatility parameter σ can be estimated from a sample of historical price data following the standard procedure from classical statistics. Under the assumption that (I.2) is strictly true, the sample variance (annualized) is a consistent estimate of σ^2 and the sampling error can be made arbitrarily small by using a large number of data points, even if they are only observed over very short intervals. This property apparently makes the intraday transactions data that are available for many financial markets very useful, since even a few days= observations can produce a huge number of sample points.

Unfortunately, recorded prices from actual securities markets violate the assumptions embodied in (I.2) in a number of ways, and the deviations are more apparent the shorter the sampling interval. The result is that the pure classical statistics estimation procedure does not automatically produce the best volatility forecasts, and a variety of alternative techniques are commonly used to deal with the data problems that crop up in real market data. For example, biases induced by the presence of Anoise@ from the effects of market bid-ask spreads, noncontinuous trading, and serial correlation over very short intervals make most intraday data, and in some cases data from daily and even longer observation intervals, unusable for the calculation. Another obvious problem is that since volatility changes over time, old data eventually become obsolete. This suggests that accuracy may be improved by eliminating data points from the sample when they get too old. It is also reasonable to hypothesize that the optimal length of the historical sample may be a function of the forecast horizon, so that longer sample periods are appropriate when a long term forecast is needed.

Finally, the statistical properties of the estimators for the variance and the mean of a time series are quite different. In particular, volatility is calculated from the deviations of the sample data from the mean, but the sample mean can be a very inaccurate estimate of the true mean when the sample spans a short time period. Classical statistics assumes that the analyst knows nothing at all about the true parameters of the price process except what is obtained from the data, so calculating volatility by taking deviations from the sample mean, however inaccurate it may be, is still the best that can be done.

But, knowing that the data represent returns on securities traded in a financial market gives a financial economist additional insight into what reasonable parameter values may be. For example, even though the sample mean return on the Standard and Poor=s 500 stock index over a few months (a typical sample period for a market maker=s volatility calculation) could easily turn out to be -75% per year, it would make no sense economically to treat that value as the best possible estimate of the true mean return. Indeed, such an estimation result is a product of volatility, not a reasonable estimate of the mean. One common alternative to taking deviations from a noisy sample mean is to fix the mean at zero. Although zero is clearly not likely to be the true mean either, we may feel that in many cases it will be a lot closer to the true mean on average than the sample mean is.

Chapter II explores some of these alternative estimation procedures empirically to see which ones work the best in practice. We concentrate particularly on the problem of forecasting volatility over longer horizons, from six months to five years ahead. As maturities of traded derivative instruments have lengthened from a few months to 5 years, 10 years, and even longer, the need for long term volatility estimates for use in pricing and hedging such contracts has grown rapidly. It certainly seems possible that the best way to forecast volatility over the whole lifetime of a ten year Deutschemark (DM) cap contract might differ from the methods used in pricing a two week DM call option. In addition, the need to use valuation models in trading such long term contracts means that the error in the volatility forecast will be an important determinant of derivatives risk. We analyze historical data series from several important financial markets to determine, for different maturities, which procedure gives the most accurate forecasts and how big the typical forecast error is.

The results we obtain in these investigations are rather surprising. Briefly, we find that, while different markets are different, volatility forecasts are generally more accurate for longer horizons than for shorter ones, and the most accurate method for both long and short horizons is to use a historical data sample that is much longer than what is normally chosen. Moreover, restricting the mean to be zero rather than taking deviations around the sample mean seems to increase accuracy for most of the combinations of market and forecast horizon that we examine.

Estimating historical volatility and projecting it forward is a very common approach to volatility forecasting in practice, but there is an obvious conceptual inconsistency in trying to predict volatility because it changes over time, using a procedure which is based on the assumption that it doesn't. Formal models of time-varying volatility of the ARCH family (autoregressive conditional heteroskedasticity) have been widely applied to a number of problems in economics. The second part of Chapter II examines the forecasting performance of a simple but general model of this type, the GARCH (1,1).

We discover that the GARCH model fitted with monthly data is not well-suited to this application. First, fitting the model parameters statistically requires a much larger number of data points than simply calculating historical volatility. This necessitates either very long historical samples or high frequency data. In attempting to make long horizon predictions, we wish to avoid a model that focuses too narrowly on day to day market behavior, so we attempt to construct forecasts from monthly data. But this proves to be impossible for the GARCH (1,1), which only rarely is able to converge without difficulty on acceptable estimates for the three model parameters even with five years of monthly data. A second problem with ARCH family models is that they are not designed for forecasting many steps ahead and their performance, when parameter estimates can be obtained, tends to degrade rapidly as the forecast horizon is extended.

A problem with the historical volatility approach using monthly data is that for relatively short horizons of a year or less, the small number of sample points will entail relatively large sampling error. To explore the effect of our choice of a monthly observation interval, at the end of Chapter II, we also examine using daily data to forecast both daily and monthly volatility for medium term horizons (out to 24 months). The results are largely consistent with what we observe using monthly data, but with a couple of differences. One important one is that the GARCH methodology is found to perform much better with daily data, when the forecasting horizon is quite short.

Chapter III examines the other basic approach to forecasting volatility: implied volatility. While volatility is the one parameter in the standard option pricing model that can not be observed directly, for traded contracts the option=s market price is available. This allows the analyst to solve backward through the model to obtain the volatility value that investors must be using to arrive at the option prices currently seen in the market. Both academic researchers and options market participants use these implied volatilities extensively. Academics have strong prior beliefs that financial markets are informationally efficient with respect to widely available information, so that if the implied volatility is the market=s expectation of future volatility, it should be an unbiased and well-informed estimate that incorporates all of the information that can be obtained from observed past price behavior, as well as all other public information.

Chapter III begins with an elaboration of this argument, but then observes that market participants actually have quite a different perspective on what implied volatility means, due to the fact that the trading strategy upon which nearly all theoretical valuation models are based is highly impractical in real world markets. Traders tend to regard implied volatility as a measure of how the market is currently pricing a given option relative to its underlying asset, without worrying too much about whether it is an accurate forecast of the actual volatility that will be realized over the option=s future lifetime.

The major issue we examine in Chapter III is what information implied volatility does contain about the true future volatility of the option=s underlying asset. We are especially interested in assessing whether it is reasonable to assume implied volatility is an efficient forecast that accurately reflects all widely available information. This would actually require two separate conditions to hold. First, the implied volatility must be the market=s actual volatility forecast, and second, the market must make efficient use of all available information in forming that forecast. We begin by considering the impact of several important data problems, such as bid-ask spreads and nonsynchronous prices, that will cause calculated implied volatilities to differ from market expectations. We then look at statistical tests of forecast rationality as applied to implied volatilities from a number of option markets.

If implied volatility is an informationally efficient forecast of future volatility, the realized volatility should equal the implied volatility plus a (small) zero mean random error. This relationship, which must hold for any rationally formed forecast, is easily tested for a given data sample by a simple regression of the realized values on the forecasts. We refer to this standard procedure as the Arationality test@ regression. A second test for the relative information content of two or more competing forecasts, e.g., implied volatility and historical volatility, is also easily performed using an Aencompassing regression.@ We discuss the details of executing such tests, and then present some striking results from a study of the Standard and Poor=s 100 Index options market by Canina and Figlewski [1993], which found that in a large sample of data drawn from one of the most active options markets in the world, implied volatility contained no information at all about future volatility.

This extreme result calls into serious question the academics= strongly held beliefs that implied volatility provides a better forecast of future volatility than does historical volatility. In attempting to explain these apparently anomalous results, we offer a new hypothesis, that the information content of implied volatility will be a function of how easy it is for an options trader to engage in the arbitrage strategy that would allow him to profit from a mispricing in the market. To shed some light on this hypothesis, we gather results from running the rationality test regression on data from a variety of options markets, as reported in the finance literature. Briefly, the hypothesis that implied volatility will contain more information about future volatility in markets with easier options arbitrage is generally supported by these studies. However, in virtually every case, the statistical evidence indicates that implied volatility is not a fully rational forecast, even though it may appear to contain the most information among several alternatives.

The final issue discussed in Chapter III is how these results should affect the way implied volatility is used in forming predictions of future volatility. In particular, we argue strongly against the common practice of taking implied volatility as the best available volatility forecast because it does appear to contain more information about future volatility than do competing forecasts based only on historical prices. It must also pass the test of rationality. Accurate option valuation requires an accurate volatility input. But we show that both in theory and in practice, greater information content need not translate into more accurate forecasts in a root mean squared error sense unless the bias is corrected first. Unfortunately, to correct the bias it must first be estimated from historical data, which leads to yet another difficult forecasting problem. We are left with a series of results that illuminate some of the major problems in obtaining good volatility predictions from implied volatilities, but no final resolution of these problems.

The final chapter sums up our major conclusions and suggests directions for future research.

Chapter II.
FORECASTING VOLATILITY USING HISTORICAL DATA

II.1 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 process. 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 have had maturities that were typically measured in weeks or months rather than years.[2] 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 of 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.[3] Today, maturities of 5 to 10 years are not uncommon.

[2] By contrast, maturities of embedded options have always been potentially quite long. One common example is a call provision in a long term corporate bond, that applies until the bond=s maturity date.

[3] 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 are available at different exchanges.

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 chapter is to explore these issues empirically for several important financial instruments. We will examine and contrast different procedures that base volatility estimates on historical data, 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 II.3 we discuss several tricky issues in implementing it in practice. Section II.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 increase 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 II.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 do more sophisticated, but more fragile, approaches. In fact, when we examine the performance of the GARCH(1,1) model fitted to monthly data in Section II.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 it could be used.

Our focus in this chapter is on predicting long horizon volatility. To avoid any effects from market microstructure noise and other self-correcting short run phenomena, we do the estimations with monthly data in the first part of the chapter. This does not seem entirely appropriate for either historical sample periods or forecasting horizons of 1 year or less, since there are so few monthly data points to work with. The GARCH model, in particular, can be greatly improved by using daily data, that permits a large increase in the number of data points. In Section II.7, we examine long horizon forecasting performance of historical volatility and GARCH models that are fitted to daily data. We consider using daily data to predict both daily and monthly volatilities over horizons of up to 2 years.

The final section summarizes our results.

II.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 (I.2), which for convenience is repeated here as (II.1).

$$\frac{dS}{S} = \mu\, dt + \sigma\, dz \qquad \text{(II.1)}$$

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

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

and R has a Normal distribution, with

$$\text{Mean} \quad = \quad (\mu \ - \ \sigma^2 / 2 \ ) \ T$$

$$\text{Standard deviation} \quad = \quad \sigma \sqrt{T}$$

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. Thus, what 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 <u>average</u> annualized return variance over the option's lifetime, as shown in equation (I.4) above.

When an asset's price follows the constant volatility lognormal diffusion model of equation (II.1), σ can be estimated easily from historical data. The difficulty arises because actual prices do not follow (II.1) exactly, so that price behavior may change over time and differ over intervals of different lengths. Moreover, the <u>ways</u> in which (II.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 (II.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 model that adjusts for random volatility formally.

<u>The Standard Historical Volatility Estimate</u>

Consider a set of historical prices for some underlying asset that follows the process defined in equation (II.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 $R_t = \ln (S_t / S_{t-1})$, for t from 1 to T.

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

$$\overline{R} = \frac{\Sigma\ R_t}{T} \qquad \text{(II.2)}$$

The variance of the R_t is given by

$$v^2 = \frac{\Sigma\ (R_t - \overline{R})^2}{(T - 1)} \qquad \text{(II.3)}$$

The denominator in (II.3) is (T - 1) because the information contained in one observation is effectively used up in calculating the sample mean. On the other hand, if the mean is known (or is constrained by the analyst to be some particular value, such as zero), this information is not lost and the sum of squared deviations should be divided by T.

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

$$\sigma = \sqrt{N\ v^2} \qquad \text{(II.4)}$$

If the constant parameter diffusion model of (II.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 becomes the forecast for volatility going forward, over a time horizon of any length.

II.3. Problems with the Lognormal Diffusion Model

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

Time-Varying Volatility

One major problem is that volatility clearly changes over time. As an illustration consider Figure II.1, which plots the volatility of the U.S. Treasury 20 year bond yield. Taking monthly data on the bond yield from January 1971 through July 1993, we used equations (II.2) through (II.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 based on the historical data available at each point in time) certainly do not appear to be only a constant parameter plus random 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 (II.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 (II.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 have a large effect on the measured volatility. Second, the fact that prices do not obey equation (II.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 long term volatility forecasts.[4]

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

---

[4] This comment is not meant to deny the value of the large amount of research that is currently being done to understand high-frequency financial data. However, the models and methods of that line of research are sufficiently complex, and the focus on market behavior at the shortest time intervals is so different from what we are looking at here, that it is doubtful much improvement in long term volatility forecasts can be obtained by using intraday price data.

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.[5]

Sampling at longer intervals is an easy way to limit the effect of serial dependence at 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 is 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 mostly use monthly observations. However, in Section II.7, we also analyze daily data for comparison. The empirical results exhibit clearly the tradeoff between increasing accuracy by using daily data with a larger number of observations and losing accuracy because of the relatively greater effect of transitory phenomena on daily prices. In principle, it should be possible to correct for the effects of a known degree of serial dependence in high frequency data. However, doing so would require estimating autocorrelation coefficients from historical data and assuming they were constant over time, and we wish to avoid depending on such assumptions.

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 in stock prices over periods of several years. While this will have little effect on volatility forecasting for most exchange traded equity options, whose maturities are well under a year, 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. Long run 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

[5] 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.

that an option that 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 (II.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 attempted to deal with the empirical returns distribution by fitting constant elasticity of variance models or other specifications that allow for this.[6] 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 obvious reason that the degree of tail fatness should 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 sense. Dealing with the effect of an outlier like October 19, 1987 in estimating stock volatilities is a prime example of the difficulty. Figure II.2 shows a rolling volatility for the S&P 500 Index, calculated at each point from daily closing prices over the previous 500 trading days, i.e., about 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

[6] See Macbeth and Merville [1980], for example.

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 in Figure II.2: exactly 501 trading days after October 19, 1987, that data point drops out of the calculation and the Ahistorical@ 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 (II.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 (II.2):

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

All of the prices observed in between S_0 and S_T drop out of the calculation. Moreover, under equation (II.1), the standard deviation of ($\ln S_T - \ln S_0$) is $\sigma\sqrt{T}$ 0, so the standard error of $\overline{R}$ 0 as an estimate of μ is $\sigma/\sqrt{T}$ 0. This only depends on the length of the sample period, T, 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

$20 / \sqrt{4}$ 0 = 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 daily 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 ($.25 \; / \; \sqrt{1/12}$ 0 = 0.866). 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 <u>impose</u> a value for the mean 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 as the assumed mean. Fortunately, the estimate of the volatility does not depend very heavily on the mean. (The bias is proportional to $(\hat{\mu} - \mu)^2$ 0, which is a very small number if $\hat{\mu}$ 0-- either the sample mean or a mean imposed by the analyst -- is not too far from the true mean.) Thus, while it is very difficult to obtain an accurate mean estimate from the data, the main thing 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 first adopt the approach of imposing a mean return of zero and then later examine the quantitative effect of the constraint 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 (II.2) - (II.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.

II.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 (II.2) - (II.4), (making ad hoc adjustments when the procedure appears to be giving inappropriate answers). But the idea that it may be better to adjust the length of the historical sample for different forecasting horizons suggests that it is worthwhile examining 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.[7] The volatility estimate from that procedure is given in equation (II.6)

$$\sigma_t = \left(\frac{\sum_{s=1}^{k} (R_{t+1-s} - \overline{R})^2}{k-1} \right)^{\frac{1}{2}} \qquad \text{(II.6)}$$

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. Here 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 1971, after the era of floating rates. Table II.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 (RMSE). 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

---

[7] Since beginning to study the performance of this approach to volatility forecasting, I have come to think of it as a reasonable benchmark for assessing the incremental value of the more elaborate ARCH family of autoregressive conditional heteroskedasticity models. To forecast time-varying volatility, the approach computes the unconditional variance, but optimizes over the length of the historical sample period. I therefore suggest it be called Optimized Unconditional Conditional Heteroskedasticity, with the acronym OUCH.

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

and

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

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

Both historical and realized volatilities are computed around an assumed value of zero for the mean returns. We present results later to show how much difference computing volatility from the deviations from the sample means would make to forecast accuracy. We have made no adjustment for October 1987, or any other unusual events. However, the distorting effect of the Crash is more limited here than in Figure II.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 the assumed mean of zero, i.e., by averaging the squared returns and taking the square root. 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 we are examining, and the forecast errors are recorded. The starting time period is then 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. Finally, the whole process is repeated for each value of k.

This procedure unfortunately results in autocorrelation in the forecast errors that is potentially quite large 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).[8] 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 II.2 to II.5 contain the results for the four series we are examining here. Each table shows the forecast RMSE for different combinations of forecast horizon and number of months in the historical data sample. The last line in each table gives the average realized volatilities which, because of the calculation method, will differ slightly across the different horizons.

In Figures II.3 to II.6, some of the RMSEs are displayed visually, expressed as percentages of the realized values, for ease of comparison across strategies and markets. To illustrate them, consider Figure II.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 ovals 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. The percent root mean squared error is 50.9% (forecast RMSE is .0692 relative to an average realized volatility of .1360 at this horizon).

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 historical sample size considered, with percent RMSE of 41.2%. 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

[8] On the other hand, the standard error on the MSE computed under the incorrect assumption of independence would be biased, so we do not attempt to calculate standard deviations for the root mean squared forecast errors.

observations. If we really need a 6 month volatility, we should probably use daily or weekly data in the calculations. This issue is explored further in Section II.7. Forecasting at the one year horizon is more accurate, again reaching the minimum percent 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 rather 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 any case, all of the methods show poor performance for this market, with percent RMSEs of 70% or more. One reason to expect greater difficulty in forecasting the volatility of short term interest rates is that they do not fluctuate freely: they are both stabilized and managed by the Federal Reserve in its conduct of monetary policy.

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 .0574 at 6 months, .0417 at two years, and as low as .0310 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. However, mean reversion in volatility does not explain why 6 and 12 month horizon forecasts (that span time intervals too short for much mean reversion to occur) are more accurate when the historical data sample is extended to include data from several years earlier.

Another clear result for all of these series (except T-bills), as well as for 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 3.1 percent, seems quite remarkable.

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 will typically 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 II.2 to II.5, with volatility computed around the sample mean.

Table II.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 forecast over the next 12 months, the RMSE is 7.3 percent lower when the mean is not estimated, i.e., RMSE(zero mean) / RMSE(sample mean) - 1 = -0.073. Only in a few cases did constraining the mean lead to an increase in forecast RMSE. As one would anticipate, the difference between the two methods is larger when shorter time periods are involved, but these results indicate that for the more volatile markets even over quite long sample periods, more accurate forecasts may be obtained by computing volatility around zero rather than around the sample mean.

II.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.

$$R_t = E[R_t] + \varepsilon_t \qquad \begin{array}{l} E[\varepsilon_t] = 0 \\ Var[\varepsilon_t] = \sigma_t^2 \end{array} \tag{II.7}$$

Although variants exist in which the mean in equation (II.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

$$\sigma_t^2 \equiv C \tag{II.8}$$

for which the standard variance fitting procedure in equations (II.2) - (II.4) applied to all 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 modeled 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 (II.9) shows an ARCH(3) model.

$$\sigma_t^2 = C + a_1 \varepsilon^2_{t-1} + a_2 \varepsilon^2_{t-2} + a_3 \varepsilon^2_{t-3} \tag{II.9}$$

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 (II.10).

$$\sigma^2_t = C + a_1 \sigma^2_{t-1} + b_1 \varepsilon^2_{t-1} \qquad \text{(II.10)}$$

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

$$\sigma^2_{\text{Long Run}} = \frac{C}{1 - a_1 - b_1} \qquad \text{(II.11)}$$

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

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.

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.[9] Stock return volatility increases following a sharp price drop, but a price rise of the same size may even lead to lower volatility.

---

[9] Black [1976a] is one of the first articles to make this point and comment upon its potential importance for option pricing.

To deal with these problems, Nelson [1991] proposed Exponential GARCH (EGARCH), which models the log of variance, so that the explanatory variables can take negative values 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.[10]

Problems with ARCH-family Models

These models have been widely examined and applied in economics and finance.[11] The bulk 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 and deriving parameter estimates by maximum likelihood estimation. The normal distribution is by far the most common for log price relatives, but the Student t distribution is also used, in order to capture the fat tails effect.

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 best 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 stochastic 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 typically require quite a large number of observations before they behave well. Likelihood surfaces may be quite flat, making finding a maximum difficult, or the

---

[10] 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]. Day and Lewis [1993] also compare volatility forecasting performance of several models, including GARCH and EGARCH, for crude oil futures prices. This study will be discussed further in Chapter III.

[11] See, Bollerslev, et al [1992] for a review of their applications in finance.

maximum for a given sample may lie outside the theoretically acceptable region (with coefficients that are negative or sum to values greater than 1.0, which implies 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 value multiplied by a negative coefficient may drive the entire fitted variance negative for a given period. A GARCH formulation has the advantage that, while disturbances over all recent periods can enter into the calculation, one fits only a small number of parameters, which increases the likelihood that they will all be well-behaved,.

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.

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

$$E_t[\sigma^2_{t+1}] = C + a_1 \sigma^2_t + b_1 E_t[\varepsilon^2_t]$$

$$= C + (a_1 + b_1)\sigma^2_t \qquad \text{(II.12)}$$

$$\vdots$$

$$E_t[\sigma^2_{t+k}] = C \sum_{s=0}^{k-1} (a_1 + b_1)^s + (a_1 + b_1)^k \sigma^2_t$$

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 (the beginning 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 $(a_1 + b_1)$.

II.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 1947 through December 1951. Those parameters would be used to construct out-of-sample GARCH forecasts for the first 6 months of 1952 (using the procedure shown in equation (II.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.[12] 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 1952 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 1952 into the sample and dropping the same number of observations from the beginning, to keep a window of fixed size.

This procedure turned out to be infeasible, because it was extremely difficult to fit the model on as few as 60 data points even though we were only trying to fit three parameters. Of the first 36 five-year periods, the estimation routines in GAUSS (the software package we were using) were unable to converge on acceptable 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

[12] 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.

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 five 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, for which the data sample only begins in 1971.

Table II.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 were different for the GARCH predictions of Treasury bond yield volatility. Here, the GARCH post-sample forecasts were distinctly less accurate than historical volatility, and they got 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. 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.

In fairness to the GARCH methodology, these results plainly do not constitute a definitive test of its performance in forecasting volatility. Difficulty in achieving convergence to coefficient estimates using a specific software package is hardly an insuperable problem. Different estimation routines may be more successful, and convergence can always be obtained with Ahammer and tongs@ grid search methods (though not necessarily to acceptable, i.e., non-explosive coefficient values). Rather, our conclusion from these results is that attempting to allow for predictable time-variation in asset volatilities with a GARCH specification using monthly data 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. In the next section, we reconsider GARCH along with the optimized historical volatility estimator in estimations using daily data.

II.7. Volatility Forecasting with Daily versus Monthly Sampling

Up to this point, we have concentrated on forecasting volatility over relatively long horizons and have conducted the investigation using monthly price data. This was to avoid potentially spurious results that might be associated with short term self-correcting phenomena (such as bid-ask bounce, as well as other, less-easily-specified effects). But before leaving this topic it is worthwhile to explore somewhat the effect of this choice of periodicity on our results. For one thing, it makes sense to use more frequent observations when sample or forecast periods are a year or less, because of the sampling error in the monthly estimates that contain only 6 or 12 data points for those horizons. Introducing daily data also allows us to investigate the performance of the GARCH methodology further.

We obtained long samples of daily data for the four series examined above, with the exception that the best available daily long term interest rate series was for 10 year rather than 20 year Treasury bond yields. The data sources and the time periods covered are shown in Table II.1. The daily series span fewer years than the monthly series, but obviously, they contain a large number of observations. As before, in order to allow direct comparison of forecasting accuracy for estimates that use historical samples of up to 5 years, or forecast over longer horizons (up to 2 years, in this case), we set the first and last data points for which forecasts would be computed (t_{1st} and t_{last}) as follows, relative to the series beginning and ending data points (t_{beg} and t_{end}):

$$t_{1st} = t_{beg} + 1259$$

and

$$t_{last} = t_{end} - 504$$

In order to be able to compare results with daily data to those that would be obtained with monthly data over the same periods, we have constructed Amonthly@ series from the daily series by assuming a Amonth@ is exactly 21 (trading) days. Thus, the estimation starts at the data point 5 years (that is, 60 21-day months) from the beginning of the sample, at day 1260 for each series. It ends 24 x 21 = 504 days before the last observation. The exact dates covered in each case are shown in the relevant tables.

The first experiment is to forecast daily volatility over various horizons using different amounts of daily historical data. In keeping with the spirit of our investigation of long term volatility using less frequent sampling where possible, we consider forecasting daily volatility only out to 24 months in the future. However, we are able to look at shorter sample periods with daily than with monthly data, so we introduce 1 and 3 month estimates. Table II.8 contains the daily historical volatility results equivalent to Tables II.2 through II.5.

In addition, Table II.8 shows the RMSE forecasting performance of GARCH(1,1) models fitted to the daily series. In the GARCH estimation, we use a rolling sample of 5 years of historical data. GARCH coefficients are refitted every 6 months, and multistep ahead out-of-sample average volatility forecasts for 6, 12 and 24 month horizons are computed as described above. With the much larger sample size, there was much less difficulty in obtaining convergence to reasonable parameter estimates than with the monthly data sample. Lastly, the average realized volatility for each series and forecasting horizon are shown.

With daily observations we see some of the features we had expected in the performance of volatility estimates for different sample periods and horizons, but did not find with monthly data. Here it does appear that there is an optimal amount of past data to use, that varies with the forecast horizon. For example, just looking at the S&P 500 index results, the most accurate forecasts of volatility over the next month come from limiting the sample to the most recent one month of data. For a 3 month horizon, the lowest RMSE comes with 12 months of historical data, while forecasting 12 or 24 months in the future, the greatest accuracy comes from using 5 years of data. This general pattern is visible in the other series as well.

These results can not be directly compared to those with monthly data because of the different time periods covered by the samples. However, the overall picture is similar for the two data sets. For the most part, forecast accuracy is higher for longer horizons than for shorter horizons, and at a given horizon the most accurate forecasts are produced from historical samples that are quite a bit longer than the forecast horizon. For example, with all four series the most accurate 3 month volatility forecasts come from 12 or 24 months of historical data.

One of the largest differences between the daily and monthly data results is in the performance of the GARCH model. For the S&P 500 volatility, the daily GARCH model forecasts are much better than any of the historical volatilities at every horizon. Note, of course, that the GARCH estimation is based upon the previous 5 years of data. However, while that means that the coefficients are derived from a long history of prices, the volatility forecasts produced by the GARCH model actually depend most heavily on recent prices.

Unfortunately, the excellent GARCH model performance for the S&P index is not duplicated for the other series. In forecasts over the next month, the GARCH model is most accurate, or almost so, for all four series. But at a three month horizon, the GARCH has the highest RMSE, or close to it, and performance declines sharply for longer horizons for all series but the S&P index. This suggests that the basic GARCH(1,1) model fitted to daily data may be quite useful in forecasting volatilities for equity markets, but that it may be of considerably less value for making out-of-sample projections in other markets beyond short horizons.

One reason to focus on volatility computed from monthly data is that frequent rebalancing of a hedge on an option position is costly, so that those that will be held over long maturities are unlikely to be rebalanced every day. If a 5-year derivative hedge is to be rebalanced once a month, say, volatility computed from monthly data ought to be more indicative of the hedging cost (and therefore the option=s replacement value, estimated as the cost of replicating its payoffs with a dynamic trading strategy) than volatility of the daily price movements over the next 5 years would be.

Therefore, the final experiment we report on in this chapter involves using forecasts computed from daily historical data to predict monthly volatility in the future. Using the same daily data as in Table II.8, we constructed series of realized Amonthly@ volatilities for 6, 12, and 24 month horizons, where again, a Amonth@ is defined to be 21 trading days. These realized volatilities were then forecasted using both daily and monthly historical samples of different lengths. In addition, we used the GARCH models, fitted on 5 years of daily data, to produce average volatilities for monthly data at the same horizons.

Table II.9 displays the results. They are pretty consistent with what we have seen previously. In each column, we consider forecasts over a specified horizon that are computed from differing amounts of either daily or monthly historical data. In the results presented above, we found that the most accurate forecasts tended to come from using the maximum amount of historical data possible, except when a very short horizon was used with daily data. That pattern also holds here, to a degree, but since the shortest horizon in this table is 6 months, the 60 month historical samples give the lowest RMSE forecasts except for the two interest rate series at the shortest horizons. With regard to calculating historical volatilities from daily versus monthly data, the table shows that for the longest horizon, 24 months, computing the volatility forecast from 5 years of monthly historical data gives the most accurate forecast, while for the 6 month horizon, forecasts constructed from (some amount of) daily historical data had the lowest RMSE. At the 12 month horizon, results were mixed, with monthly beating daily for 2 series, daily beating monthly for one, and one tie. Again, the GARCH model performed very well for the S&P 500 index volatility, but not for the other series, with RMSEs increasing sharply for longer horizons.

II.8. 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 chapter, we have examined 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 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, except for very long time periods in relatively low volatility markets.

* The GARCH(1,1) model requires quite a large data sample for easy estimation of its

coefficients, which makes monthly data hard to use. It is also not designed for multistep ahead forecasting. GARCH performed reasonably well when daily data were used in fitting the model and forecasting horizons were under 3 months. The exception was the S&P 500 index, for which a daily GARCH model performed well for all horizons out to two years.

* The brief investigation of data periodicity suggests that there is a difference between the behavior of volatility estimated over the same time period with daily and with monthly observations. In using daily historical data to predict daily volatility, there may well be a payoff to trying to optimize the length of the historical sample, though for longer horizons like 24 months, the longest available set of past data seemed to give the best results. In forecasting the volatility of monthly prices, accuracy was improved for short horizons of 6 months by employing daily historical data, but for the longer horizons, predicting monthly volatility with monthly data seemed to be the best. These results on data periodicity should be considered suggestive only; further investigation of the issue would be worthwhile.

TABLE II.1
Dates and Sources of Data Series

Monthly Data

Series	Dates	Source
Standard and Poor's 500 Stock Index	1/47 - 12/92 1/93 - 12/95	CRSP Bloomberg
3 Month U.S. Treasury Bill Yield	1/47 - 3/95	Federal Reserve
20 Year U.S. Treasury Bond Yield	1/50 - 12/92 1/93 - 7/93	Salomon Bros. Bloomberg
Deutschemark Exchange Rate (DM per $)	1/71 - 11/95	Harris Bank

Daily Data

Series	Dates	Source
Standard and Poor's 500 Stock Index	July 2, 1962 - Dec. 29, 1995	CRSP
3 Month U.S. Treasury Bill Yield	Jan. 2, 1962 - Dec. 29, 1995	Federal Reserve
10 Year U.S. Treasury Bond Yield	Jan. 2, 1962 - Dec. 29, 1995	Federal Reserve
Deutschemark Exchange Rate (DM per $)	Jan. 4, 1971 - Nov. 30, 1995	Federal Reserve

Table II.2

Forecast Accuracy of Historical Volatility Estimates for Various Sample Periods and Forecast Horizons

Standard and Poor=s 500 Stock Index
Jan 1952 - Dec 1990

Root Mean Squared Forecast Error for
Volatility Calculated around Mean of Zero

	Forecast Horizon (Months)					
Months in Sample	6	12	24	36	48	60
6	0.0692	0.0626	0.0635	0.0634	0.0616	0.0593
12	0.0629	0.0574	0.0579	0.0562	0.0529	0.0509
24	0.0640	0.0582	0.0549	0.0503	0.0461	0.0448
36	0.0631	0.0561	0.0501	0.0447	0.0413	0.0400
48	0.0603	0.0522	0.0450	0.0399	0.0363	0.0346
60	0.0574	0.0489	0.0417	0.0362	0.0320	0.0310
Average Realized Volatility	0.1360	0.1395	0.1425	0.1437	0.1441	0.1439

Table II.3

Forecast Accuracy of Historical Volatility Estimates for Various Sample Periods and Forecast Horizons

3 Month Treasury Bill Yield
Jan 1952 - Mar 1990

Root Mean Squared Forecast Error for
Volatility Calculated around Mean of Zero

Months in Sample	Forecast Horizon (Months) 6	12	24	36	48	60
6	0.1767	0.1862	0.1898	0.1906	0.1850	0.1830
12	0.1839	0.1843	0.1836	0.1828	0.1755	0.1752
24	0.1838	0.1806	0.1792	0.1738	0.1686	0.1696
36	0.1921	0.1861	0.1786	0.1732	0.1693	0.1704
48	0.1957	0.1866	0.1797	0.1753	0.1718	0.1724
60	0.2005	0.1918	0.1835	0.1784	0.1746	0.1738
Average Realized Volatility	0.2345	0.2440	0.2533	0.2558	0.2566	0.2569

Table II.4

Forecast Accuracy of Historical Volatility Estimates for Various Sample Periods and Forecast Horizons

20 Year Treasury Bond Yield
Jan 1955 - Jul 1988

Root Mean Squared Forecast Error for
Volatility Calculated around Mean of Zero

Months in Sample	Forecast Horizon (Months) 6	12	24	36	48	60
6	0.0595	0.0585	0.0577	0.0584	0.0596	0.0608
12	0.0573	0.0540	0.0529	0.0533	0.0543	0.0556
24	0.0555	0.0513	0.0493	0.0498	0.0509	0.0507
36	0.0550	0.0505	0.0482	0.0485	0.0482	0.0471
48	0.0559	0.0512	0.0486	0.0475	0.0461	0.0451
60	0.0572	0.0523	0.0481	0.0460	0.0448	0.0432
Average Realized Volatility	0.1119	0.1152	0.1182	0.1196	0.1204	0.1211

Table II.5

Forecast Accuracy of Historical Volatility Estimates for Various Sample Periods and Forecast Horizons

Deutschemark Exchange Rate
Jan 1976 - Nov 1990

Root Mean Squared Forecast Error for
Volatility Calculated around Mean of Zero

Months in Sample	Forecast Horizon (Months): 6	12	24	36	48	60
6	0.0564	0.0496	0.0473	0.0485	0.0480	0.0467
12	0.0499	0.0410	0.0396	0.0393	0.0390	0.0366
24	0.0474	0.0396	0.0358	0.0347	0.0326	0.0292
36	0.0479	0.0396	0.0347	0.0306	0.0271	0.0231
48	0.0489	0.0404	0.0316	0.0257	0.0212	0.0176
60	0.0461	0.0353	0.0253	0.0189	0.0149	0.0126
Average Realized Volatility	0.1131	0.1178	0.1224	0.1245	0.1244	0.1238

Table II.6

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

Past Obs	Forecast Horizon	S&P 500	3 Month T-Bills	20 Year T-bonds	Deutsche-mark
6	6	-13.7	-0.4	-6.0	-16.3
12	12	-7.3	-4.2	-3.9	-8.1
36	12	-5.1	-4.0	-1.8	-3.9
36	36	-3.6	-3.5	1.3	0.7
60	12	-5.2	-2.4	-1.2	-4.2
60	36	-3.3	-2.3	0.8	1.2
60	60	-1.9	-1.5	0.0	2.3

Note: For each selected pair of historical sample period and forecast horizon, the table shows the percentage difference between the root mean squared forecast error of volatility estimates computed from deviations from the sample mean versus estimates computed around an imposed mean of zero. A negative entry indicates that the zero mean estimates achieved the lower RMSE.

Table II.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 1959 - Dec 1993
Observations:	420
Successful estimations:	28
Failed estimations:	7

Horizon	GARCH RMSE	5 Year Historical RMSE	10 Year Historical RMSE
6 months	.0645	.0648	.0626
12 months	.0547	.0557	.0538
24 months	.0482	.0497	.0477

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).

Table II.8

Out-of-Sample Accuracy of Using Daily Data to Predict Daily Volatility

The table shows root mean squared forecast error for daily volatility calculated around a mean of zero, for different forecast horizons and historical sample lengths. GARCH models are estimated using daily data from the previous 5 years. (One Amonth@ is defined to be 21 trading days.)

S&P 500 Stock Index, June 19, 1967 - Jan. 6, 1996

Months in Sample	Forecast Horizon (Months) 1	3	6	12	24
1	0.0691	0.0669	0.0703	0.0729	0.0728
3	0.0697	0.0674	0.0668	0.0686	0.0675
12	0.0709	0.0666	0.0652	0.0666	0.0641
24	0.0719	0.0672	0.0648	0.0655	0.0607
60	0.0735	0.0688	0.0656	0.0633	0.0571
GARCH	0.0579	0.0538	0.0519	0.0513	0.0509
Average Realized	0.1297	0.1329	0.1352	0.1375	0.1400

3 Month Treasury Bill Yield, Jan. 6, 1967 - Dec. 30, 1993

Months in Sample	1	3	6	12	24
1	0.0615	0.0566	0.0588	0.0642	0.0677
3	0.0572	0.0500	0.0508	0.0564	0.0589
12	0.0576	0.0492	0.0494	0.0541	0.0548
24	0.0610	0.0525	0.0512	0.0543	0.0511
60	0.0677	0.0596	0.0564	0.0549	0.0462
GARCH	0.0602	0.0631	0.0902	0.2186	1.3838
Average Realized	0.1483	0.1523	0.1547	0.1570	0.1597

Table II.8 continued

10 Year Treasury Bond Yield, Jan. 6, 1967 - Dec. 30, 1993

Months in Sample	1	3	6	12	24
1	0.0451	0.0429	0.0444	0.0483	0.0527
3	0.0434	0.0395	0.0386	0.0429	0.0470
12	0.0443	0.0392	0.0372	0.0415	0.0445
24	0.0462	0.0401	0.0380	0.0423	0.0425
60	0.0510	0.0455	0.0435	0.0452	0.0410
GARCH	0.0433	0.0435	0.0528	0.0980	0.6045
Average Realized	0.1105	0.1144	0.1167	0.1189	0.1212

Deutschemark Exchange Rate, Jan. 8, 1976 - Dec. 7, 1993

Months in Sample	1	3	6	12	24
1	0.0412	0.0398	0.0382	0.0401	0.0403
3	0.0400	0.0362	0.0317	0.0338	0.0332
12	0.0394	0.0334	0.0284	0.0301	0.0285
24	0.0387	0.0320	0.0274	0.0283	0.0246
60	0.0413	0.0344	0.0280	0.0252	0.0192
GARCH	0.0391	0.0368	0.0392	0.0532	0.1032
Average Realized	0.0993	0.1021	0.1038	0.1055	0.1082

Table II.9
Out-of-Sample Accuracy of Using Daily versus Monthly Data to Predict Monthly Volatility

The table compares root mean squared forecast error for monthly volatility using daily versus monthly historical data. Volatility is calculated around a mean of zero, for forecast horizons of 6, 12, and 24 months and various historical sample lengths. GARCH models are estimated using daily data from the previous 5 years. (One Amonth is defined to be 21 trading days.)

S&P 500 Stock Index, June 21, 1967 - Jan. 13, 1994

Months in Sample	Using Daily Historical Data 6	12	24	Using Monthly Historical Data 6	12	24
1	0.0811	0.0759	0.0737			
3	0.0759	0.0702	0.0681			
6	0.0723	0.0663	0.0646	0.0789	0.0721	0.0687
12	0.0702	0.0640	0.0623	0.0720	0.0655	0.0608
24	0.0701	0.0636	0.0602	0.0693	0.0613	0.0553
60	0.0699	0.0621	0.0574	0.0676	0.0580	0.0509
GARCH	0.0600	0.0529	0.0497			
Average Realized	0.1379	0.1417	0.1443	0.1379	0.1417	0.1443

3 Month Treasury Bill Yield, Oct. 26, 1967 - Jan. 13, 1994

Months in Sample	Using Daily Historical Data 6	12	24	Using Monthly Historical Data 6	12	24
1	0.0916	0.0894	0.0911			
3	0.0869	0.0843	0.0845			
6	0.0861	0.0830	0.0825	0.0979	0.0969	0.0959
12	0.0888	0.0829	0.0822	0.0958	0.0903	0.0884
24	0.0905	0.0846	0.0813	0.0935	0.0861	0.0820
60	0.0910	0.0830	0.0765	0.0882	0.0776	0.0683
GARCH	0.1120	0.2355	1.5097			
Average Realized	0.1839	0.1898	0.1947	0.1839	0.1898	0.1947

Table II.9 continued

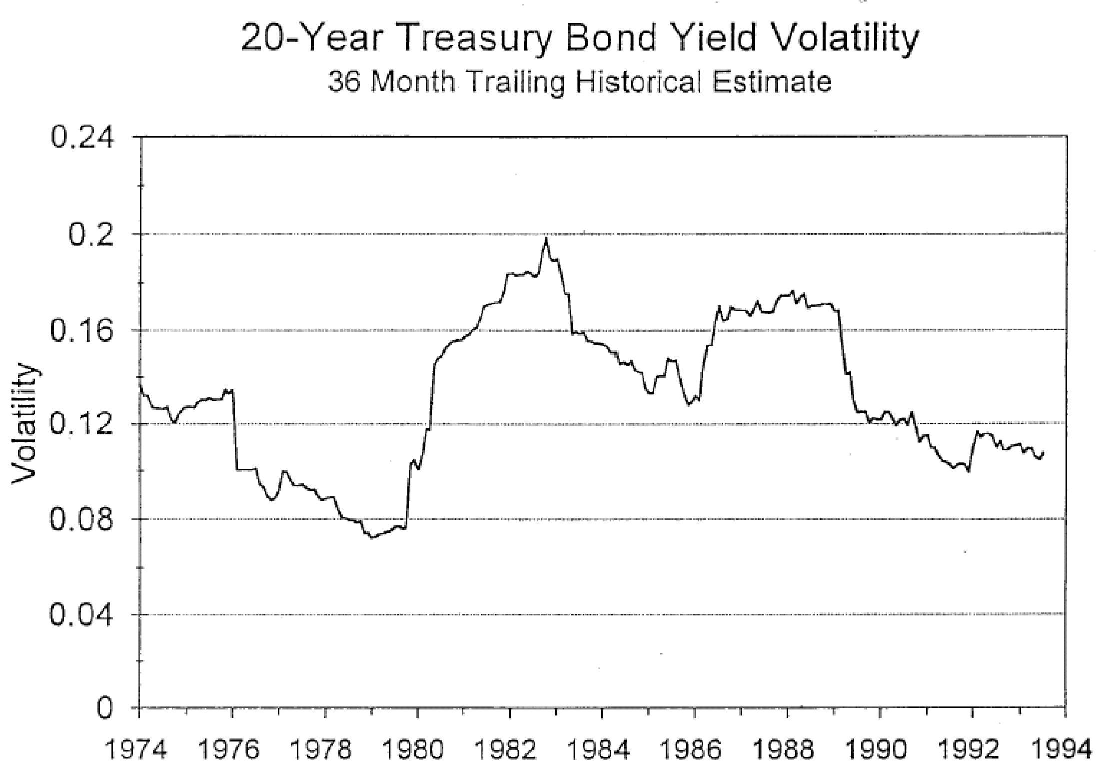
10 Year Treasury Bond Yield, Oct. 26, 1967 - Jan. 13, 1994

Months in Sample	Using Daily Historical Data 6	12	24	Using Monthly Historical Data 6	12	24
1	0.0637	0.0588	0.0589			
3	0.0595	0.0536	0.0533			
6	0.0574	0.0517	0.0510	0.0657	0.0607	0.0605
12	0.0566	0.0501	0.0499	0.0603	0.0539	0.0535
24	0.0574	0.0514	0.0502	0.0591	0.0524	0.0497
60	0.0584	0.0517	0.0483	0.0579	0.0500	0.0450
GARCH	0.0619	0.0951	0.6155			
Average Realized	0.1313	0.1355	0.1384	0.1313	0.1355	0.1384

Deutschemark Exchange Rate, Feb. 2, 1976 - Dec. 15, 1993

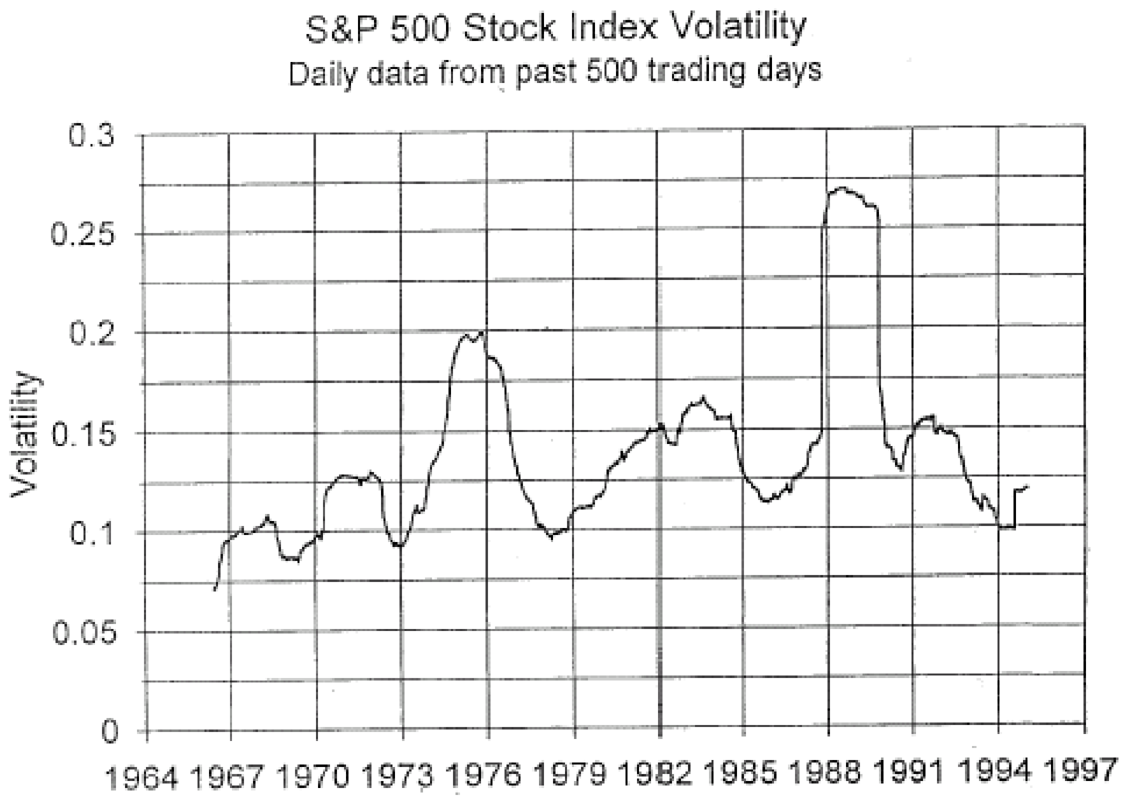
Months in Sample	Using Daily Historical Data 6	12	24	Using Monthly Historical Data 6	12	24
1	0.0494	0.0441	0.0442			
3	0.0438	0.0373	0.0371			
6	0.0391	0.0333	0.0332	0.0494	0.0446	0.0405
12	0.0378	0.0313	0.0313	0.0450	0.0368	0.0334
24	0.0383	0.0322	0.0291	0.0403	0.0329	0.0267
60	0.0366	0.0285	0.0223	0.0400	0.0304	0.0213
GARCH	0.0456	0.0509	0.1013			
Average Realized	0.1097	0.1131	0.1158	0.1097	0.1131	0.1158

FIGURE II.1



20-Year Treasury Bond Yield Volatility
36 Month Trailing Historical Estimate
Volatility
0.24
0.2
0.16
0.12
0.08
0.04
0
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994

FIGURE II.2



S&P 500 Stock Index Volatility
Daily data from past 500 trading days
Volatility
0.3
0.25
0.2
0.15
0.1
0.05
0
1964 1967 1970 1973 1976 1979 1982 1985 1988 1991 1994 1997

FIGURE II.3

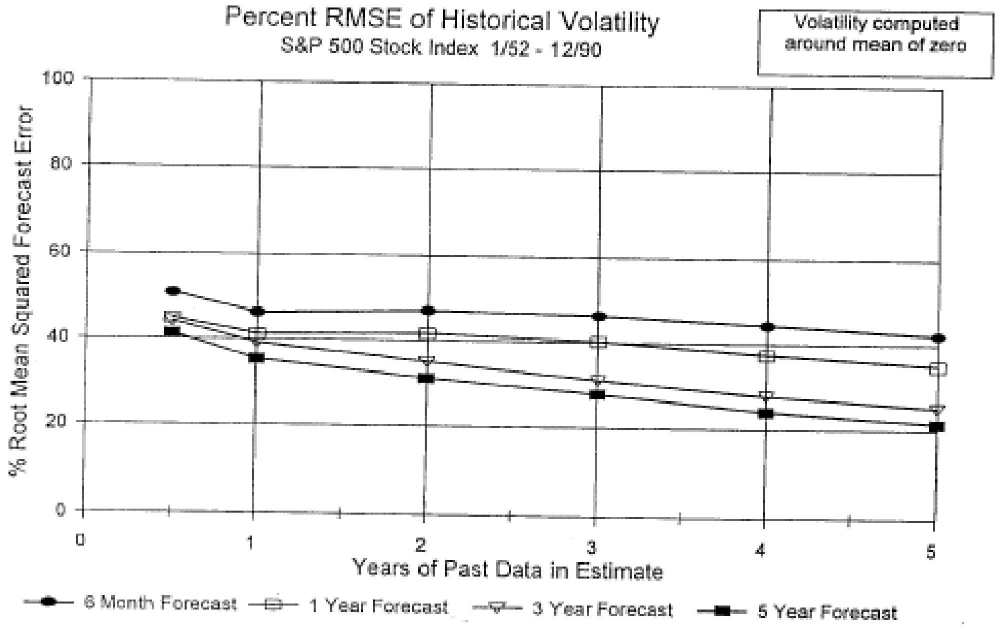
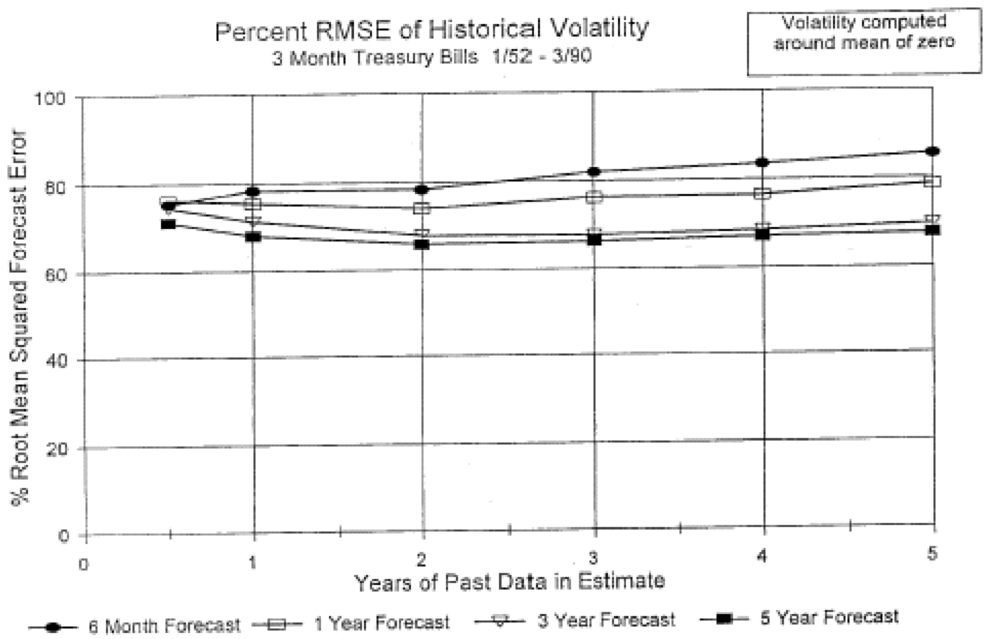

Percent RMSE of Historical Volatility
S&P 500 Stock Index 1/52 - 12/90
Volatility computed around mean of zero
% Root Mean Squared Forecast Error
100
80
60
40
20
0
0
1
2
3
4
5
Years of Past Data in Estimate
6 Month Forecast
1 Year Forecast
3 Year Forecast
5 Year Forecast

FIGURE II.4



Percent RMSE of Historical Volatility
3 Month Treasury Bills 1/52 - 3/90
Volatility computed around mean of zero
% Root Mean Squared Forecast Error
100
80
60
40
20
0
0
1
2
3
4
5
Years of Past Data in Estimate
6 Month Forecast
1 Year Forecast
3 Year Forecast
5 Year Forecast

FIGURE II.5

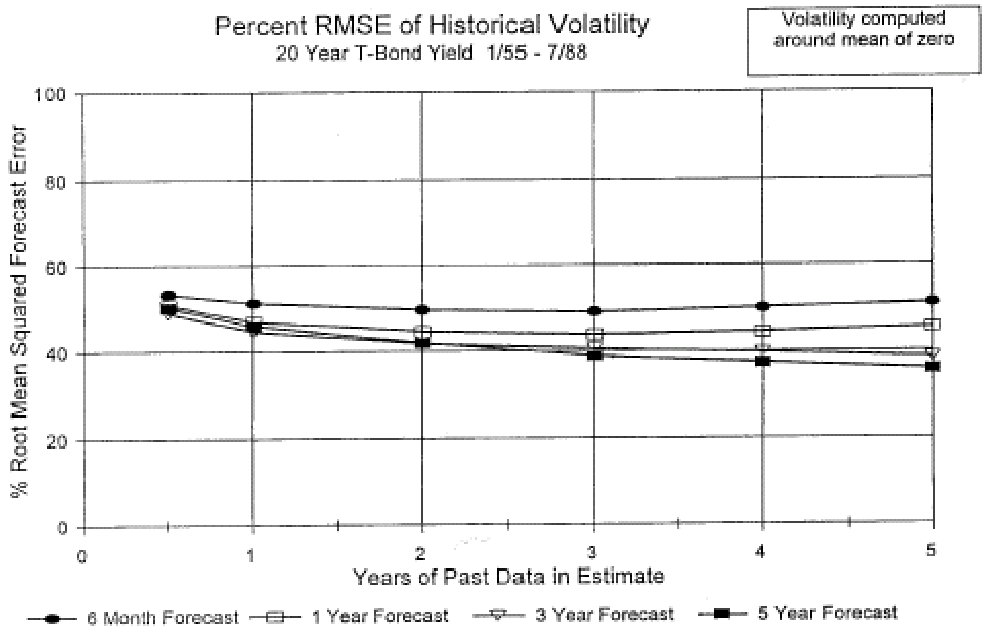
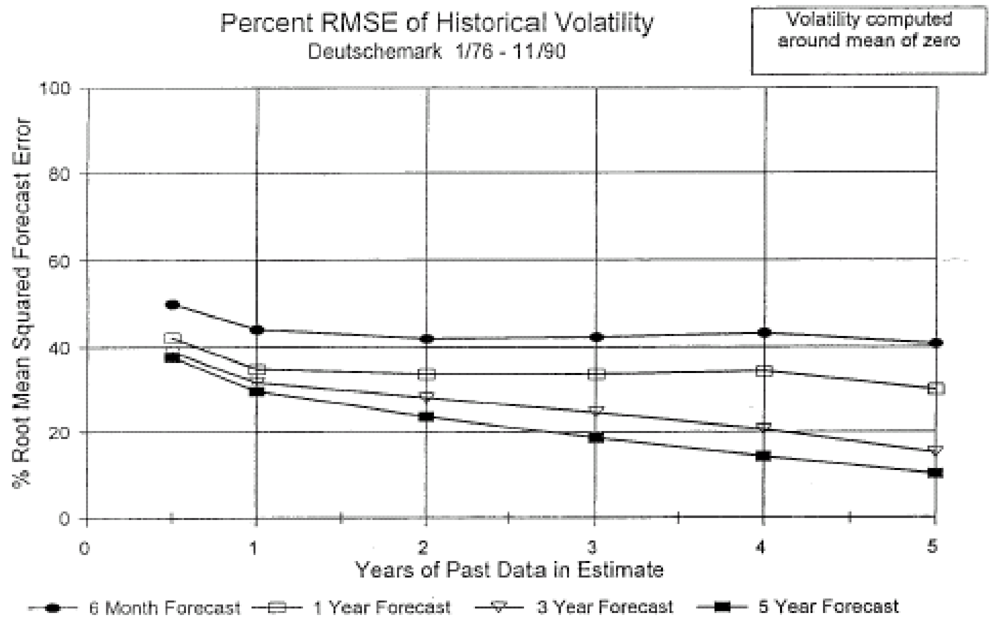

Percent RMSE of Historical Volatility
20 Year T-Bond Yield 1/55 - 7/88
Volatility computed around mean of zero
% Root Mean Squared Forecast Error
100
80
60
40
20
0
0
1
2
3
4
5
Years of Past Data in Estimate
6 Month Forecast
1 Year Forecast
3 Year Forecast
5 Year Forecast

FIGURE II.6


Percent RMSE of Historical Volatility
Deutschemark 1/76 - 11/90
Volatility computed around mean of zero
% Root Mean Squared Forecast Error
100
80
60
40
20
0
0
1
2
3
4
5
Years of Past Data in Estimate
6 Month Forecast
1 Year Forecast
3 Year Forecast
5 Year Forecast

Chapter III.
IMPLIED VOLATILITY

An alternative method of obtaining a volatility forecast for cases in which traded option contracts exist is to use Aimplied@ volatility. An option pricing equation like the Black-Scholes model gives the fair value for an option as a function of the price of the underlying asset, the option=s strike price and time to expiration, the riskless interest rate, and the volatility parameter. Of these variables, only volatility is not directly observable but must be estimated. However, while an investor can not observe volatility in the same way that the stock price can be seen, the market does reveal option prices. Thus, if the market is pricing options according to the valuation model, it is possible to solve the model backwards from the observed price to determine what the market=s volatility input must be. This is the implied volatility (IV).

Denoting the market price and model value for a given option as C_{MARKET} and C_{MODEL} , respectively, and implied volatility as σ_{IV} , we can write

$$C_{MODEL}\,(\sigma_{IV}) = C_{MARKET} \qquad \text{(III.1)}$$

The option=s implied volatility is computed by putting the observable variables into the C_{MODEL} pricing equation and (effectively) inverting it to solve for σ_{IV}. Although the inverse function can generally not be written out explicitly, the Black-Scholes model and other related valuation equations are monotonic in volatility, so it is very easy to find the implied volatility by numerical search methods.[13]

[13] Perhaps the easiest search method is the Method of Bisection. One begins with the option=s market price, C_{MKT}, and two trial volatilities that will bracket the true value, $\sigma_{LOW} < \sigma_{IV} < \sigma_{HIGH}$. If $C(\sigma_{LOW}) > C_{MKT}$ even when $\sigma_{LOW} \approx 0$, the option price violates its lower boundary condition (i.e., $C_{MKT} < S - Xe^{-rT}$). This means implied

III.1 Implied Volatility as the Market=s Volatility Forecast

Both option traders and academic researchers consider implied volatility to be of great importance. They typically feel it is a more valuable estimate of the true volatility than can be obtained from historical returns data alone, although for different reasons.

<u>The Academic Perspective on Implied Volatility</u>

Academic financial economists have a strong belief that financial markets are informationally efficient, in the sense that market prices accurately reflect all widely available information that is relevant for valuing securities. For options, this includes all information that can be used to predict the future returns volatility of the underlying asset. When the historical price record is available, the Amarket=s@ volatility estimate should correctly impound all information that can be gleaned from past returns. Where appropriate, this will include not only what can be obtained from a straightforward calculation of sample volatility, but also any more complex behavior that could be uncovered by GARCH or some other statistical technique. In addition, the market will have access to other historical information, from returns in other financial markets, past news events, and so forth, as well as knowledge about current market conditions and anticipated future events (e.g., current Federal Reserve policy, national and international financial

---

volatility can not be computed. Otherwise, one will have $C(\sigma_{LOW}) < C_{MKT} < C(\sigma_{HIGH})$. Now bisect the range: define the midpoint $\sigma_{MID} = (\sigma_{LOW} + \sigma_{HIGH})/2$ and compute $C(\sigma_{MID})$. If $C_{MKT} < C(\sigma_{MID})$, then σ_{IV} is less than σ_{MID} so we replace σ_{HIGH} by σ_{MID} and bisect the new narrower range. Otherwise, if $C(\sigma_{MID})$ is below C_{MKT}, σ_{IV} lies in the upper half of the range, so we keep the same σ_{HIGH} but replace σ_{LOW} by σ_{MID}, and continue the bisection. Each iteration cuts the possible range for σ_{IV} in half, and the procedure continues until the desired accuracy is obtained.

A second common approach to computing IV is <u>Newton-Raphson Search</u>. This can achieve convergence in a very few steps, but it requires calculation of both the option=s price and its Avega,@ or partial derivative with respect to volatility, at each iteration. Let σ_{TRIAL} denote some initial trial value for IV and $v = \partial C/\partial\sigma$ be the vega evaluated at σ_{TRIAL}. We want to know how much to change σ_{TRIAL} to get σ_{IV}. A linear approximation, Δ, can be calculated as $\Delta = (C_{MKT} - C(\sigma_{TRIAL})) / v$. Set the new value of σ_{TRIAL} equal to the old $\sigma_{TRIAL} + \Delta$, and repeat the process, until convergence to the desired accuracy is attained (typically only a couple of iterations). Newton-Raphson search converges much faster than bisection, but requires enough extra computation at each stage that it may not be faster in practice, especially if numerical derivatives are needed to obtain v.

conditions, upcoming news releases, elections, etc.). In other words, the volatility parameter implied by an option=s current market price in an efficient market should accurately impound all relevant information that can be obtained from historical returns data, plus much more. In that case, once implied volatility is known, any volatility estimate based on past prices alone should be redundant.

This chain of reasoning can be expressed mathematically as follows (supressing the time subscript t for convenience).

$$\sigma_{IV} = E_{MKT}[\sigma] \qquad \text{(III.2)}$$

and

$$E_{MKT}[\sigma] = E[\sigma \mid \Phi_{MKT}] \qquad \text{(III.3)}$$

with

$$\{S_{t-1}, S_{t-2}, \ldots\} \subseteq \Phi_{PUBLIC} \subseteq \Phi_{MKT} \qquad \text{(III.4)}$$

In words, (III.2) says that implied volatility is the market=s expectation of future volatility, and (III.3) says that the market=s expectation is the true conditional expected value of volatility given the market=s information set, Φ_{MKT}. Moreover, by (III.4) the set of historical prices is a subset of all publicly available information, $\Phi_{PUBLIC,}$ which is, in turn, a subset of the market=s information set (since that may include some nonpublic information).

Equations (III.2) and (III.3) actually express two fairly strong assumptions about the way options are priced in the market. The first says that the implied volatility is an accurate representation of the market=s expectation about future volatility. Among other things, this requires that the investigator must compute implied volatility from exactly the same model the market is using in pricing options (which still leaves aside the thorny issue of what it means to treat Athe market@ as having a single volatility prediction). Early papers that looked at the information content of implied volatility, such as Latané and Rendleman [1976] and Chiras and Manaster [1978] were flawed in this regard because they used the Black-Scholes European option model to compute implied volatilities for American options.

While this shortcoming is easily remedied by using an American option valuation model, other deviations from Black-Scholes pricing are harder to deal with. Given what we know about security price processes, it would not be surprising if options market prices embodied some effects relating to stochastic volatility, to the possibility of discrete price jumps, to mean reversion in both price levels and volatility, to fat-tailed distributions, and to other non-Black-Scholes-type behavior. Any such impact on prices in the options market will be impounded in implied volatilities if they are computed using the Black-Scholes model, and produce anomalous IV patterns across strike prices and maturities.

Similarly, since transactions costs and other market frictions allow some (theoretical) arbitrage opportunities to persist, mispricing due to fluctuations in option supply and demand from any source whatsoever will end up being impounded in implied volatility. For example, despite our strong belief that people are risk averse, they are observed to participate willingly, even eagerly, in risky lotteries that have negative expected values. This is presumably because they have a taste for the payoff pattern that consists of an extremely high possible payout combined with a limited, though highly probable, loss. Such a payoff pattern also characterizes deep out-of-the-money options. If individuals= preferences for lottery-like payoff patterns extend to the options market, out-of-the-money options will be priced higher in the market than the model says they should be (a phenomenon that is commonly observed in practice). This price effect would arise not from investors= beliefs that the expected payoff on such an option was great enough to produce a Afair@ return, but rather from the fact that investors were willing to accept a lower than fair expected return in order to obtain the desired payoff pattern. Yet because the Black-Scholes pricing model does not allow for such taste effects, the higher market prices would be entirely impounded in the volatility Aforecasts@ when implied volatility was computed for out-of-the-money options. These arguments illustrate some of the things that equation (III.2) assumes away.

Equation (III.3) says that the market makes informationally efficient volatility forecasts from the available information. Fortunately, the contention that a competitive financial market produces informationally efficient prices is on fairly solid ground. A large amount of empirical work over the years has supported the hypothesis of market efficiency, although not without some caveats, as well as a small number of Aanomalies@ that seem to contradict the general principle.

For implied volatility to have the optimality properties assumed of it as a forecast, both equations (III.2) and (III.3) must hold. But we are more comfortable with the second relationship than the first. And, reversing the argument, if in some markets or under certain conditions we find that implied volatility is not an efficient forecast, it need not imply that investors are irrational, i.e., a failure of equation (III.3). If (III.3) holds but (III.2) does not, it may simply mean that the option price in the market is different from the model value based only on investors= volatility expectations, because it is incorporating factors that are not properly accounted for in the pricing model from which implied volatility is computed.

The Option Trader=s Perspective on Implied Volatility

Academicians value implied volatility because they think it is the market=s volatility forecast, and the market is efficient so its forecast accurately impounds all widely available information. Option traders and other market participants also focus heavily on implied volatility, but for quite different reasons. For them, the great value of a pricing model is that it allows them to estimate what the option price will do when the underlying asset price moves to a different level.

The significance of implied volatility to a trader is not that it predicts price variability over the whole lifetime of the option, but rather that it indicates how the market is currently pricing options relative to the underlying asset. Option traders do not know which direction market prices will go, but because implied volatility tends to remain fairly constant over short time periods while the underlying asset and option prices fluctuate, from the model they will have a good idea of how the option price will respond to a given a change in the asset price. Thus, knowledge of current implied volatility allows traders both to hedge individual options against the underlying asset, and to price different options consistently relative to one another. Since both of these major uses of the model in actual trading involve relative pricing, in many circumstances it will not matter very much if the volatility input is not the true future asset price volatility, as long as the same input is used consistently.

This discussion brings out a significant disparity between how builders of theoretical option pricing models assume investors behave and what option traders do in reality. The logic behind the option pricing model is that if a trader knows the true volatility but the market option price differs from the theoretical value (which means the implied volatility is not the true volatility), then a riskless arbitrage trade is available. Theoretically, the trader stands ready to trade the option against the underlying asset, taking an infinitely large position if necessary, until prices move into the proper alignment. Of course, no actual trader would do this, if only because no one would be that certain their volatility prediction was right. Moreover, the transactions costs to maintain a hedged position over a long time period are potentially very large, and the fact that continuous rebalancing is impossible when the market is closed at night and on weekends means there will always be risk that can not be hedged away.

In contrast to this theoretical trader, an actual options market maker attempts to run a largely balanced position, taking long and short positions in different option contracts that mostly offset one another in terms of market risk exposure. This requires quoting bids and offers that are close to current market prices, so that buying and selling transactions approximately even out during the trading day. For example, if the market is currently pricing options such that implied volatility is about 15%, the market maker may quote bids based on a volatility of 14.8% and offers on 15.2%. Even if she were certain that the true volatility over the option=s lifetime would turn out to be 20%, the market maker would very likely continue to make markets based on 15% volatility, because that would produce more frequent turnover from which she would earn the bid-ask spread. If, on the other hand, she were determined to take positions based on the true 20% volatility, she would immediately buy as many contracts as she could afford to carry, hedge the position, and then have no further role as a market maker. There would be no revenue earned from the bid-ask spread until the contracts expired or investors eventually changed their views about volatility.

This discussion highlights the fact that in the eyes of active options market participants, the implied volatility need have little to do with the best possible prediction for the price variability of the underlying asset from the present through option expiration, while it has everything to do with the current and near term supply and demand conditions expected in the options market. Thus, the logical foundation for the financial economist=s belief that implied volatility is the efficient market=s forecast of the true volatility may be rather weak in reality.

III.2 Computing Implied Volatilities in Practice

One very common feature of implied volatilities in practice that causes difficulty for the efficient markets point of view is that they often differ substantially for different options on the same underlying asset. While, by definition, the underlying has only one volatility, an implied volatility can be obtained from every option traded on that asset. Of course, if volatility is expected to vary over time, two options with different maturity dates may both be priced correctly according to the model and still have different implied volatilities (as long as they are not too different).[14] However, since each option with the same maturity should be priced based on the

[14] If volatility is expected to fall in the future, a longer maturity option may be priced on a lower implied volatility than a nearby option, but there is a limit to how much lower it can be. If the discrepancy is too large, the only solution that would permit both options to embody the same volatility for the underlying during the first option=s

average volatility anticipated over that time period, there should be no systematic differences in implied volatilities across different strike prices. And yet, such differences are extremely common in practice.

It is to be expected that estimates of the same parameter drawn from different sources will not be identically equal, due to sampling noise. One major reason for noise in implied volatilities across different options is that the input data may not be synchronous. In particular, options that are away-from-the-money or have long maturities tend not to trade as frequently as at-the-money nearby contracts. If closing prices are used, for example, it is perfectly possible for the option=s price to have been recorded much earlier in the day than the price of the underlying, which introduces an error when the two prices are combined in the model calculations. A solution to this problem is to use intraday time-stamped transactions data, which has become increasingly common in recent years. Greater availability of intraday options data permits a much more precise examination of implementable arbitrage possibilities in the market. This has also led to an enormous expansion of the number of data points available for empirical studies, so that tests involving data sets with more than 100,000 observations are not unusual.

Another identifiable noise factor entering implied volatility calculations is the effect of bid-ask spreads on both the options and the underlying asset. For example, if the recorded option price comes from a transaction executed at the market=s ask price, it will appear to be relatively expensive and its implied volatility will be high, while if the option price was the bid, implied volatility will be lower. Similarly, if the contemporaneous underlying asset price used in the calculation of implied volatility comes from a trade at the market=s ask, a call will appear to be deeper in-the-money than it Areally@ is, and IV will be artificially reduced.

The magnitude of this effect varies with option maturity and moneyness, and can be

---

lifetime could require volatility to be negative thereafter, which is obviously impossible.

surprisingly large even for bid-ask spreads that are commonly observed in practice. Tables III.1 and III.2 provide an illustration. In Table III.1, we examine the implied volatilities calculated for one- and six-month European call options on a non-dividend paying underlying asset when there is a bid-ask spread and transactions prices are only permitted at discrete price Aticks@ (1/8 or 1/16 of a point, for option prices above or below 1, respectively). We assume the true volatility is 0.250 (25.0 percent annually), the interest rate is 8.0%, option strike prices are 45, 50 and 55, and there is no bid-ask spread to buy or sell the underlying asset at its current price of 50.

The table displays results for the two maturities in separate panels. The first three lines of each panel show the effect of price discreteness alone. For example, the 1-month 45 strike in-the-money call has a Black-Scholes theoretical value of 5.386. If this were the price in the market, implied volatility would be exactly 0.250. However, as is the practice in U.S. equity option markets, the minimum tick for an option in this price range is 1/8, so the market price would likely be rounded to 5 3/8 (5.375). At this price, IV is 0.243. The at-the-money 50 strike call would be rounded up from 1.617 to 1 5/8 (1.625), leading to an IV slightly above the true value, and the out-of-the-money 55 strike call, trading below a price of 1 where tick size is smaller, would go to 3/16 and an IV of 0.244. The effect of this rounding error on the IV calculation is smaller for 6-month options because they have greater time value. This means a price difference of a given size makes a smaller proportional impact on the time value and therefore on the implied volatility.

Introducing a bid-ask spread into the options market leads to considerable noise in the computed IVs. A spread of 1/4 point on a 5 dollar option is smaller than what is normally seen in the U.S. equity options market. Even so, with quotes of 5 1/4 bid to 5 1/2 ask, a trade at the ask in the 1-month in-the-money call would produce an IV of 0.321, more than 7 percentage points too high. A more customary spread of 1/2 point would yield an extremely high IV of 0.398 at the ask price of 5 5/8.

More problematical still is the fact that the IV can not be computed at all for a trade at the bid price even when the spread is only 1/4 point. The problem is that deep-in-the-money short maturity contracts are relatively high priced, but most of their value comes from the option=s intrinsic value, and very little is related to volatility. For example, with no volatility at all in the underlying (σ = 0.000), the 1-month 45 strike call would be priced at its arbitrage-based lower boundary value. A call option is always worth at least the stock price less the present value of the strike price, or $S - X(1+r)^{1/12}$ = 5.293 in this case. Thus, the effect of 25% volatility is only to raise this option=s fair value from 5.293 to 5.386, i.e., less than 1/8 of a point. There is no positive volatility that can make 5 1/4 the correct price because that would violate the lower bound. (Buying the option at 5 1/4 and investing the present value of the 45 dollar strike price in riskless bonds would create a position that would cost less than buying the stock alone but would have a better payoff; such portfolio dominance is inconsistent with market equilibrium.) The other options in Table III.1 also exhibit substantial noisiness in the IV calculation due to the bid-ask spread, but it is quantitatively smaller.

In Table III.2, we examine the effect of a bid-ask spread on the underlying asset. Here we assume that the option is priced at its (rounded) Black-Scholes value, without a spread, while the stock quote is 50 ± half of a bid-ask spread of 1/4, 1/2, or 1 point. The combined effect of spreads in both markets may be additive (e.g., if the stock price is at the bid and the call price is at the ask), or offsetting (e.g., if both are at the bid). Once again, we see that the effect on IV can be sizable, particularly for nearby deep-in-the-money contracts.

This strong effect of market price noise on IVs calculated from in-the-money options can explain at least some of the differences across strike prices. If recorded transactions from deep-in-the-money calls come half from trades at the ask and half from trades at the bid, IVs computed from the ask prices will tend to be very large, while those from the bids will, in many cases, be impossible to calculate, so those data points will be discarded from the sample. As Canina and Figlewski [1993] observed in their extensive data set on S&P 100 stock index options, this can lead to a serious upward bias in the Aaverage@ IV for these options. They found that simply entering IV values of 0.0 for those Abad@ data points instead of dropping them from the sample produced average IVs for in-the-money calls that were very close to the values found for at-the-money contracts.

In principle, random noise from the bid-ask spread can be eliminated by using not transactions prices, but the midpoints between bid and ask in each market. This is not often done, however, due to lack of intraday bid and ask price data. The value of using such quotes when they are available also depends upon the quality of the data. While transactions are real market events, in the absence of trades, posted quotes may become stale and no longer representative of where the market really is. Thus, attempting to eliminate noise by employing bid and ask quote data may simply substitute one form of noise for another, without producing much improvement overall.

Along with these factors, there are any number of less clearly identifiable sources of noise in market supply and demand conditions that will affect measured option prices, and therefore enter into the calculation of implied volatility. Researchers frequently try to deal with noise in implied volatilities by averaging across a number of options on the same underlying. In some cases, a simple average is used (often known by the acronym AISD for Aaverage implied standard deviation@). But, for several reasons, taking a weighted average (WISD, for Aweighted implied standard deviation@) is more common. One reason is that some options are traded with much greater liquidity than others, so their reported prices are expected to contain more reliable information.

Another reason for weighting is to adjust for differing sensitivities of option values to the volatility parameter. As we have just seen, for options that are either deep in- or out-of-the-money, a small, and perhaps economically insignificant, price change has a big impact on implied volatility. Since price noise will then be amplified into relatively large inaccuracies in IV, these options are downweighted in the averaging. Chiras and Manaster [1978], for example, used the elasticity of each option=s price to volatility in weighting its IV in their WISD calculation.

A third approach to creating a combined IV estimate is to used least squares. The IV estimate is taken to be the σ* that minimizes the following expression:

$$\sum_{i=1}^{K} (\sigma_i - \sigma^*)^2$$

where each σ_i represents the IV computed from one of the K available options. Both Beckers [1981] and Whaley [1982] adopt this technique.

To illustrate a typical case with a multiplicity of IVs from different options on the same underlying, we computed implied volatilities as of the close of trading on November 16, 1992 for a set of Treasury bond futures call options based on the Chicago Board of Trade=s March 1993 bond futures contract. The futures price on that date was 102 4/32 and option strike prices were 2 points apart, ranging from 94 to 114, that is, from deep in-the-money to deep out-of-the-money. Unlike options on bonds themselves, which are complex because theoretical values may depend on the dynamics of the whole yield curve, options on futures can be priced easily using Black=s variant of the Black-Scholes model.[15] Thus, there is little problem here with the option pricing model itself. Even so, implied volatilities range from around 9% to over 12% across different strike prices.

Figure III.1 displays the very regular pattern of implied volatilities from these T-bond futures options. At-the-money options have the lowest implied volatilities, and IVs rise monotonically as one moves to lower (in-the-money) or higher (out-of-the-money) strikes. This Aclassic@ U-shaped relationship between IV and moneyness is known as the volatility Asmile.@

---

[15] Black=s [1976b] model for the value of a call option on a futures contract is given by

$$C_{FUT} = e^{-rT} (F\, N[d] - X\, N[d - \sigma\sqrt{T}]),$$

where F is the futures price, $d = (\ln F/X + \sigma^2 T/2)/\sigma\sqrt{T}$, and the other symbols have their standard meanings.

Although precise details vary from market to market and over time within a given market, a smile is very common, to the point that it is unusual to find a market that does not exhibit something like it. In some cases, only one side will have a strong upward curvature, making a Askew@ or Asmirk.@

Objective consideration of this pattern, however, must lead one to the conclusion that it seriously calls into question the ideas, and standard operating procedure just discussed. Suppressing IV differences across different options on the same underlying by an averaging procedure such as we have described can only be appropriate when the differences are due entirely to random sampling noise. But where a regular and pervasive smile relationship exists between relative implied volatility and moneyness, it can not be attributed to randomness; it shows that option prices are systematically different from what the pricing model (i.e., the one used to compute the IVs) says they should be. This must be considered strong evidence that the market is valuing options using a different model from the one the analyst is assuming. If so, there is no reason to think that implied volatilities computed from the wrong model, whether examined individually or combined into a weighted average, will yield the market=s true estimate of the volatility of the underlying asset.

III.3 Testing Whether Implied Volatility is a Rational Forecast

If implied volatility is the best possible forecast of the volatility of the underlying asset given all widely available information, a large amount of time and effort can be saved that would otherwise be devoted to gathering and analyzing historical data and other information to make independent volatility predictions. Naturally, therefore, many tests of forecast rationality have been conducted on implied volatilities in different markets. The most common test is easily performed in the context of a simple regression.

Given an information set, a rational forecast is the (true) expected value of the variable in question conditional on that information set. That is the hypothesis about implied volatility expressed formally in equations (III.2) and (III.3). By definition, the realized value of a stochastic variable is its expected value plus a zero mean random disturbance. So if implied volatility is to be a rational forecast of future volatility, it must be the conditional expected value of future volatility given the market=s information, and we will have

$$\sigma_{ACTUAL} = \sigma_{IV} + \varepsilon \qquad \text{(III.5)}$$

with

$$E[\varepsilon] = 0; \quad E[\sigma_{IV}\ \varepsilon] = 0$$

The forecast error has mean zero and is uncorrelated with the forecast.

Note that different information sets will naturally give rise to different conditional expectations, and a more inclusive information set will yield a more accurate forecast. Nevertheless, the equivalent of equation (III.5) must hold for every rationally formed forecast. Better information will simply reduce the variance of the forecast error ε.

Equation (III.5) leads naturally to the following regression model.

$$\sigma_{REALIZED}(t) = \alpha + \beta\,\sigma_{IV}(t) + \varepsilon(t) \qquad \text{(III.6)}$$

The statistical test for unbiasedness is then the joint test that $\alpha = 0$ and $\beta = 1.0$.[16] In what follows, we will refer to this equation as the Arationality test@ regression.

The regression-based approach to testing forecast rationality can easily be extended to look at the relative information content of different data sets. In particular, let $\sigma_{HIST}(t)$ denote the

16 Theil [1966] is credited with introducing this test for forecast unbiasedness, and it has been widely applied since then. Brown and Maital [1981], for example, present results of running this test on survey data covering forecasts for a broad variety of economic variables. Below, we discuss results of this test on implied volatilities drawn from a number of papers in the finance literature.

sample volatility calculated at time t from a set of past prices. Since the set of historical price data is a subset of the market=s information set, if both σ_{IV} and σ_{HIST} are rational forecasts, it still must be the case that $E[\ \sigma \mid \sigma_{HIST}\ ,\ \sigma_{IV}\ ]\ =\ E[\ \sigma \mid \sigma_{IV}\ ]$. The expected value of realized volatility given both historical and implied volatilities is no different from the expected value given implied volatility alone. This can be examined by running a so-called Aencompassing regression@ like (III.7):[17]

$$\sigma_{REALIZED}(t) \;=\; \alpha \;+\; \beta_1\,\sigma_{IV}(t) \;+\; \beta_2\,\sigma_{HIST}(t) \;+\; \varepsilon(t) \qquad \text{(III.7)}$$

If both implied and historical volatilities are rational conditional forecasts, but historical volatility is based on only a subset of the market=s information, regression (III.7) should yield coefficient estimates of $\alpha = 0$, $\beta_1 = 1.0$ and $\beta_2 = 0$.

Regression (III.6), the equivalent Arationality test@ regression with σ_{HIST} in place of σ_{IV}, and (III.7) have been fitted to implied volatilities in many published studies, some of which we will discuss in greater detail in the next section. Here we will describe the results from one large study, by Canina and Figlewski [1993] (CF), of call option implied volatilities based on the S&P 100 stock index, commonly known by its ticker symbol as the OEX index. This is one of the most actively traded option contracts in the U.S.. One reason to focus on this particular study is that it deals carefully with a number of important statistical and data problems that all such studies are exposed to. A second reason is that it examines one of the most active and important options markets, and the results obtained are very striking.

Conceptually, the regression tests for forecast rationality are simple, but implementation raises a variety of potential problems with both the data and the estimation technique. Let us begin by reviewing the basic arbitrage mechanism by which the market=s volatility forecast is supposed

---

[17] Fair and Shiller [1990] present the encompassing regression methodology in the context of an examination of different macroeconomic models.

to be incorporated into market options prices. In the theoretical world of the Black-Scholes pricing model, investors are assumed to know the volatility of the underlying asset. They use the model to compute fair values for all available option contracts, and any mispriced option is sold if it is overpriced, bought if underpriced, and then hedged against the underlying asset. A dynamic hedging strategy locks in the mispricing as an arbitrage profit that will be earned over the lifetime of the option. By this arbitrage trading, the market=s volatility forecast is incorporated into option prices.

This brings out a number of important points regarding the data to be used in testing forecast rationality, some of which we have already mentioned. First, the prices for the option and the underlying asset must reflect the terms under which an actual arbitrage trade could have been executed. Infrequent trading of less liquid options can cause reported closing prices to be nonsynchronous, but the fact that many options markets remain open later than the markets for their underlying assets (15 minutes later, in the case of U.S. equity options) introduces timing problems even for actively traded contracts. Second, prices drawn from recorded transactions will produce different implied volatilities, depending on whether they were done at the bid or the ask. This introduces noise into the calculated IVs, that in turn creates an Aerrors-in-variables@ problem for the regression.

Third, naturally, accurate values for the other model inputs must be obtained. The level of the underlying, strike price, and time to expiration are all unambiguous, but the interest rate and dividend payout (required for options on underlying assets that pay out cash during the option=s lifetime) are less certain. It is common to use U.S. Treasury bill yields to measure the riskless interest rate because they are clearly free of default risk. They are also available with maturities every week for the next six months, so precise matching of the interest rate and option maturities is possible. One can argue, however, that T-bill rates are not the most appropriate choice to measure the relevant interest rate in an options arbitrage, for several reasons. T-bills are, at best, only a rate at which funds can be lent; they do not reflect the cost of funds to an arbitrageur who must borrow money to finance a position. Also, T-bill interest is tax exempt at the state and local level, which depresses the rate slightly relative to comparable maturity fully taxable interest rates. Finally, bill rates are also lowered artificially due to the additional market demand for T-bills as collateral, that other money market instruments do not have. As an alternative, Canina and Figlewski use the average of the Eurodollar deposit rate and the brokers call rate as a proxy for the riskless interest rate facing an options arbitrageur.

Measuring the expected dividend payout over the option=s lifetime also may present problems. Since equity dividends tend to be smoothed by firm management, they are fairly predictable over the short run, but IVs from longer term options will be exposed to noise from dividend uncertainty (and time-variation in interest rates, as well). CF use realized dividends on the OEX index, assuming that these were fully known to traders, at least for the next four months.

Another class of issues has to do with the choice of the proper option valuation model to use in calculating the implied volatilities. This is very much a problem in examining interest rate options, since there is no general agreement on the best way to model term structure behavior. For European equity options, the Black-Scholes model, with adjustment for dividends, is universally used, and Black=s variant for futures options is adopted for options on futures and forwards. Although early researchers tended to use Black-Scholes even for American options, it is not hard conceptually (and only a little cumbersome in practice) to take account of the value of early exercise of American options by using any of a variety of numerical approximation methods, the most common of which is probably the binomial model.

Much more problematical conceptually is the fact that the reason one is interested in forecasting volatility in the first place is that it is expected to vary randomly over time, but the Black-Scholes and related models all treat volatility as a known parameter. There is clearly a logical inconsistency in using a fixed volatility model to analyze options whose prices have been established under conditions of stochastic volatility in the market. Nonlognormality--most importantly in the form of discontinuous jumps in the price series and fat tails of the returns distribution--is another reason that the standard Black-Scholes family of models can not be expected to reflect everything that is going into market options prices.

The difficulty is that models that incorporate stochastic volatility or nonlognormality are much more complicated and harder to use than Black-Scholes. Nor is there general agreement on the particular model to adopt as an alternative.[18] Such models are therefore not widely used by market participants, who typically prefer instead to compute BS values and then adjust them in various ad hoc ways to deal, subjectively, with the known shortcomings of the basic model. Note that if one wishes to obtain the market=s volatility estimate, it is necessary to compute IV from the pricing model the market is using, even if in theory this is not the most accurate model available.

Lastly, in computing realized volatility, the dependent variable in regression (III.6), it is necessary to match the horizon to the option=s lifetime, since that is the volatility parameter that should go into the pricing model. Early researchers such as Latané and Rendleman [1976] did not always do this, nor is it universally done in such studies even today. More ambiguous is how much past data to include in the calculation of historical volatility, σ_{HIST}. Traders often consider only data from the very recent past, e.g., the last one to three months, and academic researchers frequently do likewise. Yet, from the results presented in Chapter II, this may produce far from the most accurate forecasts that can be obtained from historical prices.

It is a common shortcoming among academic studies on this subject to devote a great deal of time and effort in attempting to obtain the best possible implied volatilities from individual option prices and to combine them into a single optimally weighted IV, but then to pick a procedure for computing historical volatility seemingly at random, without any apparent effort to

[18] Ball [1993] gives an excellent review of stochastic volatility models and option pricing.

find the best performer among the possible alternatives. If one wants an honest Ahorse race@ between approaches, one ought to compare the best methods in each category.

Of course, the issue of whether to estimate volatility by taking deviations from the sample mean also arises. CF compute historical volatility, around the sample mean, from the previous 60 calendar days, but report that experiments with both longer and shorter historical samples found little difference in the results.

One final major issue arises in this kind of empirical testing. An enormous amount of price data is produced by options trading in an active market, even without using intraday observations. For a single underlying asset, there will normally be concurrent trading in a number of options, both calls and puts, with different strike prices and maturities. CF, for example, look at closing OEX call option prices for 8 different strikes and 4 maturities on each day for 4 years, for a total of over 17,000 observations. However, since the test involves examining the forecast errors in predicting volatility over horizons of up to 127 days in the future, these multiple observations are far from independent. Indeed, the forecast errors for any two options whose remaining lifetimes overlap even partially may be expected to be correlated.

Researchers have traditionally dealt with this problem by combining IVs from options with the same maturity but different strikes by taking a weighted average, as discussed above, and by discarding observations with overlapping forecast horizons. Treated in this way, a data set like that analyzed by CF would produce only 6 observations per year for two-month options, a total of 24 data points. CF, however, introduce a statistical procedure to use all of the available data and to correct the regressions for the effects of cross correlation in the residuals. This allows them to employ over 850 observations from at-the-money options with 1 to 2 months to expiration.

Briefly, the procedure is as follows. Ordinary least squares is well-known to produce consistent coefficient estimates even with correlated residuals, but the estimated coefficient standard errors will be biased by the lack of independence. Following the work of White [1980] and Hansen [1982], however, the OLS residuals can be used to construct a consistent estimate of the coefficient variance-covariance matrix.

Let u_n , n = 1,...,N denote the OLS residual from the nth data point and X_n be the row vector of right hand side variables for that observation. For example, $X_n = (1\ \sigma_{IV})_n$ for regression equation (III.6) and $X_n = (1\ \ \sigma_{IV}\ \ \sigma_{HIST})_n$ for equation (III.7). Define Q(k,n) to be an indicator function that takes the value 1 if there is an overlap in the forecast horizons of observations k and n, and 0 if there is no overlap. Then compute

$$\Psi \;=\; \sum_{n=1}^{N} u_n^2\, X_n{}'X_n \;+\; \sum_{k=1}^{N-1}\sum_{n=k+1}^{N} Q(k,n)\, u_k\, u_n\, (X_n{}'X_k + X_k{}'X_n) \qquad \text{(III.8)}$$

(Note that this is a corrected version of eq.(6) in Canina and Figlewski [1993]). The consistent estimator Ω of the true covariance matrix for the coefficients is then given by

$$\Omega \;=\; (X'X)^{-1}\,\Psi\,(X'X)^{-1} \qquad \text{(III.9)}$$

where X denotes the N x 2 (for eq. (III.6)) or N x 3 (for eq. (III.7)) data matrix of the X_n.

Using a Monte Carlo simulation, CF show that this procedure works well to produce consistent standard errors and leads to a substantial increase in the amount of information that can be extracted from a data set such as theirs. For example, for 1 to 2 month options, the standard errors on the estimated coefficients are reduced by more than a factor of 6 relative to what would be obtained by discarding all overlapping observations.

Table III.3 reproduces some of the results reported in CF. Running regression (III.6) on the entire sample of 17,606 observations, yields values of $\alpha = 0.136$ and $\beta = 0.022$. This represents a total rejection of the hypothesis that implied volatility in this market is a rational forecast of the volatility that will be realized over the option=s lifetime. The constant term is significantly greater than 0 and the slope coefficient is (highly) significantly less than 1.0. Indeed, the estimate on β is not even close to being significantly greater than zero! These regression results indicate that instead of being the best possible forecast, in this market implied volatility appears to contain no information at all about future realized volatility.

To analyze this surprising result further, CF estimate equation (III.6) on subsamples of the options data, to determine whether the lack of fit is due to the effects of less liquid contracts that were deep in- or out-of-the-money, or far from maturity. Generally the options that trade with the greatest liquidity, and are therefore expected to be the most efficiently priced, are those that are at- or slightly out-of-the-money, and near, but not too near, to expiration. The second part of Table III.3 reports results from regressions on implied volatilities from 1 to 2 month options that are at-the-money to 5 points out-of-the-money. Although the α coefficient falls somewhat and β rises, these results are hardly any better than those from the full sample, and still indicate a strong rejection of the rationality hypothesis.

If OEX volatility were extremely hard to predict accurately, perhaps no forecast could achieve a significant slope coefficient. But this possibility is contradicted by the rationality test regression run on historical volatility. Although rationality is still rejected (with both $\alpha > 0$ and $\beta < 1.0$, significantly), historical volatility does contain some information about future realized volatility, since the slope coefficient is at least significantly positive. Finally, the failure of the rationality hypothesis is reconfirmed in the encompassing regression test, where the estimated slope on implied volatility is negative and insignificant, while historical volatility ends up with virtually the same coefficient as in the univariate regression.

How should one interpret these extraordinary results?

One possibility is that option traders are irrational, and they systematically ignore readily available information that would permit them to compute more accurate theoretical option values.

In other words, one might attribute the results in Table III.3 to a failure of the relationship embodied in equation (III.3) and conclude that the average trader=s subjective volatility expectation is a poor estimate of the true conditional expected value of future volatility, given the publicly available information. We can not eliminate that hypothesis without direct observation of the market=s volatility expectations, $E_{MKT}[\sigma]$. But the great bulk of empirical evidence from research on competitive financial markets suggests that investors generally make good use of relevant information in pricing securities. It would be strange indeed if the stock index options market turned out to be the one major exception.

This points to failure of equation (III.2) as the probable explanation: In this market the implied volatility may not be a good measure of the market=s true expectation about future volatility. In other words, investors could be making appropriate use of available information about the volatility of the underlying stock index in forming their expectations, but these expectations do not translate properly into option prices in the market. This is our preferred explanation for the failure of the rationality test regression to show any predictive power for OEX implied volatilities.

In support of this interpretation, consider the difficulty a trader would face in trying to profit from a belief that an OEX option=s price did not accurately reflect the best forecast of future volatility. In theory, an investor who knows an option to be mispriced should lock in a riskless excess return by forming a hedged position with the option and the underlying asset, and rebalancing the proportions frequently (continuously) over the option=s lifetime. But following such a trading strategy when the underlying is a large portfolio of 100 stocks is very expensive in terms of transactions costs, and is also exposed to considerable execution risk. And, of course, no investor can be sure of knowing the true volatility. The fact is that relatively few such arbitrage trades between the cash index and an index option are actually done.[19] OEX option traders are

[19] Neal [1993] presents data on all program trades done at the New York Stock Exchange during a 3 month

much more likely to hedge their positions with S&P 500 index futures (which means, among other things, that if the futures contract is mispriced relative to its index, the options will tend to be mispriced also, relative to their underlying cash index).

Thus the arbitrage trading by which investors= volatility forecasts are assumed to be impounded in OEX option prices turns out to be quite expensive and risky to follow in practice. And at the same time trading options by betting that one=s volatility estimate is a more accurate forecast than that contained in implied volatility is unlikely to be an optimal strategy for an active options market maker. Rather, basing option bids and offers on the current implied volatility (whether or not it is an accurate prediction) will lead to transactions that are largely balanced between buying and selling, and the resulting turnover should produce an ongoing flow of profits from the bid-ask spread.

In this market, therefore, the difficulty of following the trading strategy that would cause investors= volatility expectations to enter into option prices combines with the activities of options market makers whose trading is based primarily on considerations of current supply and demand in the market. The result is that option prices may drift relatively far away from the Black-Scholes values without producing much equilibrating trading.

If this reasoning is at least a partial explanation for the failure of the rationality test regression in the CF study, it suggests a hypothesis about pricing in different option markets. In markets like this one, where the arbitrage between options and the underlying asset is hard, or impossible, one might anticipate relatively poor performance of implied volatility as a forecast of future volatility. But in markets where the arbitrage is fairly easy and inexpensive, such as those for futures options (where the option and the underlying future trade on the same trading floor with

---

period, from January through March, 1989. Those that involved trading the basket of OEX stocks totaled 650 transactions, for a total of about 125,000 contracts over the period. This includes all trades to initiate and unwind positions and also those done to rebalance hedges. But it represents only 0.77 percent of the trading volume in OEX options for that period.

low transactions costs) or currency options (in which the underlying is simply money), one would expect implied volatility to contain much more information. In the next section we present results from a variety of studies of implied volatilities in different markets that lend some support to this hypothesis.

III.4 The Information Content of Implied Volatility in Different Markets

The rationality test regression, equation (III.6), has been examined in many studies. We will not attempt to review even a representative sampling of them all here.[20] Rather, in this section we will describe the regression results from several very good articles in the literature covering a variety of different options markets. We report selected results, in most cases transforming them from the original versions into a common format for easier comparison.

Table III.4 summarizes rationality test results from six papers, covering OEX stock index options (Canina and Figlewski [1993], Fleming [1996]), options on individual stocks (Beckers [1981]), options on crude oil futures (Day and Lewis [1993]), foreign currency futures options (Jorion [1995]), and spot currency options on the Deutschemark (Ferri [1996]). The first section of the table describes the data sets used and the second section summarizes the results from the rationality test regressions. The seventh paper, by Lamoureux and Lastrapes [1993], looking at encompassing regression models for individual stock volatilities will be discussed separately.

First is the paper just discussed above, by Canina and Figlewski. The data set is drawn from more than 17,000 S&P 100 stock index (OEX) call option daily closing prices over a four year period, from 1983 to 1987. Separate regressions are fitted for 8 strike price categories (from 20 points out-of-the-money to 20 points in-the-money) and 4 maturities (effectively 1, 2, 3, and 4 months). Historical volatility is computed over the previous 60 calendar days. An important

[20] See Mayhew=s [1995] fine review of implied volatility for a more complete discussion of the literature on this topic.

innovation in this paper is using a large sample of daily data and correcting statistically for the effect of cross-correlation in the residuals due to overlapping forecast horizons, as described above.

The results for the full sample were shown in Table III.3. In repeating them here, we convert from the format of the original published article to express the statistical significance of the fitted coefficients in terms of t-statistics on the hypotheses that the constant is 0.0 and the slope is 1.0. Thus, the full sample regression shows that the estimated constant of 0.136 is 11.3 standard deviations above 0.0 ($t_0 = 11.3$), while the t-statistic is (negative) 19.6 on the hypothesis that the true value of the slope coefficient is 1.0 even though the fitted value is only 0.022.

Regressions are estimated for each strike and maturity combination, but none of them come close to passing the rationality test. For comparison, 32 regressions are fitted using historical volatilities, matched by dates to the option maturity/moneyness subsets. (For example, if the 12.5 to 17.5 points out-of-the-money calls did not trade on a particular date within the sample period, that data point is also removed from the matching regression with historical volatilities.) The second set of results from CF in Table III.4 reports the averages of the coefficients from these 32 (highly similar) regressions. Although this shows that historical volatility contains more information about future realized volatility than the IVs do, the t-statistics indicate clear rejection of rationality here, too.

There are numerous possibilities for extracting more information from a set of options prices than what CF attempted. One method, as discussed above, is to average across IVs from several simultaneously observed options with different strike prices. Another is to average over multiple intraday observations on the same options, in order to reduce the impact of price noise from the bid-ask spread. In his study of at-the-money OEX calls and puts, Fleming [1996] uses transactions data and averages over all trades during the last ten minutes of trading each day. By restricting consideration only to at-the-money contracts, he ends up with fewer data points than

CF, but each one contains information from a number of recorded option prices, and his sample spans a longer time period, from October 1985 to April 1992. One special feature in this paper is that to avoid possible estimation problems that arise if actual and implied volatilities are completely cointegrated series, the regression is fitted in first difference form.

Fleming=s results differ markedly from those obtained by Canina and Figlewski for the same market. Although call IV, put IV, and historical volatility from the previous 28 calendar days all still fail the rationality test on the fitted coefficients, at least all of the slopes are around 0.6, and are statistically significantly greater than zero. Coefficients on the averaged implied volatilities now appear comparable in size to those on historical volatilities, and the R^2 suggests that they may contain more information about future volatility.

Beckers [1981] looks at implied volatilities from traded call options on individual stocks over 10 subperiods of 3 months each (i.e., each subperiod contains one expiration for each stock) between April 1975 and July 1977. This study is the best of those done in the 1970s in terms of the way the tests are set up and the data is handled. Beckers considers different averages of implied volatilities from options with different strikes, but concludes that the at-the-money option=s implied volatility by itself is as good a predictor as the averaged IVs.

In each subperiod, IVs are computed for 5 consecutive days and then averaged, producing one observation for each stock. The regression is then run on the cross section of from 62 to 116 stocks depending on the period. Realized volatility is computed for the period up to option expiration, and historical volatility is estimated from daily prices over the previous quarter.

Table III.4 presents the averages of the coefficients, and the t-statistics, from the 10 subperiods. For this market, we now see that the 0 and 1.0 target values for the constant and slope are close to being realized, with average constant terms below 0.005 and average slopes of 0.813 and 0.673 for implied and historical volatilities. Still, these slopes are about 10 standard deviations below the rational value of 1.0. One important feature here is that implied volatilities in this market seem to be Aless biased@ than historical volatilities (the slope coefficient is higher).

These results for individual stocks are consistent with the notion that the ease of executing an arbitrage trade between the option and the underlying asset will determine how far market option prices will be allowed to deviate from their model values, and therefore, how much information about the true theoretical values will be contained in the implied volatilities computed from those prices. Another feature of this research design that tends to make Beckers=s results stronger is that it includes both time series variation and cross sectional variation within the data sample. This means that there is considerably more in-sample variability in realized volatilities that can be explained by IV differences. The substantially higher R^2 statistics for these regressions show this effect clearly.

Day and Lewis [1993] compare implied volatility from call options on crude oil futures to a simple historical volatility and also to out-of-sample forecasts from GARCH and EGARCH (Exponential GARCH) models.[21] Each of the conditional heteroskedasticity models is fitted in a form that also specifies a time-varying mean return, but this does not have much effect on the volatility forecasts from those models. The data sample consists of daily closing prices for oil futures and at-the-money calls. Historical volatility is estimated over a time period of equal length to the time remaining to option expiration, and the parameters for the GARCH-M(1,1) and EGARCH-AR(1)(1,1,1) models are refitted for each date using only historical data from the previous 500 days= futures prices. The (1,1) and (1,1,1) notation indicates that the models each contain one lagged variance and one lagged squared disturbance (ε) term. The EGARCH model is also set up to allow an asymmetric volatility response to positive and negative price shocks.

Table III.4 shows the results from the rationality regression for the second nearest to

[21] EGARCH is another member of the ARCH family of models for time-varying volatility. Its most important difference from GARCH for this application is that it permits positive and negative disturbances to have different impacts on subsequent volatility. See Nelson [1991] for a full exposition of the EGARCH model.

expiration contracts, with average maturity of 32 trading days. Here the slope coefficient for implied volatility is 0.880 and the constant is 0.003, neither of which is significantly different from its theoretical value. The joint hypothesis on the two coefficients together is also satisfied (p-value = 0.62), and the regression R^2 is 0.718, which indicates extremely good forecasting performance for implied volatility in this market. The simple historical volatility estimate also performs much better in this market than in those we have just looked at, passing the joint rationality test (p-value = 0.10) even though the slope coefficient is only 0.61. (Note that the t-statistic we have calculated for the constant term in this case is subject to substantial rounding error.) Both ARCH family models, however, fail the rationality test at better than the 0.01 level, with EGARCH doing substantially worse than GARCH.

This is the first example among these papers in which IV (or any volatility forecast) has passed the rationality test. It should be noted, however, that none of these models passes the test for the longer maturity 4 month option contracts in this study. Even though the estimated IV slope only falls to 0.829, rationality is rejected at the 5% level for this maturity. This market, where the option and the underlying are traded with very low transactions costs on the same trading floor, is one where arbitrage is perhaps the easiest. And, of course, the shorter the option maturity, the easier it will be to maintain a hedged position over the remaining lifetime of the contract. Thus both results tend to support our hypothesis that the information content of implied volatility is directly related to the ease of executing an arbitrage between the option and the underlying asset. The estimated slopes are close to 1.0, but only the prices for the short maturity oil option contracts reflect fully rational use of the market=s information about future volatility.

In the next study reported in Table III.4, Jorion looks at the information content of implied volatility in futures options on three foreign currencies. The data set consists of daily closing prices for contracts traded on the International Monetary Market of the Chicago Mercantile Exchange, from various points in 1985 through February 1992. Implied volatilities are computed for the at-the-money call and put on the nearest to expiration future. These are averaged to produce one IV per currency per day. Historical volatility is estimated over the previous 20 trading days. A GARCH(1,1) model is also tested, but not in a true forecasting mode: The model is only fitted once, over the full sample, and is used to compute in-sample 1-day volatility estimates for each date, not a series of volatilities predicted for the option=s remaining lifetime.

Table III.4 reports averages for the constant and slope coefficients across the three currencies. Although the study considers futures options, which ought to be easy to arbitrage, the results of the rationality test are more similar to those from Fleming=s study of stock index options than to the crude oil futures options examined by Day and Lewis. Implied volatility does appear to have predictive ability, but the slope coefficient only averages 0.521, which is significantly less than 1.0, and R^2 is only 0.133 on average. Historical volatility (from only 20 trading days) does substantially worse, with an average slope of only 0.169, and a lower R^2. While the GARCH results appear to be fairly good, that is a misleading impression. First, recall that these are in-sample estimates, not what would have been produced in practice in a trading situation. Second, the estimated standard errors are much larger on these coefficients than for the implied or historical volatilities, which contributes to failure to reject the hypothesis that 0.715 is significantly less than 1.0. The difference in information content is revealed by the low R^2 and by the results of encompassing regressions with the IVs, in which the GARCH forecasts end up with a negative weight for 2 out of 3 currencies.

We looked for published results from running the rationality regression with traded options on spot exchange rates, rather than FX futures, such as those traded on the Philadelphia Stock Exchange (PHLX), but did not find anything suitable. However, Ferri has been examining this market in his Ph.D. dissertation that is currently in progress, and has obtained preliminary results for PHLX options on the Deutschemark-dollar exchange rate. The data set consists of all transactions from the period Jan. 3, 1984 through Feb. 23, 1995 (with two months missing). The results reported here are drawn from DM calls and puts with strike prices from 10 percent out-of-the-money to 10 percent in-the-money and next-nearest to expiration maturities, i.e., from 1 to 2 months. The estimation method used by Canina and Figlewski is adopted, which allows over 70,000 observations each for calls and puts to be analyzed. Like most of the other studies whose results are reported in Table III.4, Ferri finds that the IVs from these options contain information, but do not pass the rationality test. The constant terms are significantly greater than zero and the slopes are between .0.5 and 0.6, significantly less than 1.0. The regression on historical volatility calculated from the previous 60 days exchange rates, shows a distinctly smaller slope coefficient and R^2 statistic than the IV regressions.

III.5. Forecasting Volatility with Biased Forecasts

The results reported in the last section show that, in the statistical sense of equations (III.2) and (III.3), implied volatility is not generally a rational forecast of the volatility that will be realized in the future. In this section, we consider what it is appropriate to do next.

Virtually all published studies reporting results from the rationality test regression find estimates for the constant and slope coefficient to be significantly different from 0 and 1.0, respectively, for all volatility forecasts. This is true whether the forecasts are obtained by simple or complex means from historical data, as implied volatilities from market option prices, or in any other way. These pervasive negative results have led most researchers to focus on different properties in assessing which prediction method to adopt. Most tests of forecast rationality do indicate that implied volatility contains information about future volatility even though IV as a forecast is biased.

It is common, for example, to compare R^2 values or t-statistics on the slope coefficients from equation (III.6) run on alternative volatility forecasts and to conclude that the one with the best fit is the preferred forecast. The winner by this criterion is frequently the implied volatility, as

Table III.4 suggests. Comparing competing forecasts in the context of an encompassing regression like equation (III.7) is also common, and again IV often appears to dominate forecasts based on historical data alone, both the simple sample standard deviation computed from recent data, as well as the more sophisticated forecasts from ARCH-family and other time-varying volatility models.

These comparative results that generally favor IV over statistical estimates tend to bolster the preconceptions of academic researchers who expect financial markets to be informationally efficient. From past experience, they are not surprised that IV fails the rationality test, since other prediction methods fail, too, and they are comforted that IV at least appears to be the best of the available forecasts. They also emphasize the many sources of noise and other potential data problems that might cause apparent failure of IV rationality, even if it is true. Thus, it is widely believed that, although it is somewhat flawed, implied volatility is the best available forecast of future volatility and also the best measure of the market=s volatility expectations. Academics use it in valuing options and for other purposes where a proxy for expected volatility is required.

Unfortunately, these conclusions are wrong. Viewed objectively, it should be obvious that if IV is not a rational forecast, it can not be rational to value options using IV as one=s prediction of volatility. Further, taking IV as a proxy for the market=s assessment of risk on the underlying asset inherently involves building into one=s model of expectations the assumption that investors are irrational (i.e., by this assumption the market is informationally inefficient).

The following slightly contrived example demonstrates why producing a much higher R^2 than historical volatility in the rationality test regression and receiving most or all of the weight in an encompassing regression still need not make IV a better volatility forecast than an alternative predictor.

Consider a stock whose volatility σ_t ranges between 0.15 and 0.25 with a uniform distribution, and let there be two different methods that produce forecasts F_{1t} and F_{2t}. The relationship between the forecasts and the true volatility is given by:

$$F_{1t} = 2\,\sigma_t$$

$$F_{2t} = 0.20 + \eta_t ,$$

where η_t represents a very small zero mean random perturbation (introduced here only to prevent perfect multicollinearity between the forecast and the constant term in the rationality regression).

Clearly, the first forecast contains much more information about the true volatility than the second forecast does. Running the regression (III.6) with F_1 will produce $\alpha = 0.0$, $\beta = 0.5$, and $R^2 = 1.0$, while the regression on the essentially constant F_2 will yield $\alpha \cong 0.20$, $\beta \cong 0.0$, and $R^2 \cong 0.0$. An encompassing regression with both forecasts will give $\beta_1 \cong 0.5$ and $\beta_2 \cong 0.0$. With these results, many analysts would declare IV to be the superior forecast and proceed to use it confidently as the volatility input to option valuation models and as a measure of the market=s (informationally efficient) expectation of future volatility.

But consider the <u>accuracy</u> of the predictions made by the two models. When $\sigma_t = 0.15$, $F_{1t} = 0.30$ and $F_{2t} \cong 0.20$, so forecast 1's error is $(F_{1t} - \sigma_t) = 0.15$, while forecast 2's error is only 0.05, or one third as large. At the upper end of its range, $\sigma_t = 0.25$, and we will have $F_{1t} = 0.50$ and $F_{2t} \cong 0.20$. Forecast 1's error is now 0.25, which is five times larger than forecast 2's error of -0.05.

It is easy to see that at every possible value for the true volatility, the error produced by the Agood@ forecast number 1 is much larger than that produced by the Abad@ constant forecast of 0.20. The problem, of course, is that F_1 is biased, and even though it contains much more information about the true volatility than F_2 does, it is not necessarily more accurate unless the bias is corrected. Simply putting F_{1t} into an option pricing formula as the volatility parameter will result in highly inaccurate answers.

This artificial example illustrates what can happen hypothetically, but is this a real problem in practice? A close look at the results presented by Lamoureux and Lastrapes [1993] (LL) shows that it can be. LL examine the information content and out-of-sample forecasting performance of three different types of volatility estimators for 10 individual stocks. Using two years of bid and ask quotes from a transactions data base, they attempt to minimize the effects of bid-ask bounce, nonsynchronous prices, and other noise in the data. Implied volatilities are calculated from at-the-money medium maturity (approximately 1 to 4 months) call options, with each day=s IV taken to be the average of the values computed from all price quotes during the day (about 50 quotes, on average). A GARCH (1,1) model is also fitted for each stock, and refitted each day so that only past data is used in each estimation. The third volatility estimator examined is the historical sample standard deviation calculated over all prior days in the sample (a minimum of 300 observations).

LL analyze the three estimators both in-sample and out-of-sample. The in-sample results suggest that both GARCH forecasts and IVs may contain valuable information, but of course forecasting performance can only be measured correctly in an out-of-sample trial. Unfortunately, LL do not report results from the rationality test regression on each volatility predictor separately. However, they do run encompassing regression tests on the 10 stocks in their sample, and they also compare the forecasts in terms of RMSE.

Table III.5 summarizes some of the results of these tests. First, LL run encompassing regressions with all three predictors. For 9 out of 10 stocks, the coefficient on daily average implied volatility from all at-the-money call option quotes is positive, and significant at the 95 percent confidence level for 7 of them. The GARCH predictions are constructed in the appropriate manner by computing a multi-step ahead conditional GARCH forecast for each date from the present through option maturity and cumulating them into a forecasted average variance to expiration. The GARCH forecast coefficients are positive in 7 cases out of 10, but only statistically significant in one of them. Lastly, the simple historical sample volatility receives a negative coefficient estimate for every stock, and 8 of them are significant at the 95 percent confidence level. These results are strong evidence that given the IV and GARCH forecasts together, there is no further relevant information about future volatility to be extracted from the historical volatility. LL therefore eliminate the historical estimator and rerun the encompassing regressions with just IV and GARCH forecasts. In the 2-variable encompassing regressions, the coefficient on IV is positive and statistically significant for 9 of 10 stocks, while the GARCH forecasts now have negative coefficients in 9 cases, of which 8 are statistically significant. Most researchers would conclude from these results that the average IV is the best forecast of future volatility for these underlying stocks, or at least that it is definitely superior to both GARCH(1,1) and the simple historical volatility.

However, this seemingly obvious conclusion is wrong. It does not distinguish between the information content of implied volatility, which is what the encompassing regression addresses, and forecast accuracy, which depends on how close the forecast is to the realization. The critical distinction, seen in our simple example above, is that a biased estimate may contain a great deal of information about the target variable, but it only becomes an accurate forecast when the bias is corrected.

The second part of Table III.5 compares the same forecasts in terms of their root mean squared error in predicting future realized volatility for these stocks. By this standard, IV now appears to be the worst predictor, showing the lowest RMSE in only two cases and the highest in 8. The GARCH model is best 3 times and worst twice, but the simple historical volatility provides the most accurate predictions in 5 cases out of 10 and is never the worst.

Thus, the success of IV in an encompassing regression analysis does not mean that one will obtain the best estimates of option fair values by simply inserting it as the volatility parameter into the Black-Scholes valuation equation. The bias must be corrected first. Worse still, although the results from the rationality test regression indicate how to do this, due to nonstationarity such attempts at bias correction need not be successful.

Suppose historical data are available on IVs and realized volatilities, and running the equation (III.6) regression has produced coefficient estimates *a* and *b* for the constant and slope, respectively. One way of describing what this regression calculation does is that it finds the linear function F = $a + b\,\sigma_{IV}$ that minimizes the squared deviation from $\sigma_{REALIZED}$. That is, in the data sample, F is the minimum root mean squared error estimate of $\sigma_{REALIZED}$ given σ_{IV}. By construction, F will pass the rationality test for this sample.

Therefore, if the characteristics of the bias in implied volatility are stable over time, one can fit *a* and *b* values on past data and use them to correct the bias in the current IV. The difficulty is that like all of the other elements of this problem, these parameters may vary over time. That is especially likely in this case, as market participants attempt to correct any biases in expectations formation that they become aware of.

Day and Lewis [1993] examine such a bias correction scheme in their analysis of crude oil futures options. They fit regression (III.6) on the first portion of their data sample and then use the fitted coefficients to Acorrect@ the biases for their last 150 observations. Table III.6 summarizes some of their results. Only for the GARCH model forecasts did this attempt at bias correction improve prediction accuracy, but this result was of little consequence, since the corrected GARCH forecasts were still less accurate than the uncorrected forecasts from the other models, except for EGARCH at the near horizon.

Thus the statistical evidence from many studies on different markets indicates that the commonly used volatility estimators are normally biased, but Day and Lewis=s results show that attempting to correct the bias may not work, since the bias itself varies over time. While this is rather discouraging in terms of our desire to obtain accurate volatility predictions, from another perspective it can be somewhat comforting, as we will explain below.

III.6. Concluding Comments

This chapter has offered a critical look at the widely held belief that implied volatility computed from market options prices is an informationally efficient forecast of the volatility that will actually be experienced by the underlying asset from the present through expiration date. We began by noting that this was actually two separate hypotheses. The first holds that the market price for an option will be its model value based on the market=s expectation of future volatility, so that the implied volatility reveals the market=s true volatility forecast. The second is that investors as a group are rational in evaluating the information available to them, so that the market=s volatility forecast is the correct conditional expected value of the future volatility, given the available information.

The second hypothesis relates to the rationality of expectations formation. We are loath to question investor rationality, both because it is one of the fundamental principles of all of financial theory, and also because of the large body of empirical evidence that supports market efficiency in other markets. But the first hypothesis relates to the performance of the trading mechanism in the options markets. In theory, the reason the implied volatility computed from an option=s market price should reveal the market=s forecast of future volatility is that arbitrageurs stand ready to take positions aggressively based on their (rational) volatility beliefs. Their trading will stabilize the market so that imbalances in supply and demand do not push option prices away from their model values. Otherwise, if the market price for the option is not its model value, the implied volatility will not be the market=s volatility forecast.

We then described some of the difficulties in obtaining a clean test of these hypotheses given the nature of the data generating process, particularly the noise introduced by bid-ask spreads and nonsynchronous trading in options and their underlying assets. Volatility only affects the time value of an option, so for deep-in-the-money and deep-out-of-the-money contracts, the price impact of a small bid-ask spread in either the option or the underlying can translate into a large effect on implied volatility. This is especially a problem for the in-the-money options, whose relatively high prices and illiquidity cause spreads to be wide. It is easy, for example, for the option bid price in the market to be below the arbitrage-based lower bound on the option price, which would correspond to an implied volatility less than 0.

The widely used regression-based tests for forecast rationality and for relative information content in competing forecasts were then discussed. A close look at empirical results from running these regressions on implied volatilities from a large sample of OEX call options, published in Canina and Figlewski [1993], revealed that at least for that market and time period, implied volatility from the options prices appeared to contain no information at all about the future volatility of the underlying stock index. We argued that these highly negative results should be interpreted not as evidence that OEX option traders are irrational, but rather, that the aggressive arbitrage trading needed to hold option market prices close to their theoretical values is especially hard, costly, and risky to do in this market. At the same time, we described how the volatility-based options arbitrage, that plays such an important role in the theoretical market environment in which the valuation model holds, may be less appealing than other trading strategies in the real world. Actual options market makers may well find such arbitrage trading based on their volatility expectations to be much less profitable than simply setting bids and offers around the current market prices, even when those prices are very different from the theoretical model values.

This interpretation of the CF results led to the hypothesis that the performance of implied volatility in the rationality test regression should vary across markets according to how difficult the arbitrage trade is to execute. Stock index options represent a polar case, where the arbitrage trade is complicated to execute at the outset, and the resulting position is both costly and risky to hedge over time. At the other extreme would be futures options, where the option and the underlying are traded side by side on the same trading floor, with low transactions costs (along with other practical advantages not present in other options markets, such as more favorable margin treatment for hedged positions). A selection of results from the finance literature covering a variety of markets provided some support for this hypothesis. However, only for a single maturity in a single market--nearby at-the-money options on crude oil futures--did the statistical analysis not reject the hypothesis that implied volatility was a fully rational forecast.

Lastly, we addressed the issue of how to obtain the necessary volatility input for an option valuation model when the available choices all seem to be irrational. In particular, we argued strongly against the common practice of running the rationality test regression on implied volatilities and alternative forecasts from models based on historical price data, finding that IV has the highest R^2 and receives the largest weight in an encompassing regression, and concluding that IV is therefore the best volatility input for the valuation formula. If the bias in the volatility forecast is not corrected, information content, as indicated by a high R^2 in these regression tests, does not translate directly into forecast accuracy and correct model option values. The results from Lamoureux and Lastrapes [1993] illustrated that point clearly. Unfortunately, attempting to correct for bias based on the results of the rationality regression need not be successful, since the bias can vary over time. This is an area where more research is needed.

The rationality test regression shows that implied volatility is biased, as are the volatility forecasts from other prediction methods. This is somewhat disturbing to our belief in efficient markets. Even though we can explain this result without abandoning the assumption that investors evaluate information rationally, we also can not rule out irrationality without actually observing and testing their expectations.

However, if the evidence had shown the bias in implied volatility to be persistent and easily correctable, it would have been much stronger evidence against investor rationality than what we actually observe. Bias that is time-varying and hard to adjust for is consistent with a financial market in which investors act rationally, but can only learn about new market informational factors gradually over time. Once evidence of a systematic expectations error accumulates, they will try to correct for it in making predictions, so that last period=s mistakes are not perpetuated. But investors can not adjust immediately to each new factor as it arises. This will lead to a continual series of short run biases, but things that persist long enough will become understood and incorporated into future expectations.

Unfortunately, the general failure of implied volatility to pass the rationality test, and the difficulty in correcting IV to turn it into a rational forecast, means the question of how to obtain the best prediction of future volatility from observed implied volatilities remains open.

Table III.1 Effect of Option Price Discreteness and Bid-Ask Spreads on the Calculation of Implied Volatility

European call options on a non-dividend paying stock are valued using Black-Scholes. Prices are rounded to the nearest tick (1/16 for prices under $1, 1/8 for prices over $1). Implied volatilities are computed for Amarket@ option prices incorporating a bid-ask spread around the Arounded price.@ Assumptions: Stock price = 50; Strike prices = 45, 50, 55; Maturities: 1 and 6 months; Interest rate = 8.0 percent; True volatility = 0.250. *** indicates implied volatility could not be computed because the market option price is too low even at a volatility of 0.0.

1 Month Options

Strike	Black-Scholes Value	Rounded to Nearest Tick	Rounded Price IV	Bid-Ask Spread	Bid Price	Bid IV	Ask Price	Ask IV
45	5.386	5 3/8	0.243	0	5 3/8	0.243	5 3/8	0.243
50	1.617	1 5/8	0.251	0	1 5/8	0.251	1 5/8	0.251
55	0.206	3/16	0.244	0	3/16	0.244	3/16	0.244
45	5.386	5 3/8	0.243	1/4	5 1/4	***	5 1/2	0.321
50	1.617	1 5/8	0.251	1/8	1 9/16	0.241	1 11/16	0.262
55	0.206	3/16	0.244	1/16	3/16	0.232	1/4	0.255
45	5.386	5 3/8	0.243	1/2	5 1/8	***	5 5/8	0.398
50	1.617	1 5/8	0.251	1/4	1 1/2	0.230	1 3/4	0.273
55	0.206	3/16	0.244	1/8	1/8	0.221	1/4	0.265
45	5.386	5 3/8	0.243	3/4	5	***	5 3/4	0.475
50	1.617	1 5/8	0.251	3/8	1 7/16	0.219	1 13/16	0.284
55	0.206	3/16	0.244	3/16	1/8	0.210	5/16	0.276

Table III.1 continued

6 Month Options

Strike	Black-Scholes Value	Rounded to Nearest Tick	Rounded Price IV	Bid-Ask Spread	Bid Price	Bid IV	Ask Price	Ask IV
45	7.668	7 5/8	0.245	0	7 5/8	0.245	7 5/8	0.245
50	4.488	4 1/2	0.251	0	4 1/2	0.251	4 1/2	0.251
55	2.354	2 3/8	0.251	0	2 3/8	0.251	2 3/8	0.251
45	7.668	7 5/8	0.245	1/4	7 1/2	0.232	7 3/4	0.259
50	4.488	4 1/2	0.251	1/4	4 3/8	0.242	4 5/8	0.260
55	2.354	2 3/8	0.251	1/8	2 5/16	0.247	2 7/16	0.256
45	7.668	7 5/8	0.245	1/2	7 3/8	0.219	7 7/8	0.272
50	4.488	4 1/2	0.251	1/2	4 1/4	0.232	4 3/4	0.269
55	2.354	2 3/8	0.251	1/4	2 1/4	0.242	2 1/2	0.261
45	7.668	7 5/8	0.245	1	7 1/8	0.192	8 1/8	0.299
50	4.488	4 1/2	0.251	3/4	4 1/8	0.223	4 7/8	0.279
55	2.354	2 3/8	0.251	1/2	2 1/8	0.233	2 5/8	0.270

Table III.2 Effect of Underlying Asset Bid-Ask Spread on Calculation of Implied Volatility

European call options on a non-dividend paying stock are valued using Black-Scholes. Option prices are rounded to the nearest tick (1/16 for prices under $1, 1/8 for prices over $1). Implied volatilities are computed for underlying asset prices incorporating a bid-ask spread. No bid-ask spread is assumed for options. Assumptions: Stock price (center of spread) = 50; Strike prices = 45, 50, 55; Maturities: 1 and 6 months; Interest rate = 8.0 percent; True volatility = 0.250. *** indicates implied volatility could not be computed because the market option price is too low even at a volatility of 0.0.

Stock Bid-Ask Spread	Option Strike	Option Price (Rounded)	Stock Bid Price	Option IV at Bid	Stock Ask Price	Option IV at Ask
			1 Month Options			
1/4	45	5 3/8	49 7/8	0.310	50 1/8	***
	50	1 5/8	49 7/8	0.263	50 1/8	0.239
	55	3/16	49 7/8	0.249	50 1/8	0.239
1/2	45	5 3/8	49 3/4	0.370	50 1/4	***
	50	1 5/8	49 3/4	0.275	50 1/4	0.227
	55	3/16	49 3/4	0.254	50 1/4	0.234
1	45	5 3/8	49 1/2	0.471	50 1/2	***
	50	1 5/8	49 1/2	0.297	50 1/2	0.201
	55	3/16	49 1/2	0.264	50 1/2	0.225
			6 Month Options			
1/4	45	7 5/8	49 7/8	0.256	50 1/8	0.234
	50	4 1/2	49 7/8	0.257	50 1/8	0.245
	55	2 3/8	49 7/8	0.255	50 1/8	0.248
1/2	45	7 5/8	49 3/4	0.267	50 1/4	0.223
	50	4 1/2	49 3/4	0.262	50 1/4	0.239
	55	2 3/8	49 3/4	0.259	50 1/4	0.244
1	45	7 5/8	49 1/2	0.287	50 1/2	0.200
	50	4 1/2	49 1/2	0.273	50 1/2	0.227
	55	2 3/8	49 1/2	0.266	50 1/2	0.237

TABLE III.3. Regression Tests for Forecast Rationality of Implied Volatility and Historical Volatility for S&P 100 Index Options

Source: Canina and Figlewski [1993]

Results are shown for regression Equation (III.6), the equivalent univariate regression using σ_{HIST} (historical volatility over the previous 60 calendar days), and the Aencompassing regression@ Equation (III.7) on S&P 100 stock index call options, for the period March 15, 1983 - March 28, 1987. The study uses reported closing prices for all options from 20 index points out-of-the-money to 20 points in-the-money (relative to an index level above 200 for most of the sample period) having time to maturity from 7 to 127 calendar days. Implied volatilities are computed using a binomial model with 500 time steps, taking into account the value of early exercise of these American options. Further details of the methodology are given in the text.

Consistent standard errors are shown in parentheses, adjusting for cross-correlations in the residuals due to overlapping maturities, as described in the text.

<u>Results using full data sample (NOBS = 17,606):</u>

$$\sigma_{REALIZED}(t) = \underset{(0.012)}{0.136} + \underset{(0.050)}{0.022}\,\sigma_{IV}(t) + \varepsilon(t) \qquad R^2 = 0.002$$

<u>Results from at-the-money, near term option subsample (NOBS = 852)</u>
[($-5.00 \le S - X \le -0.01$) and ($29 \le T\text{-}t \le 63$ days)]

$$\sigma_{REALIZED}(t) = \underset{(0.017)}{0.113} + \underset{(0.101)}{0.163}\,\sigma_{IV}(t) + \varepsilon(t) \qquad R^2 = 0.053$$

$$\sigma_{REALIZED}(t) = \underset{(0.024)}{0.074} + \underset{(0.165)}{0.464}\,\sigma_{HIST}(t) + \varepsilon(t) \qquad R^2 = 0.151$$

$$\sigma_{REALIZED}(t) = \underset{(0.024)}{0.074} - \underset{(0.097)}{0.008}\,\sigma_{IV}(t) + \underset{(0.203)}{0.473}\,\sigma_{HIST}(t) + \varepsilon(t) \qquad R^2 = 0.149$$

TABLE III.4. The Rationality Test Regression: Evidence from the Literature

PART A: THE ARTICLES

Paper: Canina and Figlewski, Review of Financial Studies, 1993
Underlying: S&P 100 Stock Index (OEX)
Data set: 17,000+ observations, Daily, 3/84 - 3/87
Variables: **Implied volatility** - Call option closing prices, 8 strikes, 4 maturities each day, not averaged
Historical volatility - previous 60 calendar days

Paper: Fleming, Rice University, Working Paper 1996.
Underlying: S&P 100 Stock Index (OEX)
Data set: Transactions data, 1664 days, 10/85 - 4/92
Variables: **Implied volatility** - At-the-money calls and puts, nearest contract with >15 days to expiration, average over all transactions during 10 minutes around the market close
Historical volatility - previous 28 calendar days
Special feature: Regression is run in 1st difference form.

Paper: Beckers, Journal of Banking and Finance, 1981
Underlying: Individual stocks
Data set: 10 3-month subperiods from 4/75 to 7/77, 1 observation per stock per period, 62-116 stocks in each period
Variables: **Implied volatility** - Closing prices, at-the-money call with approximately 3 months to maturity, average IV over 5 days
Historical volatility - from previous quarter

Paper: Day and Lewis, Journal of Derivatives, 1993
Underlying: Crude Oil Futures
Data set: Daily closing prices 11/86 - 3/91,
Variables: **Implied volatility** - Closing prices, at-the-money calls, 2nd and 4th month (on average 32 and 72 trading days, respectively)
Historical volatility - number of days set equal to option maturity
GARCH-M(1,1) - rolling estimate on 500 past days
EGARCH-AR(1)(1,1,1) - rolling estimate on 500 past days

TABLE III.4 continued

Paper: Jorion, Journal of Finance, 1995
Underlying: Foreign Currency Futures, (Deutschemark, Japanese yen, Swiss franc)
Data set: Daily closing prices, starting1/85 (DM), 7/86 (JY), 3/85 (SF); ending 2/92
Variables: **Implied volatility** - Average of call and put IVs from at-the-money nearest maturity contracts (3 to100 calendar days)
Historical volatility - average of squared returns over previous 20 trading days
GARCH(1,1) - 1-day ahead forecast, GARCH parameters estimated over full sample

Paper: Ferri, Unpublished Ph.D. dissertation research,1996
Underlying: Deutschemark exchange rate
Data set: All option transactions, Jan. 3, 1984 - Feb. 23, 1995, approximately 150,000 observations
Variables: **Implied volatility** - IVs from 2nd maturity (1-2 months) calls and puts from 10 percent out-of-the-money to 10 percent in-the-money
Historical volatility - past 60 days

Paper: Lamoureux and Lastrapes, Review of Financial Studies, 1993
Underlying: 10 Individual non-dividend paying stocks
Data set: Bid/ask quotes from transactions data base, 4/82 - 3/84
Variables: **Implied volatility** - At-the-money 2nd maturity call, average IV from all Bid/Ask midpoint quotes during the day
Historical volatility - from the beginning of the sample
GARCH(1,1) - estimated over all past days, 300 minimum

TABLE III.4 continued

PART B: THE RATIONALITY TEST REGRESSION RESULTS

	Constant (t_0)	Slope (t_1)	Variable	R^2
Canina and Figlewski				
Full sample	0.136 (11.3)	0.022 (19.6)	σ_{IV}	0.002
Average of 32 subsamples	0.075 (2.6)	0.461 (2.7)	σ_{HIST}	0.142
Fleming				
All regressions GMM estimation on 1st differences	-.017 (3.2)	0.567 (9.3)	$\sigma_{CALL\ IV}$	0.026
	-.023 (3.2)	0.640 (9.3)	$\sigma_{PUT\ IV}$	0.024
	.001 (0.1)	0.577 (3.8)	σ_{HIST}	0.012
Beckers				
Averages over 10 subperiods	.003 (2.48)	0.813 (10.5)	σ_{IV}	0.533
	.004 (3.8)	0.673 (9.7)	σ_{HIST}	0.495

TABLE III.4 continued

	Constant (t_0)	Slope (t_1)	Variable	R^2
<u>Day and Lewis</u>				
2nd month (average maturity 32 days)	.003	0.880	σ_{IV}	0.718
	(1.0)	(0.8)		
	.008	0.607	σ_{HIST}	0.392
	(2.0)	(1.8)		
	.007	1.160	σ_{GARCH}	0.371
	(1.2)	(0.3)		
	.018	0.183	σ_{EGARCH}	0.022
	(3.6)	(4.1)		
<u>Jorion</u>				
Average values over 3 currencies	.347	0.521	σ_{IV}	0.133
	(2.7)	(2.9)		
	.606	0.169	σ_{HIST}	0.044
	(7.4)	(8.3)		
	.184	0.715	σ_{GARCH}	0.053
	(0.9)	(1.1)		
<u>Ferri</u>				
	.043	0.544	$\sigma_{CALL\ IV}$	0.161
	(3.6)	(5.3)		
	.035	0.582	$\sigma_{PUT\ IV}$	0.212
	(2.7)	(4.0)		
	.074	0.338	σ_{HIST}	0.090
	(6.1)	(6.4)		

Note: t_0 is the t-statistic on the hypothesis that the constant term is 0.0; t_1 is the t-statistic on the hypothesis that the slope coefficient is 1.0.

TABLE III.5. Information Content versus Forecasting Accuracy for Volatility Predictors for 10 Stocks

Source: Lamoureux and Lastrapes [1993]
See Table III.4 and the text for information about the data set.

Encompassing Regression Coefficient Estimates:

3-variable: $\sigma_{REALIZED} = \alpha + \beta_1 \sigma_{IV} + \beta_2 \sigma_{GARCH} + \beta_3 \sigma_{HIST}$

σ_{IV}: 9 β_1 estimates positive, 7 significant at 95% confidence

σ_{GARCH}: 7 β_2 estimates positive, 1 significant at 95% confidence

σ_{HIST}: 10 β_3 estimates negative, 8 significant at 95% confidence

2-variable: $\sigma_{REALIZED} = \alpha + \beta_1 \sigma_{IV} + \beta_2 \sigma_{GARCH}$

σ_{IV}: 9 β_1 estimates positive, 9 significant at 95% confidence

σ_{GARCH}: 9 β_2 estimates negative, 8 significant at 95% confidence

Comparison of Root Mean Squared Forecast Errors

	Lowest RMSE	Highest RMSE
σ_{IV}:	2	8
σ_{GARCH}:	3	2
σ_{HIST}:	5	0

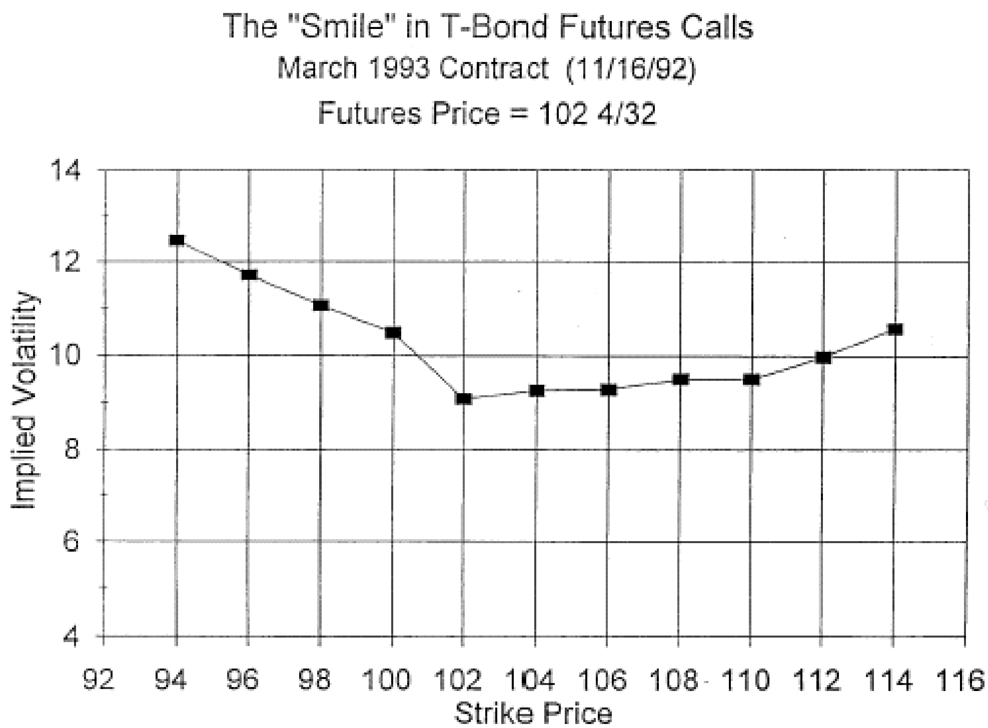
TABLE III.6. Performance of Bias-Corrected Forecasts of Crude Oil Futures Volatility

Source: Day and Lewis [1993]

The table reports out-of-sample mean error and root mean squared error for different volatility predictors, where forecast error $\varepsilon = (\sigma_{REALIZED} - \sigma_{FORECAST})$. Bias-corrected forecasts are constructed from the raw volatility predictions by $\sigma_{CORRECTED} = a + b\sigma_{RAW}$, where a and b are the fitted constant and slope coefficients from the regression equation (III.6). The coefficients are estimated from the first 845(Near term) or 710 (Distant term) sample data points, leaving 150 observations for the out-of-sample analysis. Near (Distant) term contracts average 32 (68) days to expiration. Figures are in annualized percents. See Table III.4 for further information on the data set.

	Unadjusted Forecasts		Bias-Corrected Forecasts	
Forecast Method	Mean Error	RMSE	Mean Error	RMSE
	Near Term Contracts			
Implied volatility	0.006	0.128	0.092	0.145
Historical	-0.024	0.243	0.207	0.265
GARCH	0.215	.0284	0.196	0.257
EGARCH	0.256	.0305	0.287	0.323
	Distant Term Contracts			
Implied volatility	0.099	0.179	0.190	0.219
Historical	0.049	0.230	0.221	0.257
GARCH	0.271	0.305	0.217	0.252
EGARCH	0.168	0.218	0.240	0.268

FIGURE III.1



The "Smile" in T-Bond Futures Calls
March 1993 Contract (11/16/92)
Futures Price = 102 4/32
Implied Volatility
14
12
10
8
6
4
92 94 96 98 100 102 104 106 108 110 112 114 116
Strike Price

Chapter IV:
SUMMARY AND CONCLUSIONS

In this final chapter we will summarize the major points covered in this investigation and indicate a variety of implications and conclusions that we believe are warranted.

IV.1 On the meaning of volatility

AVolatility,@ as the term is used in practice, does not refer to a single parameter, but to a set of related concepts. In an option pricing model, volatility is the square root of the average variance of return on the underlying asset over every instant of the option=s remaining lifetime. An options market maker, on the other hand, may use the term to mean the variability of return over the immediate short run. But much of the time, when a market maker uses the term volatility he actually means implied volatility. By contrast, a risk manager for a financial institution may calculate volatility in order to estimate the probability distribution for the value of a borrower=s collateral at the maturity of a loan, and from that, her firm=s exposure to the risk of default. Only under restrictive assumptions (notably those commonly adopted in contingent claims valuation models) can these related, but different, risk measures be summarized with a single number.

A clear implication of this observation is:

* It is important to specify exactly what risk one needs to be concerned with, and to choose an estimation framework that is suited to it. This will not necessarily be the same for all applications, even if what is needed in each case is the standard deviation of some asset=s return.

IV.2. On the philosophy of modeling the behavior of a financial market

A financial market is an institution set up by human beings, and the security prices established in it are artifacts of human activity. It is very different from a physical system, such as that governing the movements of a celestial body, even though we may use some of the same tools to analyze the data a market produces. Classical statistics is based on a conceptual framework in which there is a fixed underlying structure, whose characteristics are revealed gradually as its behavior is observed over time. This is appropriate for a physical system, since physical laws, like the equation governing gravitational attraction, are unchanging and exact (at least until one gets to the sub-atomic level) and they tend to have relatively simple functional forms that can be uncovered by statistical analysis of noisy observational data.

By contrast, the behavior of a financial market entails nothing like the potential for predictability of a physical system. As a man-made institution, its Aunderlying structure@ is no more fixed and immutable than human behavior is. Unlike physical laws, economic relationships are inherently noisy since they represent an attempt to apply mathematical descriptions to explain how people, rather than inanimate objects, will act. Financial markets can be expected to change continuously as economic and social conditions evolve, sometimes sharply and without warning. Forecasting market behavior is possible only to the extent that change is relatively gradual most of the time, so extrapolation from the recent past can yield some information about the near future.

Recognizing the inherent difference between a physical system and a financial market leads to several implications about the strategy for estimating future volatility.

* In contrast to the models of classical statistics, out-of-sample forecasting of the behavior of a financial market is much different from in-sample estimation, because the underlying structure is changing. In-sample goodness-of-fit statistics are not a dependable gauge of how successful a model will be in forecasting.

* Limiting one=s expectations about how much predictive power can be obtained from a financial model is both a safer posture than being overly ambitious, and also more likely to yield a realistic assessment of probable performance.

* Overfitting is a great danger in financial modeling. Simple but robust models are likely to do better than complex ones that depend heavily on specific details of model structure, such as an assumption about the particular functional form of a probability distribution.

* All models should be expected to go off track over time. Frequent refitting, and eliminating obsolete data from the estimation sample are appropriate. At times a formerly successful model may need to be discarded entirely.

IV.3. On volatility forecasting with historical data

We were particularly interested in examining how best to obtain volatility forecasts from historical data for use in pricing longer maturity options. The standard approach is to treat volatility as a constant parameter, selecting a sample of recent prices and applying the estimation technique from classical statistics that would be appropriate under the assumptions of the Black-Scholes option pricing model. This may be modified slightly in recognition of known differences between actual price behavior and that of a pure logarithmic diffusion, for example, not attempting to make use of intraday data because of noise from market microstructure effects. Some of the difficulties that arise with this approach include the following: Volatility is not constant, it evolves and may exhibit occasional discontinuous price jumps that produce very high measured volatility at particular points in time; Volatility is meant to measure the variability of the market-clearing equilibrium price, but price data from transactions will exhibit spurious variation due to bid-ask spreads and infrequent trading of less liquid instruments; Estimation can be affected by serial correlation in returns and mean reversion; The standard technique is based on deviations of returns from the sample mean, which will be a highly inaccurate estimate of the true mean except in very long data samples.

The conceptual inconsistency of using a fixed volatility framework to forecast a parameter that changes over time can be dealt with by adopting a formal model of time-varying volatility. The most common are those from the ARCH family. These models have several drawbacks for long-term volatility prediction, however, including the fact that they require a large data sample and a relatively complicated estimation procedure, and they are not designed for multi-step ahead forecasting.

These are the major conclusions we draw from our investigation of the forecasting performance of historical volatility estimates.

* Market Amicrostructure@ effects, like those arising from the use of transactions data subject to a bid-ask spread, are important and interfere with accurate estimation. Intraday data in most cases will contain too much noise from this source to be useful for longer horizon forecasting. Serial correlation over periods of a day or longer may also be present in some markets, and will cause estimation problems. The easiest way to deal with serial correlation is to limit the observation frequency in each case to be no higher than the highest frequency at which returns are serially independent. In theory, an analyst could use high frequency data and correct statistically for these effects, but this would require information about the exact properties of trading noise and nonindependence to be accurately estimable from past data. We did not explore this line of investigation.

* Since volatility changes over time, if one uses a fixed-volatility estimation approach, it may be appropriate to limit the amount of past data used. The tradeoff will be between losing accuracy because throwing away old data makes the sample size smaller than it might be, and losing accuracy by including old data that may contain relatively little information about the current state of the system. One way to downweight old data without completely discarding it is to use a weighted average of historical observations with weights that decline with the age of each data point. We did not examine that alternative here.

* Surprisingly, with monthly data the greatest accuracy in terms of RMSE for both long (5 years) and short (6 months) horizons, was generally achieved by using long historical price series in constructing volatility estimates. Five years of monthly data gave better results than shorter samples in most cases. We found the same results for longer horizon forecasting using daily data, but for shorter horizons (6 months or less), the most accurate forecasts were obtained from historical samples that were substantially longer than the forecast horizons, but not as long as 5 years.

* Also surprisingly, the accuracy of the forecast of average volatility increased for longer forecasting horizons, so that there were smaller errors in predicting average volatility over the next five years than over the next 6 months.

* The sample mean return is a highly inaccurate estimate of the true mean except in very long data samples, while economic theory allows us to place relatively tight restrictions on plausible values for expected asset returns. This means that in most cases it will reduce RMSE to compute volatility from deviations around a Amean@ imposed by the analyst instead of the sample mean. Since the problem can be largely eliminated just by avoiding using extreme sample mean estimates in the calculation, imposing a Amean@ of zero and computing variance as the average of the squared

returns in the sample is normally an adequate correction for this problem.

* ARCH family models require large data samples and are difficult to adapt to a multi-step forecasting application. Using monthly data, they were found to be hard to estimate and not useful for volatility prediction over the relative long horizons we examined. With daily observations GARCH(1,1) models fitted to 5 years of data worked well for the S&P 500 index volatility, even for 24 month forecasts, but GARCH was successful in forecasting volatilities of the other data series only at the shortest horizon of 1 month. We did not actively explore the many possibilities for enhancing performance with this class of volatility models.

IV.4. On implied volatility as a forecast of future volatility

Because all of the parameters that enter the Black-Scholes model and similar valuation equations are observable except for volatility, one can solve for the volatility that would make the market price equal to the model value. Both academics and practitioners regard implied volatility as important information.

Many academics consider IV to be the best forecast of future volatility, because it properly accounts for all publicly available information, including everything that can be gleaned from historical price data. This is because they think of IV as a direct measurement of Athe market=s@ expectation of future volatility, and the market is informationally efficient. By contrast, traders use an option=s implied volatility as a gauge of how the market is currently pricing it relative to the underlying asset, without worrying too much about whether IV is an accurate forecast of how volatile the underlying will be over the option=s lifetime. The two groups are largely unaware of how differently they think about what information IV contains.

The calculation of IV is seriously affected by a variety of data problems, including the effects of bid-ask spreads in both the option and the underlying, nonsynchronous prices, and transactions costs and other problems that prevent option mispricing induced by imbalances in supply and demand from being arbitraged away. Some data problems can be corrected, or at least partially mitigated by averaging IVs from different options or from multiple transactions in the same option. However, we argue that it is inappropriate to suppress a regular volatility structure, like the Asmile,@ by such averaging. Rather, the existence of a smile pattern implies that the model used to calculate the IVs is not a correct description of how the market is pricing options.

Forecast Arationality@ of IV can be tested by regressing realized volatility on implied volatility. For any informationally efficient forecast, the regression constant should be zero and the slope coefficient should be one. The relative information content of two different forecasts, such as IV and historical volatility, can be examined by putting both into an encompassing regression and comparing their coefficients. Such tests have been run for a large number of options markets, using a wide variety of data selection and cleaning techniques. The surprising result from an extensive study of S&P 100 stock index options conducted by Canina and Figlewski [1993] was that not only was IV not a fully rational forecast of future volatility in that market, it appeared to contain no information about it at all. By contrast, historical volatility, while not a rational forecast either, clearly contained more information than IV.

This very negative result led us to reconsider the mechanism by which investors= volatility expectations are incorporated into option prices, through the trading of arbitrageurs who attempt to exploit option mispricing in the market. If the arbitrage trade is hard to execute, or risky, or entails large transactions costs, this mechanism will be weak. Relatively large pricing errors may be allowed to persist, and implied volatilities computed from market prices can be very different from investors= true expectations. This suggested a hypothesis, that implied volatilities from different options markets will contain relatively more or less information depending on whether the arbitrage trade in that market is easy or hard. We presented evidence from published and unpublished articles reporting rationality test regression results for a number of different options markets and found them to be broadly consistent with the hypothesis. For the most part, the studies showed that implied volatility contained a statistically significant amount of information about future volatility, and generally more than the historical volatility measures that were examined. However, in almost no case did IV appear to be a fully rational forecast.

Lastly, in considering how the rationality test results should affect the use of IV in predicting future volatility, we argued strongly against the common practice of comparing IV against alternative forecasts in an encompassing regression, finding that it receives the highest weight, and then adopting it as one=s volatility input even though it does not pass the rationality test. Both in a simple example and in the context of Lamoureux and Lastrapes=s [1993] investigation of stock option pricing, it was shown that greater information content in the form of better performance in an encompassing regression does not necessarily produce more accurate forecasts unless the bias is first corrected.

Our conclusions regarding the information content of implied volatility and its value as a forecast of future volatility are the following.

* Data problems can be very important in obtaining accurate calculations of implied volatility. Use of intraday transactions data can be valuable to guarantee synchronous prices and allow averaging across multiple observations to reduce the effect of noise from the bid-ask spread. Care should be taken in selecting an interest rate (and dividend forecast, where relevant), especially for longer maturity options. We argue that although the Treasury bill rate is clearly a Ariskless interest rate@ in the U.S., it is actually too low to be a good proxy for the relevant rate facing an options arbitrageur, particularly as a measure of borrowing costs.

* Averaging IVs across options with different strike prices that has the effect of suppressing a smile pattern is inappropriate. A persistent volatility smile means the model being used to obtain IVs is not the model the market is using in pricing them. (That need not mean the IV it produces contains no useful information, however.)

* The effect of noise from the bid-ask spread is greatest for a deep-in-the-money option because the volatility-related time value is only a very small portion of the total price. In some cases, an observed volatility smile pattern is partly spurious, due to the fact that even a small underpricing for these options can violate the lower bound on the option price and make calculation of IV impossible: An ordinary bid-ask spread can produce very high IVs from option trades at the ask price but the offsetting low IVs from trades at the bid are eliminated from the sample because they violate the lower bound.

* IV generally does not pass the test of forecast rationality even though it may contain significant information about future volatility. For IV to be the market=s fully efficient volatility forecast, first

it must be an accurate measure of the market=s expectation of future volatility for the underlying asset, and second, investors must be rational in forming their expectations from the available information. We are inclined to seek an explanation for the negative rationality test results in a failure of the first condition rather than the second: Investors may well be rational in analyzing available data, but the trading mechanism by which their expectations become embedded in option prices is weak. Market frictions impede arbitrage trading, while market makers may well find that profitability is higher if they maximize turnover by trading at market clearing prices even when, given expected future volatility, they think these prices are incorrect. Nevertheless, while we prefer to think that investors are not irrational in using information, we have offered no hard evidence that that is, in fact, the case.

* Evidence from a variety of studies is consistent with the hypothesis that information content of IV will be positively related to the ease of performing the arbitrage trade between options and their underlying assets. However, the evidence is far from overpowering. This is an area in which further investigation would be worthwhile.

* The general result that IV has a statistically significant coefficient in the rationality test regression and frequently dominates historical volatility in an encompassing regression means that it contains useful information. It does not mean that IV is necessarily a more accurate forecast of future volatility or that it is a better volatility parameter to use as an input to a pricing model. That would only be true if it were also unbiased.

* The fitted coefficients from the rationality regression indicate how a raw implied volatility figure would need to be adjusted to correct its bias, or, more precisely, how it would have needed to be adjusted to be unbiased <u>during the sample period</u>. Attempting to correct future IVs before putting them into an option pricing model using the fitted coefficients will only be successful if the bias remains relatively constant over time. One published study found that this approach did not, in fact, lead to greater accuracy in practice. Even so, it can not be an entirely rational procedure to attempt to value options using a volatility forecast that has been shown to fail the rationality test. One should not treat IV as if it were the volatility expectation from an efficient market, but should make use of IV as a source of information and attempt to construct an efficient forecast from it. It is appropriate to adjust the raw IV to remove any correctable biases and possibly to combine it with information from other sources (perhaps historical volatility or the projections from a GARCH model). How this should best be done is a major area in which additional research is called for.

References

Ball, Cliff. AA Review of Stochastic Volatility Models with Application to Option Pricing.@ Financial Markets, Institutions & Instruments 2, 1993, pp. 55-71.

Beckers, Stan. @Standard Deviations Implied in Option Prices as Predictors of Future Stock Price Variability.@ Journal of Banking and Finance 5, Sept. 1981, pp. 363-82.

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

____________. AThe Pricing of Commodity Options.@ Journal of Financial Economics 3, March/June 1976b, pp. 167-79.

____________ and Myron Scholes. AThe Pricing of Options and Corporate Liabilities.@ Journal of Political Economy 81, May/June 1973, pp. 637-54.

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

Bollerslev, Tim, Ray T. Chou and Kenneth Kroner. AARCH Modeling in Finance.@ Journal of Econometrics 52, 1992, pp. 5-59.

Brown, B.W. and S. Maital. AWhat Do Economists Know? An Empirical Study of Experts= Expectations.@ Econometrica 49, 1981, pp. 491-504.

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

Canina, Linda and Stephen Figlewski. AThe Informational Content of Implied Volatility.@ Review of Financial Studies 6, No. 3, pp. 659-81.

Chiras, Donald and Stephen Manaster. AThe Information Content of Option Prices and a Test of Market Efficiency.@ Journal of Financial Economics 6, June/Sept. 1976, pp. 213-34.

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

Day, Theodore and Craig Lewis. AForecasting Futures Market Volatility.@ Journal of Derivatives 1, Winter 1993, pp. 33-50.

Engle, Robert. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of

United Kingdom Inflation." Econometrica 50, 1982, pp. 987-1000.

Fair, Ray and Robert Shiller. AComparing Information in Forecasts from Econometric Models.@ American Economic Review 80, 1990, pp. 375-89.

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

Figlewski, Stephen. AOptions Arbitrage in Imperfect Markets.@ Journal of Finance 44, Dec. 1989, pp. 1289-1311.

Fleming, Jeff. AThe Quality of Market Volatility Forecasts Implied by S&P 100 Index Options Prices.@ Working Paper, Jones Graduate School, Rice University, 1996.

Ferri, Arthur. Unpublished Dissertation Research. New York University Stern School of Business, 1996.

Hansen, Lars. ALarge Sample Properties of Generalized Method of Moments Estimators.@ Econometrica 50, 1982, pp. 1029-54.

Hull, John and Alan White. AThe Pricing of Options on Assets with Stochastic Volatilities.@ Journal of Finance 42, June 1987, pp. 281-300.

Jorion, Philippe. APredicting Volatility in the Foreign Exchange Market.@ Journal of Finance 50, June 1995, pp. 507-28.

Lamoureux, Christopher and William Lastrapes. AForecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities.@ Review of Financial Studies 6, No. 2, pp. 293-326.

Latané, Henry and Richard Rendleman, Jr. AStandard Deviations of Stock Price Ratios Implied in Option Prices.@ Journal of Finance 31, May 1976, pp. 369-81.

Lo, Andrew and Jiang Wang. AImplementing Option Pricing Models when Asset Returns are Predictable.@ Journal of Finance 50, March 1995, pp. 87-129.

Macbeth, James and Larry Merville. "Tests of the Black-Scholes and Cox Call Option Valuation Models." Journal of Finance 35, May 1980, 285-300.

Mayhew, Stewart. AImplied Volatility.@ Financial Analysts Journal 51, July/Aug 1995, pp. 8-20.

Morgan Guaranty Trust. RiskMetrics--Technical Document, 3rd ed. New York: Morgan Guaranty Trust Company Global Research, May 1995.

Neal, Robert. AIs Program Trading Destabilizing?@ Journal of Derivatives 1, Winter 1993, pp. 64-77.

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.

Theil, Henri. Applied Economic Forecasting. North-Holland, Amsterdam, 1966, chapter 2.

Whaley, Robert. AValuation of American Call Options on Dividend-Paying Stocks: Empirical Tests.@ Journal of Financial Economics 10, March 1982, pp. 29-58.

White, Halbert. AA Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.@ Econometrica 48, 1980, pp. 817-838.