Measuring Implied Volatility: Is an Average Better?

Louis Ederington*
Wei Guan**

August 2000

*Michael F. Price College of Business
Finance Division, Room 205 Adams Hall
University of Oklahoma
Norman, Oklahoma 73019-4005
405-325-5591
lederington@ou.edu

**School of Management
Department of Accounting and Finance
Delaware State University
Dover, Delaware 19901-2277
302-857-6919
wguan@dsc.edu

Measuring Implied Volatility: Is an Average Better?

Abstract

Based on the law of large numbers, several options researchers have proposed using (different) weighted averages of the implied standard deviations, ISD, calculated from numerous options with the same expiry to obtain a single best ISD measure. However, most commercial providers use an average (and often an equally weighted average) of just a few at-the-money ISDs. We find that the practitioners' restricted averages forecast slightly better than the broader weighted averages from the academic literature but that neither group forecasts actual volatility very well.

We suggest an adjustment to the extant models which improves their forecasting ability considerably. We also suggest a new weighting scheme which yields better estimates on an out-of-sample basis than any of the existing models (adjusted or unadjusted). However, we also find that because there is little independent noise in individual options ISDs (at least in the S&P 500 options market), there is little gain to averaging. Consequently, individual option ISDs and averages of just a few ISDs forecast almost as well as weighted averages of many ISDs and the weighting scheme choice is relatively unimportant.

Measuring Implied Volatility: Is an Average Better?

In most options markets, numerous options with the same expiry but different strike prices are traded. When plugged into the appropriate pricing model, the price of each yields an implied volatility of the underlying asset's price over the remaining life of the option, i.e., an implied standard deviation or ISD. Since they are forecasting volatility in the same asset over the same period, if markets are perfect and the pricing model is correct, implied volatilities calculated from options with the same expiry but different strike prices should be identical and so should implied volatilities from calls and puts. In reality, however, puts and calls with the same expiry and strike yield somewhat different ISD estimates and so do options with different strikes. On the one hand, there are persistent implied volatility patterns in many markets, i.e., the well known volatility smile. However, even in the absence of a smile, some implied volatility differences are expected due to market imperfections such as bid-ask spreads, non-synchronous price observations (between the option and the underlying asset), and discrete prices. These imperfections would be expected to cause the implied volatility calculated from a single option's price on day t to differ from the market's true expectation.

Based on the law of large numbers, several options researchers have advocated combining ISDs from all available options with the same expiry to obtain a single measure of implied volatility over the life of the option and have suggested several weighting schemes for doing so.[1] These authors argue that an average of the ISDs calculated from options with the same expiry should be a more efficient estimator of the market's true volatility expectation than that calculated from any single option alone since noise from market imperfections should tend to average out. In other words, in a large sample, the noise can be averaged away and the more observations (i.e., options) used to estimate implied volatility the better.

Ignoring this advice, commercial vendors of implied volatility figures normally use only a few at-the-money options to calculate implied volatility. For instance, the implied volatilities reported by Bloomberg are simple averages of the ISDs on the two closest-to-the-money calls (or puts). Knight Ridder sells implied volatility estimates which are simple averages of four ISDs, the two closest-to-the-money calls and closest-to-the-money puts while the CBOE provides a weighted average of the same four options. While not stated explicitly, the likely rationale for this parsimony is the volatility smile since persistent cross-sectional implied volatility differences imply that the ISDs at some strikes are biased estimators of the market's true volatility expectations. By choosing at-the-money strikes, practitioners are presuming that these yield more accurate measures of implied volatility. Many academic researchers use only a few strikes to obtain volatility estimates as well. For instance, among those studies basing their implied volatility estimate on only one strike are: Lamoureux and Lastrapes (1993), Kleidon and Whaley (1992), and Christensen and Prabhala (1998).

We show that neither the weighted averages of many options proposed in the academic literature nor the averages of a few ISDs from at-the-money options provided by such services as Bloomberg and Knight Ridder provide implied volatility estimates which accurately forecast actual market volatility. A relatively naive model in which recent past volatility is simply extrapolated into the future provides forecasts which are either better than or almost as accurate as those measures proposed in the literature or used in practice.

We find that both the "academic" and "vendor" models fail primarily because they do not adjust for the ISD biases implicit in the smile. Some of the academic or weighted average models also fail because they give high weights to less efficient strikes. However, we show that because the bias persists, it is possible to obtain forecasts of future volatility based on the available ISD's which forecast considerably better than the extant models. We do this by first correcting the Black-Scholes implied volatilities for their persistent bias and then averaging together the corrected ISDs. We also propose a new weighting scheme in which the various ISDs are weighted by their estimated relative efficiency. This new model forecasts better out-of-sample than the existing models (either with or without our bias correction). However, we find that, at least in our market, options on S&P 500 futures, the weighting scheme matters little (after one corrects for bias persistence) because there is very little noise in individual ISD estimates to be averaged out. While our weighting scheme is the best, it is not significantly better than other models and it matters little whether one averages together many ISDs or just a few.

The paper is organized as follows. In the next section, we review the literature on combining information from numerous options with the same expiration date to obtain a single best forecast of future volatility. In section II, we describe our data, describe the exact ISD models to be compared, and present our measures of forecast accuracy. In III, we apply these extant models to the S&P 500 options market over the 1988-1998 period and show that none forecast actual stock market volatility very well. In IV, we develop our alternatives and use them to forecast volatility in-sample. Out-of-sample results are presented in V. In section VI, we measure the efficiency gains due to averaging. Section VII concludes the paper.

I. Implied Volatility Estimators in the Literature

Most of the suggested schemes in the academic literature for obtaining a single measure of implied volatility involve a weighted average of the implied volatilities calculated from numerous options with the same expiry. The law of large numbers provides the basic rational. If the pricing model is correct and option traders are rational, the implied standard deviation, $ISD_{j,t}$, calculated from option j's price on day t equals the markets volatility expectation, EV_t, plus a random error term $\varepsilon_{j,t}$, i.e., $ISD_{j,t}=EV_t+\varepsilon_{j,t}$. If $\varepsilon_{j,t}$ is independently and identically distributed with mean zero and variance σ^2 for all options j, then (as is well known) a simple average of the J $ISD_{j,t}$, $\overline{ISD}_t$, has mean EV_t and variance σ^2/J. The more options J used to calculate the average, the more efficient is $\overline{ISD}_t$, i.e., the smaller its variance.

While simple averages are often used in practice, e.g., Schmalensee and Trippi (1978), Jorion (1995), and Weber (1996), most researchers who have addressed this issue argue that the J ISDs are not equally efficient so that a weighted average is more appropriate. Let $ISD_{j,t}$ represent the implied standard deviation calculated from the price of option j and the underlying asset on day t. Define $\overline{ISD}_t$ as a weighted average of the $ISD_{j,t}$ for J options with the same expiry:

$$\overline{ISD}_t = \sum_{j=1}^{J} w_{j,t} ISD_{j,t} \tag{1}$$

If again $ISD_{j,t}=EV_t+\varepsilon_{j,t}$ and the $\varepsilon_{j,t}$ are independently distributed with mean 0 but different variances, σ_j^2, then the variance of $\overline{ISD}_t$ is minimized by setting:

$$w_{j,t} = \frac{1}{\sigma^2_{j,t} \sum_{i=1}^{J} \frac{1}{\sigma^2_{i,t}}} \tag{2}$$

In other words, ISDs from options j with a relatively small variances, σ_j^2, should receive larger weights than high variance ISDs. However, as equation 2 illustrates, even relatively inefficient (i.e., high variance) options, help make $\overline{ISD}$ more efficient and hence should be included in the calculation of $\overline{ISD}_t$. Assuming independence, the variance of this $\overline{ISD}_t$ is $\sum_{j=1}^{J} w_j^2 \sigma^2_{j,t}$. Since $w_j<1$, the variance of $\overline{ISD}_t$ is reduced vis-a-vis the ISDs of the individual options.

If the errors, $\varepsilon_{j,t}$ are not independent, then the variance of $\overline{ISD}_t$ will be $\sum_{j=1}^{J} w_j^2 \sigma^2_{j,t} + \sum_{j=1}^{J} \sum_{i=1}^{j-1} 2 w_j w_i \sigma_{ij,t}$ where $\sigma_{ij,t}$ is the covariance between $\varepsilon_{i,t}$ and $\varepsilon_{j,t}$. If the covariances are negative (positive) the variance of $\overline{ISD}_t$ will be less (greater) than $\sum_{j=1}^{J} w_j^2 \sigma^2_{j,t}$. For instance, as pointed out by Fleming, Ostdiek, and Whaley (1995), measurement errors due to non-synchronous prices will tend to be of opposite signs for puts and calls so the variance of a simple average of the put and call ISDs should be less than half the average of the two.[2] Suppose the underlying asset trades more often than the options so that its closing price tends to be observed later.[3] If prices rise between the time the two option prices are observed and the time of the observation of the underlying asset's price, the implied volatility of the call will tend to be too low ($\epsilon_{j,t}$<0) while the ISD for the put will tend to be too high ($\epsilon_{i,t}$>0). If both options are at-the-money, the two errors should approximately offset.

While agreeing on the theoretical superiority of a weighted average, researchers have disagreed on the best weights, $w_{j,t}$. One scheme due to Latane and Rendleman (1976) is to set weights equal to the option's relative vega, i.e., $w_{j,t} = (\partial C_{j,t} / \partial ISD_{j,t}) / [\Sigma_j (\partial C_{j,t} / \partial ISD_{j,t})]$ where $C_{j,t}$ is the price of option j on day t.[4] Since vega measures the sensitivity of option j's price, $C_{j,t}$, to changes in $ISD_{j,t}$, the same noise in option prices $C_{j,t}$ due to market imperfections should create more noise in the implied volatilities of small vega options than in large vega options. Hence, weighting option j's ISD by its relative vega should give more weight to the lower variance options. Since vega is maximized when the strike price is just slightly above the underlying asset price, this scheme puts more weight on ISDs calculated from at-the-money options than options based on deep in-the-money or far out-of-the money options.

Chiras and Manaster (1978) propose a variation on this in which the weights are set equal to the option's price elasticity with respect to $ISD_{j,t}$, specifically $w_{j,t} = [(\partial C_{j,t} / \partial ISD_{j,t})(ISD_{j,t} / C_{j,t})] \div [\Sigma_j (\partial C_{j,t} / \partial ISD_{j,t})(ISD_{j,t} / C_{j,t})]$. Due to the addition of the term $(ISD_{j,t} / C_{j,t})$, this scheme puts more weight on out-of-the-money options (with low prices $C_{j,t}$) than the vega weights model. While the reasoning for vega weights is clear, why high elasticity options should be more efficient is not clear to us.

Beckers (1981) and Whaley (1982) propose estimating $\overline{ISD}_t$ by finding the single ISD which minimizes $\sum_{j=1}^{J} w_{j,t}[C_{j,t} - BS_{j,t}(ISD)]^2$ where BS refers to the Black-Scholes price according to that ISD. In Whaley's model $w_{j,t} = 1/J$ while in Becker's $w_{j,t} = (\partial C_{j,t} / \partial ISD_{j,t}) \div [\Sigma(\partial C_{j,t} / \partial ISD_{j,t})]$. According to Beckers, this is approximately equivalent to estimating equation 1 where $w_{j,t} = (\partial C_{t,j} / \partial ISD_{j,t})^2 \div [\Sigma(\partial C_{j,t} / \partial ISD_{j,t})^2]$. In other words, it is approximately equivalent to a traditional weighted average model in which the weights are the squared vegas. In fact we find that this ISD measure is most highly correlated (r=.99998) with a weighted average model where the individual ISD's are weighted by their relative vegas raised to the fourth power.

Since the weights are approximately proportional to the vegas raised to a high power, this scheme puts even more weight on close-to-the-money options than the vega weighting scheme of Latane and Rendleman.

As noted above, while these procedures all seek to obtain ISD's with little noise by averaging together ISDs from a large number of options j, commercial providers normally use just a few ISD's from at-the-money options. For instance, Bloomberg's call and put ISDs are equally weighted averages of the two nearest-the-money calls and two nearest-the-money puts respectively. Knight Ridder's implied volatilities are an equally weighted average of four individual option ISD's, the two nearest-the-money calls and the two nearest-the-money puts. The CBOE's Market Volatility Index, VIX, uses a weighted average of the same four ISDs but weights them so that the mean strike equals the underlying asset price.[5] While the procedures due to Latane and Rendleman (1976), Beckers (1981), and Whaley (1982) put more weight on the ISD's from options which are near-the-money, the Bloomberg, Knight-Ridder, and CBOE models focus on these exclusively. While the commercial providers do not provide an explicit reason for this choice, one possible rationale is the belief that ISD's of away-from-the-money options are biased, i.e., $E(ISD_{j,t})$…EV_t for away-from-the-money options. Nonetheless, restricting the set to at-the-money options only would appear to ignore a lot of information and to forego most of the efficiency benefits of averaging.

Several studies have compared some of these weighting schemes empirically with quite different results. Beckers (1981) found that, in the market for individual equity call options, the ISD for the single call with the highest vega tended to forecast slightly better than Becker's own scheme and better than a weighted average with vega weights. However, his data period was quite short: October 1975 through January 1976. Roughly consistent with Beckers, Feinstein (1989) found that in the S&P 500 options market, the single nearest but out-of-the-money call forecast marginally better than other models considered. In contrast, Gemmill's (1986) best performing ISD was the single ISD calculated from a deep in the money call - a result which is inconsistent with all suggested weighting schemes. Turvey (1990) found that in two commodity options markets, soybeans and cattle, a weighting scheme with vega weights forecast better than either an equally weighted model, elasticity weights, and a single at-the-money put. Note that in three out of four studies, a single ISD forecasts better that the averages that they consider - certainly not a good sign for the averaging models. None of these studies explore the reason for this results or why one scheme forecasts marginally better than another.

II. Data and Measures of Forecast Ability.

A. Data

Our data consist of daily closing prices of S&P 500 futures and options on S&P 500 futures traded on the Chicago Mercantile Exchange from January 4, 1988 through April 30, 1998.[6] Our data set begins January 4, 1988 for three reasons. One, trading was light in the first years (1983-1987) of the market so it may not have been as efficient. Two, until serial options were introduced in August 1987, only options maturing in March, June, September, and December were traded. After that date, we have a continuous monthly series. Three, several studies report a sizable, permanent shift in the smile pattern (from a smile to a sneer) after the 1987 market crash.

We utilize daily observations on options expiring within 10 to 35 trading days. We generally observe ISD's calculated from the closest-to-maturity options but on the ninth day before expiration we switch to the next contract since (1) the time value of short-lived options relative to bid-ask spreads is small making the ISDs less stable, and (2) in order to determine how well implied volatilities forecast actual realized volatility we need a sufficient sample of days over which to calculate realized volatility.

Each day we observe both calls and puts at a variety of strike prices for options with the same expiration date. In the S&P 500 options market, these strikes are in the increments of 5 points, e.g., 825, 830, 835 etc. Since trading is light in some far-from-the-money contracts, we restrict our sample to the first eight out-of-the-money and first eight in-the-money contracts relative to the underlying S&P 500 futures price. In summary, for each day we collect data on *up to* sixteen calls and *up to* sixteen puts. However, since all 32 options do not trade each day, we do not have 32 observations on all days.

Using Black's (1976) model for options on futures, day t closing prices for both the option and S&P 500 futures, and 3-month T-bill rates, we solve for the implied standard deviation, $ISD_{j,t}$, on each of the (32 or less) options j observed on day t.

In order to test the ability of these ISDs to forecast actual volatility, we need a measure of actual realized (or ex-post) volatility over this period, RLZ_t, which is calculated as the annualized standard deviation of returns over the period from day t through the option expiration date:

$$RLZ_t = \sqrt{252 \times \left[\frac{1}{N-t-1} \sum_{s=t+1}^{N} (R_s - \bar{R})^2\right]} \tag{3}$$

where $R_s = \ln(F_s / F_{s-1})$ and F_s is the closing futures price on day s. As noted above, $10 \le N-t \le 35$. Since all options j observed on day t expire on the same day, RLZ only has a *t* subscript.

B. Measures of Forecast Accuracy

We employ three measures of the ability of the various measures to forecast actual volatility. The first is the root mean squared forecast error defined as:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (RLZ_t - \overline{ISD}_t)^2} \tag{4}$$

The second is the mean absolute forecast error:

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |RLZ_t - \overline{ISD}_t| \tag{5}$$

Third is the mean absolute percentage error:

$$\mathrm{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|\,RLZ_t - \overline{ISD}_t\,|}{RLZ_t} \qquad (6)$$

Obviously, RMSE puts more emphasis on large errors than MAE and MAPE.

C. The ISD Models.

Using all of the (up to 32) individual $ISD_{j,t}$'s available on day t we compute measures of $\overline{ISD}_t$ for each of the major models used in practice or proposed in the literature. The ISD measures which we examine are:

ISD4 - An equally weighted average of the two closest-to-the-money calls and two closest-to-the-money puts. This average of at-the-money options is that calculated and sold by Knight Ridder and close to the measures reported by Bloomberg, who reports averages for the two closest-to-the-money calls and two closest-to-the-money puts.

ISDVIX - A weighted average of the two closest-to-the-money calls and two closest-to-the-money puts where the weights are chosen so that the average strike equals the underlying futures.[7] This is very similar to the Market Volatility Index distributed by the CBOE except that the time-to-maturity of ISDVIX matches the horizon of RLZ and the time-to-expiration of the other measures instead of being adjusted to a constant one month horizon.[8]

ISD32 - An equally weighted average of the ISDs calculated from all traded options with the same expiry. It is calculated using equation 1 with $w_{j,t} = 1/J_t$ where J_t is the number of options observed that day. If all 32 options are traded that day then $J_t = 32$.

ISDLR - A weighted average of all traded options (up to 32) calculated using equation 1 where $w_{j,t} = (MC_{j,t} / MISD_{j,t}) \div [\sum (MC_{j,t} / MISD_{j,t})]$. In other words, each ISD_j is weighted by its relative vega as proposed by Latane and Rendleman (1976).

ISDCM - A weighted average of all traded options (up to 32) calculated using equation 1 where $w_{j,t} = (MC_{j,t} / MISD_{j,t})(ISD_{j,t} / C_{j,t}) \div [\sum (MC_{j,t} / MISD_{j,t})(ISD_{j,t} / C_{j,t})]$. This is the weighted average based on elasticity weights proposed by Chiras and Manaster (1978).

ISDBW - the $\overline{ISD}_t$ which minimizes $\sum_{j=1}^{J_t} [C_{j,t} - BS(ISD_{j,t})]^2$ where J_t represents the number of options traded that day (up to 32) and BS denotes the Black-Scholes price. This is the measure proposed by Whaley (1982) and similar to that proposed by Beckers (1981).

As benchmarks, we also calculate two relatively naive non-ISD forecasting models:

HIS40 - The standard deviation of returns over the last 40 trading days.

CONSTANT - .13 which is the approximate mean of RLZ_t over the 1988-98 period.

III. Results

In Table 1, we report RMSE, MAE, and MAPE for each forecasting model over the 1988-1998 period.[9] We also rank the models from best (1) to worse (8) by each criterion. All three measures of forecast accuracy tell basically the same story. Among the ISD measures, the two best are the two commercial measures based on the four closest-to-the money options, ISD4 and ISDVIX, with the latter forecasting slightly better. Among the measures based on all traded options, the best is ISDBW, which (while it is not exactly a weighted average) is roughly equivalent to a weighted average with very high weights on at-the-money ISDs. It is followed by ISDLR , which weights at-the-money options higher than other ISDs, ISD32, and ISDCM. It is clear that forecast accuracy is directly related to how much weight the model places on at-the-money options.

However, it is clear that none of the ISD models forecast volatility very well. At the bottom of Table 1 we report RMSE, MAE, and MAPE for the two naive forecasts: CONSTANT and HIS40. The two naive forecasting models have the lowest MAE and MAPEs among the eight models (although their lead over ISDVIX and ISD4 is slight) and their RMSE is almost as low as that of ISD4 and ISDVIX. Based on these results it would appear that implied volatility contains little useful information regarding future volatility beyond the information from past volatilities (a proposition that we will show to be untrue below). According to the results in

Table 1, the contest between the various ISD models seems to be more a question of which one is the least worst - not which is the best.

IV. A Better Model

A. Obtaining unbiased ISD measures.

We suspect that one reason the ISD models do not improve on HIS40 and CONSTANT is because the $ISD_{j,t}$ are biased estimators of EV_t. We further suspect that the reason ISDVIX and ISD4 forecast better than the models based on all 32 options despite the fact that they ignore the information in all but at-the-money options is because the bias is worse in far-from-the-money options. Underlying the broad based models, such as ISDLR and ISD32, is the assumption that for all j $E(ISD_{j,t})=EV_t$, the market's volatility expectation. Given the volatility smile, it is clear that this cannot be the case for all j. In Table 2 we report time series means of $ISD_{j,t}$ over the 1988-98 period for all 32 options using the following nomenclature to identify calls and puts and the strike price grouping. The first letter, "C" or "P," indicates a **c**all or a **p**ut, the second, "I" or "O", indicates whether the option is **i**n or **o**ut of the money, and the last digit, "1" through "8", reports the strike price position relative to the underlying futures price where "1" is the closest to the money and "8" is the furthest in- or out-of-the-money. For example, CI3 indicates an in-the-money call option whose strike price is the third (or 10 to15 points) below the futures price.

As reported in Table 2, the differences in the time series means of $ISD_{j,t}$ by strike price are sizable. The mean ISDs at the top of the smile are over 50% larger than those at the bottom. The null that the means of $ISD_{j,t}$ are the same for all j is easily rejected at the .0001 level. Clearly, the expected value of all $ISD_{j,t}$ cannot be the market's volatility expectation EV_t .

In Table 2 we also report results of the regression:

$$RLZ_t = \alpha_{0j} + \alpha_{1j} ISD_{j,t} + u_{j,t} \qquad (7)$$

If $ISD_{j,t}$ equals the market's expectation of volatility over the remaining life of the option, EV_t, then market efficiency implies that α_{0j}=0.0 and α_{1j}=1.0 in Equation 7. Clearly, this null is rejected for most j. For all options j, $\alpha_{0j} > 0$ implying an upward bias and $\alpha_{1j}<1$ implying a downward bias. The latter bias normally predominates.

The regression results in Table 2 suggest a solution to the problem that $E(ISD_{j,t}) \neq EV_t$ for all j. Assuming that realized volatility, RLZ_t, varies randomly around the markets expectation, EV_t, as market efficiency suggests and that the bias is a consistent linear function of expected volatility, we define $ISD^*_{j,t}$ as:

$$ISD^*_{j,t} = \alpha_{0j} + \alpha_{1j}\, ISD_{j,t} \qquad (8)$$

where α_{0j} and α_{1j} are the parameter estimates from equation 7. In other words, if markets are efficient and the bias is consistent linear function of expected volatility, then we can use equation 8 to obtain unbiased estimators of EV_t. We do this for each j using the parameter estimates in Table 2.

B. The question of weights.

Assuming for the present that the bias in $ISD_{j,t}$ is predictable and linear so that the $ISD^*_{j,t}$ are unbiased, the logical next question is which weights, $w_{j,t}$, to use to obtain summary estimates ISD_t using equation 1. One possibility is to use the weights used in extant models. We do this for the three best ISD models from Table 2 using a * to designate that we replace the $ISD_{j,t}$ with $ISD^*_{j,t}$ in calculating the average. For instance, $ISD4^*_t$ is an equally weighted average of the four closest-to-the-money $ISD^*_{j,t}$ while $ISDVIX^*_{j,t}$ is a weighted average of the same four adjusted ISDs using the VIX weights. Applying vega weights to all 32 $ISD^*_{j,t}$, yields the revised measure: $ISDLR^*_t$.

However, the regressions in Table 2 suggest an alternative and theoretically appealing weighting scheme. The error term in equation 7 may be decomposed as follows:

$$u_{j,t} = RLZ_t - ISD^*_{j,t} = (RLZ_t - EV_t) + (EV_t - ISD^*_{j,t}) \quad (9)$$

Market efficiency implies that the expectation error (RLZ_t - EV_t) should be independent of all information available at time t including (EV_t - $ISD_{j,t}$*), hence:

$$Var(u_j) = \sigma^2 + \sigma_j^2 \quad (10)$$

where σ^2 represents the variance of the expectational error (RLZ_t - EV_t) and σ_j^2 is the variance of $\epsilon_{j,t}$ = (EV_t - $ISD_{j,t}$*) as defined in section I. As shown in section I, the most efficient weighting scheme according to statistical theory is one in which the weights are a function of σ_j^2 as specified in equation 2. Since estimates of Var(u_j) are available from the regression of RLZ_t on $ISD_{j,t}$, given an estimate of σ^2, we could obtain an estimate of σ_j^2 from equation and then derive the optimal weights using equation 2.

Unfortunately, we cannot estimate σ^2 since EV_t is unobservable. Given the size of the variance of $u_{j,t}$, we would expect the expectational error to account for most of $u_{j,t}$ but exactly how much is unclear. Consequently, we present results for four different weighting schemes based on alternative assumptions about σ^2. We proceed as follows. First, among the 32 equation 7 regressions in Table 2, we find the one with the smallest error variance, $Var(\hat{u}_j)$. Assuming that σ^2 accounts for X% of this minimum variance, we estimate σ_j^2 for all j using equation 10 and then determine the optimal weights using equation 2. We label the resulting ISD estimates, ISD(X)*$_t$. This is done for X = 75%, 90%, 95%, and 99%. A priori, we expect the latter three to be more appropriate than X=75%.

Measures of forecast accuracy, RMSE, MAE, and MAPE are reported in Table 3 for these seven models: ISDVIX*, ISD4*, ISDLR*, ISD(75)*, ISD(90)*, ISD(95)*, and ISD(99)*. For ease of comparison, we repeat the results for the comparable traditional ISD estimators from Table 1: ISDVIX, ISD4, and ISDLR, and also HIS40.

The results are striking. All seven new models based on the "corrected" ISDs predict actual volatility much better than any of the existing models in Table 1. They predict actual volatility much better than the models used by commercial vendors and much better than those proposed in the academic research literature. They also predict much better than the naive models (HIS40 and CONSTANT) in which past volatility is extrapolated into the future. The average percentage forecast error for ISD4 based on the raw ISDs is about 36% but when corrected or revised ISDs are utilized (ISD4*), this falls to 26%. ISDLR*'s MAPE is roughly 26% while ISDLR's was 44%. Clearly, there does seem to be a persistent but predictable bias in the $ISD_{j,t}$ which can be easily corrected resulting in much better estimates.

Among our six alternative models, differences in forecast accuracy are small. RMSE varies from a low of .04635 for ISDVIX* to a high of .04672 for ISDLR* - a difference of only 0.8%. MAPE ranges from a low of .2553 for ISD(99)* to .2594 for ISDLR* - a difference of only 1.6%. Apparently the weights are of secondary importance once the ISDs have been corrected. Reasons for this will be explored in section VI.

In Table 4 we report average weights according to each of four models: the vega weights model of ISDLR and ISDLR*, the elasticity weights model of ISDCM, and the error variance models ISD(75)* and ISD(99)*. We do not report weights for ISD(90)* and ISD(95)* since they fall between those of ISD(75)* and ISD(99)*. Note that, if all 32 options are traded on a given day, the equal weighting model ISD32 would assign a weight of 1/32=.03125 to each. However, all 32 options are not traded each day so all 32 do not enter into each calculation. If for instance only 25 are traded then the equal weights model would attach a weight of 1/25=.04 to all. For this reason the mean weights in Table 4 sum to more than 1.0.

Since vega is maximized when the strike price is just slightly above the underlying asset price, ISDLR and ISDLR* place the greatest weights on ISDs calculated from at-the-money options, CO1, CI1, PO1, and PI1. In contrast the elasticities model places the greatest weight on far out-of-the money options like CO8, CO7, and PO8. Our error variance models, particularly

ISD(99)* place the greatest weight on the ISDs with the lowest variances, i.e., the most efficient of the 32 options. These tend to be the options at moderately high strike prices, in other words somewhat in the money puts and out-of-the money calls like CO6, CO7, PI4, and PI5.

Although, as reported in Table 4, weights vary considerably from one model to the next, according to the results in Table 3 they do no make much difference once the ISDs are revised. The two of the best forecasting models according to Table 3 are ISD4* and ISD(99)* but they apply extremely different weights. While ISD4* gives a weight of .25 to ISDCO1, ISDCI1, ISDPO1, and ISDPI1, ISD(99)*'s average weights are: .023, .022, .021, and .025 respectively - 10% or less of ISD4*'s weights. While ISD4* assigns a zero weight to ISDCO6, ISDCO7, and ISDPI4, ISD(99)*'s average weights are: .406, .176, and .106 respectively. Yet ISD4* and ISD(99)* have virtually identical forecasting records.

In summary, the models which have been suggested in the literature for combining several ISD estimates into a weighted average, such as ISDLR, ISDCM, and ISDBW, do not forecast very well because the individual ISDs are biased measures of the market's true expectation. The ISD estimates provided by commercial providers like Bloomberg and Knight-Ridder fail for the same reason. However, we find that the bias tends to conform to a predictable linear form enabling one to use regression estimates to correct the bias. After obtaining unbiased ISD estimates, we propose combining the corrected ISDs in a weighted average in which the individual ISDs are weighted according to their estimated efficiency. However, the relative weights appear not to matter much - at least in this market.

V. Out-of-sample Results.

The results for the models in top half of Table 3 are all in-sample in that data for the entire data period 1988-1998 was used to obtain the parameter estimates in Table 2 and these parameters were then used to calculate the unbiased volatility estimates $ISD_{j,t}*$ which were then used to calculate $ISD4*_t$, $ISDVIX*_t$, $ISDLR*_t$, and the $ISD(X)*_t$ models. In other words, we use parameter estimates of "$_{0j}$ and "$_{1j}$ which are not known until 4/30/1998 to form our volatility measures each day t prior to that date. Obviously, this assumes insight which market participants do not in fact have prior to 4/30/1998. Likewise, the error variance models, ISD(75)*, ISD(90)*, ISD(95)* and ISD(99)*, use weights based on estimates of efficiency over the entire 1988-1998 period.

Consequently, we next calculate the $ISD4*_t$, $ISDVIX*_t$, $ISDLR*_t$, and the $ISD(X)*_t$ models using only information available on day t and compare the results with those for $ISD4_t$, $ISDVIX_t$, $ISD32_t$, etc. over an identical period. We use data over the four year period 1/4/88-12/31/91 to obtain initial estimates of the regression parameters in equation 7. These are used to calculate $ISD_{j,t}*$ for all 32 j for each day from ½/92 to 6/30/92. We also use the variances of the error terms from these regressions to obtain the weights used to calculate $ISD(75)*_t$, $ISD(90)*_t$, $ISD(95)*_t$, and $ISD(99)*_t$. We use the same parameter estimates to calculate each model's forecast each day from ½/1992 to 6/30/1992. On 6/30/92 we re-estimate the regressions using data from 1/4/1988 through 6/30/1992 and use these to calculate the forecast each day from 7/1/1992 to 12/31/1992. We proceed in this manner updating the parameter estimates every six months through 4/30/98.

The resulting measures of forecast accuracy, RMSE, MAE, and MAPE, are reported in Table 5 for all the ISD models. All three measures of forecast accuracy yield virtually the same results. First, as in the in-sample results in Table 3, all the new models based on regression corrected ISDs forecast much better than the traditional measures - both those touted in the literature and the simpler models used by practitioners. The mean absolute percentage forecast error among the new models is 29% to 31% versus about 37% for the best of the extant models, ISDVIX. Second, the differences in forecast accuracy among the seven new models is again relatively small though greater than the in-sample results in Table 3. The mean absolute percentage forecast error ranges from a low of 29.4% for ISD(99)* to 31.5% for ISD4*. Third, by all three measures, the best forecasting model is ISD(99)* followed by ISD(95)*, and

ISD(90)*. ISDVIX* which ranked first or second according to the in-sample measures in Table 3, forecasts slightly less well on an out-of-sample basis.

VI. The Gains to Averaging

As shown in section I, all the averaging schemes which have been suggested in the literature are based on the presumption that the $ISD_{j,t}$ calculated from any one call or put j on day t contain noise due to such factors as nonsynchronous trading, discrete prices, and bid-ask spreads and that this noise may be "averaged out" by an appropriate averaging scheme over several options. This proposition makes so much intuitive sense, that it has gone unchallenged up to now despite the fact that several studies reviewed in section I find that a single ISD forecasts better than the weighted averages that they consider.

However, the results above call this presumption into question. First, ISDVIX and ISD4, which consider only four options, forecast better than ISD32, ISDLR, and ISDCW which incorporate many more options. Likewise ISDVIX* and ISD4* forecast better than ISDLR*. Second, the best of the models ISD(X)* based on the estimated error variances of the ISD_j is ISD(99)*. This model assumes that (for the ISD_j with the lowest estimated error variance) 99% of the error variance is due to variation of actual volatility, RLZ_t, around the markets true expectation, EV_t, (an error which is common to all options j) and only 1% to variation of $ISD_{j,t}$* around EV_t. It is only this latter error which can be "averaged out" and these results imply it is minuscule.

To examine this issue more directly, we compare ISD4* to its four components. As before let $u_{j,t} = RLZ_t - ISD_{j,t}$*. As shown in equation 10 above, market efficiency implies that the variance of this forecast error term is: $Var(u_j) = \sigma^2 + \sigma_j^2$ where σ^2 represents the variance of the expectational error ($RLZ_t - EV_t$) and σ_j^2 is the variance of $\varepsilon_{j,t} = (EV_t - ISD_{j,t}$*). Let $u4_t = RLZ_t - ISD4_t$*. Since $w_{j,t} = .25$ for all j and t for $ISD4_t$*, if the four measurement error terms are independent, $Var(u4) = \sigma^2 + (1/16)\sum_{j=1}^{4}\sigma_j^2$. In other words the variance of ISD4*'s forecast error should be equal to the variance of the expectational error (which is common to all ISD_j*s) plus one quarter of the average of the four F_j^2. The gain in efficiency from averaging comes from the final term averaging out the variation in the $ISD_{j,t}$* around EV_t. If the F_j^2 are sizable and the $,_{j,t}$ are independent, then Var(u4) should be less than the average of the four Var(u_j). That is the variance of the average should be less than the average of the variances. If virtually all of Var(u_j) is due to F^2, not F_j^2, then there will be little difference between the average of the variances and the variance of the average.

In Table 6 we present estimates of Var(u4) and Var(u_j) for the four j on which ISD4* is based. We also present the average of the four Var(u_j).[10] Interestingly, we observe that the two ISD calculated from options with strikes above the underlying futures, CO1 and PI1, are more efficient than the two ISDs calculated from options with strikes below the underlying futures and more efficient than ISD4*. More important, the variance of ISD4*'s forecast error, Var(u4), is only slightly below the average of the four error variances, i.e., .0021497 versus .0021513. The implication is that either there is almost no noise, $,_{j,t}$, to be averaged out or the $,_{j,t}$ are highly correlated so that the noise cannot be averaged out. Indeed if we assume that the four $,_{j,t}$ are independent, the implication of the figures in Table 5 is that 99.7% of the variation of RLZ_t around an individual $ISD^*_{j,t}$ is due to variation of RLZ_t around EV_t and only 0.3% to measurement errors in the ISDs. In summary, contrary to the usual presumption, there is little to be gained by averaging together ISD estimates from puts and calls and different strikes.

However, a caveat is in order. The market for options on S&P 500 futures is very heavily traded so bid-ask spreads are likely to be low. Also S&P 500 futures and options on futures are traded side by side on the Chicago Mercantile Exchange and both markets close at the same time. Consequently, nonsynchronous prices are unlikely to be a problem. Averages of several ISDs may be more efficient in less heavily traded markets where the observed trades in the options and the underlying assets are likely to be at different times.

VII. Conclusions

In summary, neither the weighted averages models based on numerous ISDs, which have been proposed in the academic literature, nor the averages of just a few ISDs, which are sold by commercial vendors, forecast actual volatility very well. The vendors' restricted averages forecast slightly better than the more elaborate models proposed in the academic research literature but none forecast well.

These models fail primarily because most ISD's obtained from observed option prices using the Black-Scholes formula are biased. However, we show that these biases are persistent and predictable so that it is possible to obtain an unbiased and efficient adjusted ISD from a raw ISD with a simple linear transformation based on a regression of realized volatility on the raw ISD. Models based on these adjusted ISDs forecast actual market volatility much better than either the ISD averages sold by commercial providers or the models in the literature and much better than measures of historical volatility.

The regressions used to obtain unbiased ISDs also yield estimates of each ISD's efficiency which may be used to derive the error minimizing weights in a weighted average model. This procedure yields weighted average measures of implied volatility which forecast better than existing weighting schemes out-of-sample. However, the advantage of these new weighted averages over a simple arithmetic average based on the four adjusted or unbiased nearest-the-money ISDs is slight. Finally we show that, at least in this market, the efficiency gains to averaging are slight. Either there is little noise in the individual ISD estimates or the errors are highly correlated so that they cannot be averaged out.

References

Beckers, S., 1981, Standard deviations implied in option prices as predictors of future stock price variability, *Journal of Banking and Finance* 5, 363-381.

Black, F., 1976, The pricing of commodity contracts, *Journal of Financial Economics* 3, 167-179.

Chiras, D. and S. Manaster, 1978, The information content of option prices and a test of market efficiency, *Journal of Financial Economics* 10, 29-58.

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

Feinstein, S., 1989, A theoretical and empirical investigation of the Black-Scholes implied volatility, Ph.D. Dissertation, Yale University.

Fleming, J., B. Ostdiek, and R. Whaley, 1995, Predicting stock market volatility: a new measure, Journal of Futures Markets 15, 265-302.

Gemmill, Gordon, 1986, The forecasting performance of stock options on the London traded options market, *Journal of Business Finance and Accounting* 13, 535-546.

Jorion, Philippe, 1995, Predicting volatility in the foreign exchange market, *Journal of Finance* 50, 507-528.

Kleidon, A. W. and R.E. Whaley, 1992, One market? Stocks, futures, and options during October 1987, *Journal of Finance* 47, 851-877.

Lamoureux, C. and W. Lastrapes, 1993, Forecasting stock-return variance: toward an understanding of stochastic implied volatilities, *Review of Financial Studies* 6, 293-326.

Latané, H. and R. Rendleman, 1976, Standard deviation of stock price ratios implied by option premia, *Journal of Finance* 31, 369-382.

Schmalensee, R. and R. Trippi, 1978, Common stock volatility expectations implied by option premia, *Journal of Finance* 33, 129-147.

Turvey, C. G., 1990, Alternative estimates of weighted implied volatilities from soybean and live cattle options, *Journal of Futures Markets* 10, 353-356.

Weber, Carlene E., 1996, Ex ante stock market return volatility implied by the OEX option premium, *The Financial Review* 31, 585-602.

Whaley, R. E., 1982, Valuation of American call options on dividend-paying stocks: empirical tests, *Journal of Financial Economics* 10, 29-58.

Table 1
Measures of Forecast Accuracy for Extant Volatility Forecasting Models

For various volatility forecasting models, we report three measures of forecast accuracy: the root mean squared forecast error (RMSE), the mean absolute forecast error (MAE), and the mean percentage forecast error (MAPE). They are also ranked from the most accurate forecast (1) to the least (8). ISD4 is an equally weighted average of the ISDs for the four at-the-money options; ISDVIX is a weighted average of the four at-the-money ISDs where the weights are chosen so that the average strike is exactly at the money; ISD32 is an equally weighted average of the implied standard deviations, ISD, from all (up to 32) traded that day; ISDLR is a weighted average of all ISDs weighted by their vegas, ISDCM is a weighted average of all ISDs weighted by their price elasticities, and ISDBW is the ISD measure obtained by minimizing the sum of the squared pricing errors for the 32 options. CONSTANT=.13, and HIS40 is the actual volatility over the last 40 trading days. The forecast accuracy measures are calculated over the period 1/4/88 through 4/30/98.

Model	RMSE		MAE		MAPE	
	Level	Rank	Level	Rank	Level	Rank
ISD4	.05454	2	.04065	4	.3616	4
ISDVIX	.05441	1	.04051	3	.3597	3
ISD32	.06275	7	.05056	7	.4722	7
ISDLR	.06022	6	.04730	6	.4362	6
ISDCM	.06420	8	.05225	8	.4901	8
ISDBW	.05599	4	.04249	5	.3821	5
CONSTANT	.05522	3	.03995	1	.3587	2
HIS40	.05911	5	.04005	2	.3209	1

Table 2
The ISD Bias by Strike Price

We report time-series means of $ISD_{j,t}$. We also report estimations of the regression: $RLZ_t = "_{0j} + "_{1j}ISD_{j,t} + u_{j,t}$ where RLZ_t is the ex-post realized volatility from day t through the expiration of the options and $ISD_{j,t}$ is the implied standard deviation for option j on day t. Daily observations from 1/4/1988-4/30/1998 on S&P 500 futures options are utilized. In the “Strike Price” column, the first letter © or P) stands for a **C**all or a **P**ut; the second letter (I or O) refers to **I**n-the-money or **O**ut-of-the-money; and the last digit indicates the relative position of an option from the futures price where 1 indicates that the option is the nearest-to-the-money. * and ** designate parameters which are significantly different from zero at the .05 and .01 levels respectively. † and †† designate coefficients of $ISD_{j,t}$ which are significantly different from 1.0 at the .05 and .01 levels respectively. All are based on standard errors corrected for the serial correlation due to overlapping data. Tests of the intercept are two-tailed; those of the slope coefficients are one-tailed.

	Strike Price	**Mean ISD**	**$"_0$**	**$"_1$**	**Obs**
Calls	CI8	0.2200	0.0691**	0.2696**††	1,306
	CI7	0.2110	0.0525**	0.3559**††	1,475
	CI6	0.2008	0.0357*	0.4532**††	1,729
	CI5	0.1913	0.0281	0.5264**††	1,956
	CI4	0.1816	0.0275*	0.5491**††	2,248
	CI3	0.1723	0.0247	0.6006**††	2,447
	CI2	0.1645	0.0245*	0.6375**††	2,564
	CI1	0.1568	0.0266*	0.6574**††	2,596
	CO1	0.1497	0.0293*	0.6695**††	2,605
	CO2	0.1443	0.0293*	0.6948**††	2,605
	CO3	0.1404	0.0280**	0.7233**††	2,603
	CO4	0.1382	0.0244*	0.7610**††	2,575
	CO5	0.1380	0.0208*	0.7826**††	2,414
	CO6	0.1411	0.0182	0.7901**††	2,052
	CO7	0.1459	0.0208	0.7666**†	1,592
	CO8	0.1519	0.0306	0.7163**†	1,168
Puts	PO8	0.2269	0.0394*	0.3931**††	2,487
	PO7	0.2185	0.0345*	0.4350**††	2,574
	PO6	0.2071	0.0310*	0.4749**††	2,595
	PO5	0.1965	0.0297	0.5092**††	2,603
	PO4	0.1860	0.0274	0.5512**††	2,609
	PO3	0.1756	0.0269*	0.5869**††	2,608
	PO2	0.1658	0.0283*	0.6122**††	2,605
	PO1	0.1569	0.0297*	0.6358**††	2,598
	PI1	0.1493	0.0303**	0.6585**††	2,547
	PI2	0.1433	0.0297**	0.6838**††	2,442
	PI3	0.1405	0.0269**	0.7272**††	2,085
	PI4	0.1411	0.0221*	0.7706**††	1,625
	PI5	0.1448	0.0198	0.7805**††	1,230
	PI6	0.1520	0.0128	0.8056**†	972
	PI7	0.1650	0.0322	0.6675**††	694
	PI8	0.1765	0.0538**	0.5255**††	556

Table 3
Measures of Forecast Accuracy for Alternative Volatility Forecasting Models
In-Sample Results

For various volatility forecasting models, we report three measures of forecast accuracy: the root mean squared forecast error (RMSE), the mean absolute forecast error (MAE), and the mean percentage forecast error (MAPE). They are also ranked from the most accurate forecast (1) to the least (11). The first seven models are based on bias corrected ISDs using the regression results in Table 2. ISDLR* weights these corrected ISDs by their estimated vegas. ISD(75)*, ISD(90)*, and ISD(95)* weight the ISD's by their estimated efficiency based on the variance of the error term from the Table 2 regressions. The figure in parentheses is the percentage of the variation of actual volatility around the most efficient ISD which is assumed to be due to an expectational as opposed to measurement error. ISD4 is an equally weighted average of the four uncorrected at-the-money ISDs and HIS40 is actual volatility over the last 40 trading days. Forecast accuracy is measured over the period from 1/4/88 to 4/30/98.

Model	RMSE		MAE		MAPE	
	Level	Rank	Level	Rank	Level	Rank
bias corrected models:						
ISD4*	.04636	2	.03094	3	.2563	3
ISDVIX*	.04635	1	.03090	2	.2557	2
ISDLR*	.04672	7	.03120	7	.2594	7
ISD(75)*	.04658	6	.03115	6	.2596	6
ISD(90)*	.04647	5	.03104	5	.2580	5
ISD(95)*	.04640	4	.03097	4	.2569	4
ISD(99)*	.04637	3	.03089	1	.2553	1
traditional models:						
ISD4	.05454	9	.04065	10	.3616	10
ISDVIX	.05441	8	.04051	9	.3597	9
ISDLR	.06022	11	.04730	11	.4362	11
HIS40	.05911	10	.04003	8	.3209	8

Table 4
Option Weightings

We report the average weight each option receives in four weighted average models of the ISD: the vega weights model attributable to Latane and Rendleman, the elasticity weights to Chiras and Manaster, and two models based on the estimated error variance. In the "Strike Price" column, the first letter © or P) stands for a **C**all or a **P**ut; the second letter (I or O) refers to **I**n-the-money or **O**ut-of-the-money; and the last digit indicates the relative position of an option from the futures price where 1 indicates that the option is the nearest-to-the-money.

	Strike Price	Vega (LR)	Elasticities (CM)	Estimated Variance	
				75%	99%
Calls	CI8	0.0203	0.0040	0.0194	0.0072
	CI7	0.0229	0.0044	0.0267	0.0116
	CI6	0.0264	0.0049	0.0330	0.0169
	CI5	0.0320	0.0059	0.0281	0.0131
	CI4	0.0391	0.0072	0.0339	0.0177
	CI3	0.0478	0.0092	0.0343	0.0178
	CI2	0.0564	0.0126	0.0360	0.0192
	CI1	0.0626	0.0176	0.0384	0.0215
	CO1	0.0629	0.0251	0.0398	0.0230
	CO2	0.0556	0.0360	0.0408	0.0242
	CO3	0.0434	0.0507	0.0412	0.0247
	CO4	0.0313	0.0683	0.0425	0.0261
	CO5	0.0221	0.0845	0.0510	0.0397
	CO6	0.0168	0.0959	0.0720	0.4059
	CO7	0.0143	0.1035	0.0662	0.1760
	CO8	0.0135	0.1083	0.0393	0.0202
Puts	PO8	0.0193	0.0896	0.0281	0.0126
	PO7	0.0230	0.0816	0.0280	0.0126
	PO6	0.0274	0.0726	0.0298	0.0139
	PO5	0.0332	0.0631	0.0312	0.0149
	PO4	0.0402	0.0533	0.0330	0.0164
	PO3	0.0483	0.0435	0.0346	0.0178
	PO2	0.0565	0.0339	0.0363	0.0193
	PO1	0.0625	0.0250	0.0382	0.0213
	PI1	0.0627	0.0174	0.0416	0.0250
	PI2	0.0551	0.0116	0.0511	0.0409
	PI3	0.0424	0.0078	0.0464	0.0318
	PI4	0.0306	0.0058	0.0621	0.1058
	PI5	0.0226	0.0047	0.0566	0.0712
	PI6	0.0176	0.0038	0.0403	0.0231
	PI7	0.0159	0.0036	0.0229	0.0082
	PI8	0.0151	0.0035	0.0177	0.0058

Table 5
Measures of Forecast Accuracy for Volatility Forecasting Models
Out-of-sample Results

For various volatility forecasting models, we report three measures of forecast accuracy: the root mean squared forecast error (RMSE), the mean absolute forecast error (MAE), and the mean percentage forecast error (MAPE). They are also ranked from the most accurate forecast (1) to the least (14). The models are defined in Tables 1 and 3 and in the text. Forecast accuracy is measured on an out-of-sample basis over the period from ½/92 through 4/30/98

Model	RMSE		MAE		MAPE	
	Level	Rank	Level	Rank	Level	Rank
bias corrected models:						
ISD4*	.04122	5	.03141	7	.3149	8
ISDVIX*	.04117	4	.03136	6	.3143	6
ISDLR*	.04188	7	.03133	5	.3095	5
ISD(75)*	.04137	6	.03086	4	.3035	4
ISD(90)*	.04105	3	.03061	3	.3006	3
ISD(95)*	.04081	2	.03042	2	.2981	2
ISD(99)*	.04047	1	.03017	1	.2937	1
uncorrected models:						
ISD4	.04460	9	.03560	10	.3686	10
ISDVIX	.04446	8	.03545	9	.3665	9
ISD32	.05168	13	.04340	13	.4746	13
ISDLR	.04957	12	.04117	12	.4436	12
ISDCM	.05254	14	.04419	14	.4861	14
ISDBW	.04596	10	.03712	11	.3890	11
HIS40	.04668	11	.03449	8	.3147	7

Table 6
Measuring the Gains to Averaging

Variances of (RLZ_t - $ISD_{j,t}^*$) are reported where RLZ_t is actual volatility over the remaining life of the option and $ISD_{j,t}$* is the implied standard deviation calculated from option j on day t and corrected for its predictable bias using the parameters in Table 2. We report variances for (1) the four $ISD_{j,t}$* s calculated from the four closest-to-the money options, and (2) the average, $ISD4_t$*, of these four $ISD_{j,t}$*s.

Measure of $ISD_{j,t}$*	**Variance of (RLZ_t - $ISD_{j,t}^*$)**
ISDCI1*	0.0021670
ISDCO1*	0.0021337
ISDPI1*	0.0021369
ISDPO1*	0.0021677
Average of the four above	0.0021513
ISD4*	0.0021497

ENDNOTES

1. Prominent among these would be Latane and Rendleman (1976), Chiras and Manaster (1978), Beckers (1981), and Whaley (1982).

2. See Fleming, Ostdiek, and Whaley (1995). The same holds for errors due to imperfect measurement of the interest rate.

3. In the market for options on S&P 500 futures, this would be the normal case. In the market for options on the S&P index itself, the options would normally be observed later.

4. Actually Latane and Rendleman use a weighted geometric average not a simple weighted average as in equation 1. As a consequence, as shown by Chiras and Manaster (1978), their weights do not sum to one and the resulting $\overline{ISD}$ is biased toward zero. However, since the basic idea of vega weights was theirs, we like some others, credit the version of equation 1 with vega weights to them.

5. The VIX index also averages together these same four implied volatilities from two different times-to-maturity to maintain a constant average term-to-maturity.

6. Using options on futures rather than on the underlying index itself has a couple of advantages: (1) both the options and futures markets close at the same time alleviating the problem of non-synchronous price quotes, and (2) the option price (and therefore implied volatility) does not depend on assumed dividends. Of course, the S&P 500 index and its nearby futures prices are very highly correlated.

7. To illustrate the weighting scheme of ISDVIX suppose the underlying futures is 1004 and the two nearest strikes are 1000 and 1005. The call and put with a strike of 1005 would each receive a weight of .4 while the call and put with a strike of 1000 would receive a weight of .1. For a fuller description and analysis of the VIX index, see Fleming, Ostdiek, and Whaley (1995).

8. In calculating the VIX index, the CBOE averages together two ISDVIX measures. One for the nearby contract and one for the second nearby. The weights are adjusted so that the average term-to-expiration equals a constant 22 trading days. Obviously that is inappropriate here since the term-to-maturity of ISDVIX must match that of RLZ.

9. One of the four options used to calculate ISD4 and ISDVIX was not traded on 84 of the 2611 trading days in the 1988-1998 period so these are excluded from the sample leaving a final sample of 2527 trading days.

10. All these variances as well as the Var(u4) are calculated over days on which all four options traded so that $ISD4_t$ was based on all four individual ISDs.