Measuring Historical Volatility

Louis H. Ederington and Wei Guan

The adjusted mean absolute deviation is proposed as a simple-to-calculate alternative to the historical standard deviation as a measure of historical volatility and an input to option pricing models. We show that this measure forecasts future volatility consistently better than the historical standard deviation across a wide variety of markets. Moreover, it forecasts as well as or better than the GARCH(1,1) model. [C53, G13]

In pricing options, anticipated volatility over the life of the option is the crucial unknown. Estimation of this parameter and calculation of option prices are complicated by the fact that volatilities often differ substantially over time. This is illustrated in Exhibit 1, where we report standard deviations of returns for the S&P 500 index, short-term interest rates (Eurodollars), long-term interest rates (T-bonds), the yen/dollar exchange rate, and five individual equities for the 7/05/1967 - 9/10/2003 period divided into five subperiods approximately seven years in length. As shown there, volatilities differ substantially over time. For instance, the volatility of the S&P 500 index over the 1/1/1997-9/10/2003 subperiod was more than 80% higher than over the immediately preceding 1/1/1990-12/31/1996 subperiod. For the two interest rate series, volatility in the most volatile subperiod was more than double that in the least volatile subperiod. In every market, the null that volatility is unchanged is rejected at the .001 level. Moreover, many changes in volatility average out over seven-year subperiods. If we divide the period into subperiods shorter than seven years, the differences are even more dramatic.

The implications for option prices of different volatility estimations can be dramatic, especially for out-of-the-money options. Consider, for instance, the S&P 500 options market, the most studied and one of the most actively traded options markets. In this market, out-of-the money puts are the most actively traded contracts.[1] At the time of this writing, the S&P 500 index is approximately 1200, short-term interest rates are about 4%, and the dividend yield on S&P 500 stocks is about 1.9%. For a six-month put option with a strike of 1100, the Black-Scholes price calculated using the average 1990-96 volatility of 11.49% is $5.16. Using the 1997-2003 average volatility of 20.84%, it is almost five times as high at $25.38.

Estimation of likely future volatility is therefore a crucial step in option valuation. Although sophisticated volatility estimation procedures, such as GARCH, are popular among finance researchers, they require econometrics software that is difficult for the average undergraduate student (or casual options trader) to master and expensive to obtain, and so they have not found their way into most derivatives texts. Instead, most derivatives texts instruct students to use historical volatility, specifically, the historical standard deviation, over some recent period as the normal volatility input. Of course, implied volatility provides an alternative (and theoretically better) estimate of future volatility. However, for option

Louis H. Ederington is the Michael F. Price Professor of Finance at the University of Oklahoma in Norman, OK 73019. Wei Guan is an Assistant Professor of Finance at the University of South Florida St. Petersburg in St. Petersburg, FL 33701.

[1]This is apparently because money managers use out-of-the-money puts to hedge against a large decline in the value of their portfolios (Ederington and Guan, 2002).

Exhibit 1. Volatility of Daily Log Returns by Asset and Subperiod

The annualized standard deviation of log returns is reported for five subperiods from 10/24/67 through 9/10/03 for nine securities or series. Figures for the 10/24/67 to 12/31/75 period are not available for Eurodollar rate and yen/dollar exchange rate.

	Subperiod				
Market	**10/24/67 - 12/31/75**	**1/1/76 - 12/31/82**	**1/1/83 - 12/31/89**	**1/1/90 - 12/31/96**	**1/1/97 - 9/10/03**
S&P 500	13.86%	13.43%	17.96%	11.49%	20.84%
T-Bonds	9.43%	14.36%	13.24%	13.41%	19.36%
Eurodollars	NA	37.39%	18.12%	17.56%	17.93%
Yen/Dollar	NA	10.17%	9.63%	10.08%	11.91%
Boeing	40.53%	33.84%	27.50%	25.03%	39.22%
GM	22.61%	23.03%	25.36%	29.34%	34.96%
Int. Paper	28.18%	25.23%	29.94%	22.70%	37.90%
McDonalds	44.38%	23.66%	25.61%	23.52%	31.28%
Merck	24.18%	22.08%	22.68%	24.18%	32.42%

pricing purposes, this measure suffers from an obvious chicken and egg problem in that to calculate implied volatility requires the option price and to calculate the appropriate option price requires a volatility estimate. Hence, the historical standard deviation of log returns is the volatility estimator touted in most textbooks and is most commonly reported on options websites.

The historical standard deviation has several well-known shortcomings as an estimator of future volatility. One is that only the information in past returns is considered, ignoring other possible information sets, such as knowledge of future scheduled events that might move the markets, for example, a quarterly earnings report or an upcoming meeting of the Fed's Open Market Committee.[2] Another well-known potential problem is that all past squared return deviations back to an arbitrary date are weighted equally in calculating the standard deviation and all observations before that date are ignored (Engle, 2004, and Poon and Granger, 2003). Evidence on volatility clustering and persistence indicates that more recent observations contain more information regarding volatility in the immediate future than do older observations. Accordingly, more sophisticated models, such as GARCH and the exponentially weighted moving average model used by Riskmetrics, employ weighting schemes in which the most recent squared return deviations receive the most weight, and the weights gradually decline as the observations recede in time.[3] However, these procedures are both too complicated and too costly for the average student.

We see another potential drawback to the historical standard deviation. Because the historical standard deviation and variance are functions of *squared* return deviations, they could be unduly sensitive to outliers. In other words, if the period used to calculate the historical standard deviation contains a single highly volatile day, the standard deviation tends to be high and could overestimate actual volatility over the life of the option. We show that this is in fact the case that historical volatilities calculated over periods with an extreme observation tend to substantially overestimate actual volatility, implying substantial overestimation of an option's value.

[2]See, for instance, Ederington and Lee (1993, 2001). For an excellent review of the issues in volatility forecasting, see Poon and Granger (2003).

[3]For a nice review of the highlights of this literature, see Robert Engle's Nobel prize acceptance speech (Engle, 2004). Although the GARCH weighting scheme, in which the most recent squared return deviation receives the greatest weight and the weights decline exponentially, is theoretically more appealing than the historical standard deviation weighting scheme, in which all observations are weighted equally to an arbitrary cutoff point, evidence is mixed on whether GARCH actually forecasts better. In a comprehensive review of 39 studies comparing the historical standard deviation (or a related measure) and GARCH, Poon and Granger (2003) report that 17 studies find that GARCH forecasts better and 22 studies find that the historical standard deviation (or a related measure) forecasts better.

One way to avoid the undue impact of extreme observations is to use a longer period to measure the historical standard deviation. For instance, one might use returns over the last year instead of over the last month.[4] The problem with this approach is that it may dilute the information in recent volatility. Volatility tends to cluster. Specifically, high (low) volatility one week, or month, is often followed by high (low) volatility the succeeding week or month. Measuring historical volatility over a long period smoothes out the clusters and dilutes much of the information in recent volatility about likely future volatility.

We propose an alternative that is less sensitive to extreme observations while keeping the sample period short enough to reap the benefits of volatility clustering. Our suggestion is to measure the historical volatility of *absolute*, instead of *squared*, return observations and to use this measure to forecast future volatility. Specifically, we propose using the historical mean absolute return deviation instead of the historical standard deviation to forecast future volatility. Relatively, extreme observations are less extreme when measured in absolute rather than squared terms. Moreover, the mean absolute return deviation is easily calculated using rudimentary statistical software such as Excel.

A possible drawback of the mean absolute deviation is that it is distribution specific. That is, unlike the sample standard deviation, the relation between the mean absolute deviation and the population standard deviation differs among normal, beta, gamma, and other distributions. However, because the primary use of historical volatility is as an input into the Black-Scholes formula, which assumes log returns are normally distributed, basing the historical volatility estimate on the normal distribution is a natural choice and imposes no additional assumption in this application. We refer to the mean absolute return deviation with the adjustment factor suggested by the normal distribution as the "adjusted mean absolute deviation." We find that the volatility estimate based on the mean absolute deviation and on the normality assumption is fairly robust; that is, it generally forecasts actual volatility better than does the standard deviation even in markets where returns differ from normality.

[4]An alternative proposed by Andersen and Bollerslev (1998), among others, which gives less weight to extreme observations without lengthening the data period, is to switch from daily to intraday data. For instance, if the trading day is six hours, switching from one month of daily return observations to one month of hourly return observations increases the number of observations six times and cuts the weight on any one observation to one-sixth the weight on daily observations. However, intraday data are not readily available to students and casual traders, are costly, and are more difficult to handle. In any case, to our knowledge, no derivatives text recommends this approach.

In this article we compare the ability of the historical mean absolute return deviation and of the historical standard deviation to forecast future volatility (measured as the standard deviation of ex post log returns) across a wide variety of markets, including the stock market, long- and short-term interest rates, foreign exchange rates, and individual equities. Except for the Eurodollar and yen markets at long horizons, the adjusted mean absolute deviation forecasts future volatility better than does the historical standard deviation in most markets at almost all horizons. It is not surprising that the major exception, Eurodollars, is the market furthest from normality.

We also compare volatility forecasts generated by the historical mean absolute deviation with those generated by a GARCH(1,1) model. Conclusions on their relative forecast ability partially depend on how one measures forecast accuracy. By one of our two criteria, the mean absolute deviation dominates; by the other criterion, GARCH forecasts slightly better in most markets. Certainly, our results do not justify the extra time, effort, and cost involved in purchasing and mastering GARCH(1,1) software. It appears that an undergraduate armed with an Excel spreadsheet can forecast volatility as well as or better than an econometrician with more sophisticated software.

I. Measuring and Forecasting Volatility

Let $R_t = \ln(P_t/P_{t-1})$ represent daily returns on a financial asset.[5] The historical standard deviation over the last n days is measured as:

$$STD(n)_t = \sqrt{252\,\frac{1}{n}\sum_{j=0}^{n-1} r_{t-j}^2} \tag{1}$$

where $r_{t-j} = R_{t-j} - \mu$, the return deviation, and μ is the expected return. Often μ is replaced by the sample mean and accordingly the r^2 are divided by n-1, rather than n. The latter procedure implicitly assumes that the expected return over the coming period equals the mean return in the n-day period used to estimate STD(n). Given the low autocorrelation in returns, there is no justification for such an assumption, and Figlewski (1997) shows that better forecasts are normally obtained by setting $\mu = 0$. In our calculations, we set μ equal to the average daily return over the last five years.[6] Because volatilities are normally quoted in

[5]In the case of dividend-paying stocks, the numerator is changed to P_t+D_t, where D_t denotes any dividends paid from t-1 to t.

[6]Our results are not sensitive to this choice. Because the average daily return measured over long periods is small, it makes little difference whether μ is set equal to the mean over the last five years, the entire period, or zero. Moreover, the results are approximately the same if we use the sample mean and divide by n-1.

annualized terms, the sum of the squared daily return deviations is multiplied by 252, the approximate number of trading days in a year.

As seen in Equation 1, this volatility estimator assigns each squared lagged return deviation, r^2_{t-j}, after time t-n a weight of 1/n, and observations before t-n receive a weight of 0. An issue in applying this procedure is choosing the cutoff date n. Although setting the length of the period used to calculate historical volatility n equal to the length of the forecast period s is a common convention, Figlewski (1997) finds that forecast errors are generally lower if the historical variance is calculated over a longer period. Accordingly, we consider a variety of sample period lengths and designate the length of the sample on which STD is based with (n). For example, STD(20) indicates the annualized standard deviation of log returns over the last 20 trading days.

Our alternative measure of historical volatility is the adjusted mean absolute return deviation over the same n days. The (unadjusted) mean absolute return deviation is:

$$\mathrm{MAD}(n)_t = \frac{1}{n}\sum_{j=0}^{n-1} |r_{t-j}| \tag{2}$$

However, in this unadjusted form, MAD(n) is a biased estimator of the population standard deviation. Whereas STD converges to σ as sample size increases,[7] MAD converges to σ/k where k depends on the distribution. If the distribution is normal, $k = \sqrt{\pi/2}$ (Stuart and Ord, 1998). If the data follow another distribution, k takes a different value. Because most option pricing models (e.g., Black-Scholes) assume normality, we maintain this assumption and calculate the adjusted mean absolute deviation as:

$$\mathrm{AMAD}(n)_t = \sqrt{\frac{252\,\pi}{2}}\;\frac{1}{n}\sum_{j=0}^{n-1} |r_{t-j}| \tag{3}$$

This yields an unbiased estimator of σ if the r_t are normally distributed. Like STD(n), AMAD(n) is annualized by multiplying by the square root of 252.

In the early part of the 20th century there was debate among statisticians about the relative merits of STD and MAD as measures of volatility, but use of the sample standard deviation dominated because it is a consistent estimator of σ regardless of the data distribution and is more efficient for normal distributions. Nonetheless, it is argued by Staudte and Sheather (1990) and Huber (1996), among others, that MAD is more robust; that is, it yields better estimates of the population standard deviation if the data are contaminated.

In this article, STD(n) and AMAD(n) are compared in terms of their ability to forecast actual volatility over various future horizons, s, corresponding to likely option times to expiration. We (like all other studies in this area) measure this realized volatility as the standard deviation of daily log returns from t to t+s. Specifically,

$$\mathrm{RLZ}(s)_t = \sqrt{252\sum_{j=1}^{s}\frac{r^2_{t+j}}{s}} \tag{4}$$

Note that STD(n) and AMAD(n) are both evaluated using the same measure of ex post volatility, the ex post standard deviation of returns, RLZ(s), the measure used in virtually all previous studies of forecasting ability.

We compare the ability of STD(n) and AMAD(n) to forecast RLZ(s) using the two most common measures of forecast accuracy. The first is the root mean squared forecast error:

$$\mathrm{RMSFE} = [\,(1/T)\sum_{t=1}^{T}(\mathrm{RLZ}(s)_t - F(n)_t)^2\,]^{1/2} \tag{5}$$

where $F(n)_t$ designates the volatility forecast: $\mathrm{STD}(n)_t$ or $\mathrm{AMAD}(n)_t$. The second is the mean absolute forecast error:

$$\mathrm{MAFE} = (1/T)\sum_{t=1}^{T}\left|\mathrm{RLZ}(s)_t - F(n)_t\right| \tag{6}$$

Our main interest here is in which of these two simple volatility estimators, STD(n) and AMAD(n), forecasts realized volatility better. Both of these estimators are easily explained to and calculated by undergraduates. However, we are also interested in how AMAD(n) and STD(n) compare with the more sophisticated ARCH-GARCH models proposed by academic researchers. By far the most popular of these ARCH-GARCH models is the GARCH(1,1) model:

$$v_{t+1} = \alpha_0 + \alpha_1 r_t^2 + \beta v_t \tag{7}$$

where v_t represents the (unobserved) conditional variance at time t. Consequently, we also calculate how well this model forecasts actual volatility using RMSFE and MAFE.[8]

II. Data and Procedures

We compare the volatility forecasting ability of these three models for four financial assets with highly active

[7]Although the sample variance is an unbiased estimator of σ^2, the sample standard deviation is a biased estimator of σ in small samples. However, it is asymptotically unbiased and consistent.

[8]The model in Equation 7 yields a forecast variance for the next day, t+1. The forecast variance over the next s days is obtained by successive forward substitution following the procedure outlined in Ederington and Guan (2005).

options markets: the S&P 500 index, the 10-year Treasury bond rate, the 3-month Eurodollar rate, and the yen/dollar exchange rate. We also collect data for five equities chosen from those in the Dow Jones index: Boeing, GM, International Paper, McDonald's, and Merck. Daily return data for S&P 500 index and four of the five equities were obtained from the Center for Research in Security Prices (CRSP) for 7/3/62 to 12/31/02 and for McDonalds from 7/6/66 to 12/31/02. Daily interest rate and exchange rate data were obtained from Federal Reserve Board files for: 1/3/62-12/31/03 for the 10-year bond rate, and 1/5/71-12/31/03 for Eurodollars and the yen/dollar exchange rate.

Descriptive statistics for daily log returns are reported in Exhibit 2, where the mean and standard deviation are standardized for ease in interpretation. Skewness varies and is slight for all returns except those on the S&P 500 index. All returns, except those for Merck, exhibit excess kurtosis, which is most extreme for the S&P 500 index. Because the AMAD measure provides a consistent estimate of the population standard deviation only if the underlying distribution is normal, the Kolmogorov-Smirnov D statistic (hereafter, K-S D) test for normality shown in the last column is of particular interest. In all nine markets, the normality null is rejected at the .01 level. However, the K-S D is considerably higher for Eurodollar returns than for any of the other log return series, indicating a more serious deviation from normality in that market. This is surprising because both skewness and kurtosis are more serious for the S&P 500 index, suggesting that skewness and kurtosis do not completely account for deviations from normality. Given that the normality null is rejected in every market, it will be interesting to see how well AMAD(n) predicts.

Using the three procedures STD(n), AMAD(n), and GARCH, we forecast volatility over horizons, s, of 10, 20, 40, and 80 market days, or approximately two weeks, one month, two months, and four months. Because the best past period, n, for calculation of the STD(n) and AMAD(n) is unclear, they are calculated separately over historical periods, n, of 10, 20, 40, and 80 market days, and each is then used to forecast volatility over the four s horizons. The GARCH model is estimated using 1,260 daily return observations, or approximately five years of daily data. Consequently, our estimation periods for RLZ begin approximately five years after the beginning dates for our data sets reported previously, for example, 10/24/67 for the S&P 500 index and four of the five equities, and 12/3/71 for McDonald's. They end 80 trading days before the end of the data sets, for example, 9/6/02 for the S&P 500 index and the equities, and 9/10/03 for the interest and exchange rate series. To limit the computational burden, the GARCH model is re-estimated every 40 days.

III. The Historical Standard Deviation as a Forecast of Future Volatility

Before comparing the volatility forecasting ability of the three models, consider the hypothesized deficiency of the historical standard deviation. We posit that because it is calculated from squared surprise returns, the historical standard deviation is especially sensitive to outliers. If the period used to calculate the historical standard deviation contains a single highly volatile day, we hypothesize that the historical standard deviation will be high and will overestimate actual future volatility over the life of the option. To test whether this is the case, we consider all 20-day periods with at least one observation whose absolute return deviation is in the top 1%. For instance, for the S&P 500 index, this consists of all 20-day periods with at least one absolute return deviation greater than 2.92%. If there were no overlap, 20% of our 20-day periods would contain one of these extreme return observations, but because of volatility clustering, many contain more than one. Consequently, the actual percentage ranges from 10.34% for McDonald's to 15.41% for Boeing.

As reported in Exhibit 3, the historical standard deviation is indeed much higher when measured over 20-day periods with one or more extreme observations. The ratio of the average standard deviation over periods with an extreme observation to that over periods without one of these observations ranges from 1.717 for Merck to 2.535 for Eurodollars.

As the theory of volatility persistence predicts, volatility is indeed higher than normal for periods after one of these extreme observations. However, it is substantially and considerably less than the historical standard deviation predicts. We calculate the percentage mean volatility overprediction as (M_STD - M_RLZ)/M_RLZ, where M_STD is the mean historical standard deviation over past 20-day periods with an extreme observation and M_RLZ is the mean realized volatility over the subsequent 20-day periods. In Exhibit 3, we report both this mean overprediction and t-statistics of tests for the null that it equals zero. The mean overprediction ranges from 18.3% for Merck to 29.9% for Eurodollars, and it averages 25.9% across our nine markets. In every market, the null that the historical standard deviation does not overpredict actual volatility following one of these extreme observations is rejected at the .0001 level. When the historical period does not contain an extreme observation, the historical standard deviation slightly underestimates subsequent volatility.

Exhibit 2. Descriptive Statistics

Statistics are reported for annualized daily log returns for the following periods: S&P 500 and the five individual equities: 10/24/67–9/06/02 (McDonalds starts on 12/3/71), T-bonds: 2/20/67–9/10/03, Eurodollars and the yen/dollar exchange rate: 2/20/76–9/10/03.

	Mean	Standard Deviation	Skewness	Kurtosis	Kolmogorov-Smirnov D
S&P 500	0.06438	0.1560	-1.5263	37.949	0.0594
T-Bonds	-0.00236	0.1404	0.1261	5.011	0.0803
Eurodollars	-0.05706	0.2425	-0.3945	8.434	0.2191
Yen/Dollar	-0.03317	0.1047	-0.5084	4.551	0.0734
Boeing	0.10097	0.3379	0.1404	5.211	0.0561
GM	0.06468	0.2698	-0.1750	7.235	0.0475
Int. Paper	0.08518	0.2883	-0.4048	12.991	0.0462
McDonalds	0.11301	0.2905	-0.3224	7.953	0.0510
Merck	0.13056	0.2504	-0.0405	3.373	0.0451

Exhibit 3. Bias in the Forecasting Ability of the Historical Standard Deviation Following Periods with Extreme Returns

The ability of the historical standard deviation of returns calculated over the previous 20 trading days to forecast the standard deviation over the subsequent 20-day period is analyzed when the historical period contains at least one absolute return deviation in the top 1%. The percentage of historical periods with at least one extreme observation is reported in column 2, the ratio of the mean standard deviation for these periods to the mean for periods without an extreme observation is reported in column 3, the mean percentage prediction error for the subsequent 20-day period is reported in column 4, and t-statistics for tests of the null hypothesis that the mean prediction error is zero are reported in column 5.

	Percentage of 20-Day Periods with at Least One Extreme Observation	Ratio of STD for Periods with Extreme Observations to Those Without	Mean Percentage Overestimation of Subsequent Volatility	t-value for Test of Null That the Mean Overprediction is Zero
S&P 500	12.51%	2.143	25.77%	15.17
T-Bonds	12.11%	2.187	23.26%	21.49
Eurodollars	12.82%	2.535	29.91%	22.18
Yen/Dollar	13.87%	1.919	28.56%	23.81
Boeing	15.41%	1.840	26.98%	24.71
GM	13.97%	1.859	26.01%	21.51
Int. Paper	14.38%	1.839	26.52%	22.76
McDonalds	10.34%	2.170	18.28%	11.86
Merck	14.64%	1.717	27.66%	29.34
Average		2.023	25.88%	

IV. Results: Forecast Accuracy

Although the historical standard deviation, STD, overpredicts volatility when the calculation period contains an extreme observation, it might still predict better than AMAD and GARCH. After all, GARCH is also calculated from squared return deviations. Hence, we next compare the forecasting ability of the three models following the procedures outlined in Section II.

Calculations of how accurately the three measures forecast actual future volatility (RLZ) according to the RMSFE (Equation 5) and the MAFE (Equation 6) criteria are reported in Exhibit 4 for a 20-day forecast horizon. RMSFE statistics are reported in Panel A and MAFE in Panel B. STD(n) and AMAD(n) are calculated using sample period lengths n of 10, 20, 40, and 80 trading days. For each n, the measure, STD(n) or AMAD(n), with the lowest RMSFE and MAFE is shown in bold. Also, in each market, the cell of the model with the lowest RMSFE or MAFE (among all nine) is shaded.

As shown by cells in bold in Panel A, when forecast accuracy is evaluated in terms of the RMSFE in all nine markets for all four n (a total of 36 pairwise comparisons), the historical adjusted mean absolute deviation anticipates actual future volatility better than does the historical standard deviation. Moreover, in 14 of the 36 pair-wise comparisons, the forecasting ability of AMAD is significantly better than that of STD at the .05 level according to Diebold and Mariano's (1995) S1 statistic.

When forecast ability is evaluated using the MAFE in Panel B, the same results generally hold. However, for the 20-day forecast horizon, STD(10) has a lower MAFE than AMAD(10) for Eurodollars and GM. We regard this as of little consequence, as the results in Exhibit 4 indicate forecast accuracy is much better if a period longer than 10 days is used to forecast volatility over the subsequent 20 days. For longer data periods, AMAD has a consistently lower MAFE.

As shown by the shaded cells, the AMAD model forecasts better than the GARCH model in five of the nine markets according to the RMSFE criterion and in eight of nine markets according to the MAFE criterion. This surprising finding implies that by using the adjusted mean absolute deviation approach, an undergraduate student armed only with an Excel spreadsheet can forecast volatility at least as well as an econometrician with sophisticated software. This comparison is not quite fair to GARCH because in Exhibit 4 we compare the *lowest of four* AMAD models with GARCH. But if we restrict attention to AMAD(40) or AMAD(80), they still outperform GARCH more than half the time according to the MAFE criterion and break roughly even according to RMSFE. Of course, because it is dominated by AMAD, STD fares worse against GARCH. In most markets, GARCH forecasts better than the historical standard deviation by both criteria.

As shown in Exhibit 4, forecast accuracy is best when STD(n) and AMAD(n) are measured over 40 or 80 trading days instead of 10 or 20. This parallels Figlewski's (1997) finding that the historical standard deviation is best measured over a period longer than the forecast horizon. Therefore, to conserve space, in Exhibit 5, we report results using STDs and AMADs calculated over the past 40 days for the 10-day forecast horizon and over the last 80 days for the 40- and 80-day forecast horizons. These choices do not materially affect the comparison of STD and AMAD, and full results for all four calculation periods are available from the authors on request.

As shown by cells in bold in Panel A, when forecast accuracy is evaluated in terms of the RMSFE criterion, the historical adjusted mean absolute deviation anticipates actual future volatility better than does the historical standard deviation in all markets except for Eurodollars and yen at the 40- and 80-day horizons. At the 10-day horizon, the RMSFE of AMAD is significantly lower than that of STD in most markets, but the differences are not significant over longer horizons. Comparing AMAD with GARCH, in 17 of the 27 pair-wise comparisons in Exhibit 5, the RMSFE of GARCH is lower than that of AMAD.

As shown in Panel B, AMAD dominates both STD and GARCH according to the MAFE criterion. The MAFE of AMAD is lower than that of STD for all markets at all horizons except for Eurodollars and yen at the 80-day horizon. The MAFE of AMAD is significantly lower than that of STD at the .05 level in 15 of the 27 pair-wise comparisons in Panel B. In 19 of the 27 pairs, the MAFE of AMAD is lower than that of GARCH.

As explained in Section II, the AMAD measure is based on the assumption that log returns are normally distributed. As seen in Exhibit 2, there is evidence of non-normality for most of our series. Despite this, we find that AMAD forecasts better than STD; therefore, it seems to be robust to at least small deviations from normality. However, there is reason for caution. According to the K-S D statistic, the Eurodollar market is the furthest from normality. Along with the yen/dollar exchange rate, this is one of two markets where AMAD forecasts future volatility relatively poorly. Consequently, caution should be exercised when applying this technique to markets that differ significantly from normality. Users may want to test its forecasting ability first.

V. Conclusions

We propose the adjusted mean absolute return deviation as an alternative to the widely used historical

Exhibit 4. Volatility Forecast Accuracy for 20-Trading-Day Horizon

Root mean squared forecast errors (RMSFE) and mean absolute forecast errors (MAFE) are reported when the procedures listed in column 1 are used to forecast the standard deviation of returns over the next 20 trading days. STD(n) denotes the standard deviation calculated over the last n trading days and AMAD(n) denotes the adjusted mean absolute return deviation over the last n days. For each n, the STD(n) or AMAD(n) with the lowest RMSFE or MAFE is shown in bold. * Indicates that the RMSFE or MAFE of AMAD is significantly lower than that of STD at the .05 level according to Diebold and Mariano's (1995) S1 statistic. In each market, the cell of the model with the lowest RMSFE or MAFE is shaded.

	Markets								
Forecasting Model	**S&P 500**	**10-Year T-Bond**	**90-Day Eurodollar**	**Yen/Dollar**	**Boeing**	**GM**	**Int'l Paper**	**McDonalds**	**Merck**
Panel A. Root Mean Squared Forecast Errors (RMSFE)									
STD(10)	0.07247	0.05371	0.11795	0.04379	0.14263	0.11054	0.11977	0.11630	0.09517
AMAD(10)	**0.06910**	**0.05279**	**0.11419**	**0.04234***	**0.13631***	**0.10890**	**0.11327**	**0.11399***	**0.09380**
STD(20)	0.06898	0.04880	0.10855	0.03930	0.12464	0.10138	0.10930	0.10286	0.08405
AMAD(20)	**0.06375**	**0.04738***	**0.10415***	**0.03775***	**0.11674***	**0.09733**	**0.10260**	**0.09898***	**0.08206***
STD(40)	0.06711	0.04713	0.10101	0.03731	0.11403	0.09345	0.10076	0.09752	0.07894
AMAD(40)	**0.06184**	**0.04582***	**0.09797**	**0.03625**	**0.10763***	**0.08931**	**0.09496**	**0.09452***	**0.07712***
STD(80)	0.06827	0.04747	0.09829	0.03590	0.11015	0.09054	0.09677	0.09970	0.07773
AMAD(80)	**0.06366**	**0.04661**	**0.09793**	**0.03553**	**0.10622***	**0.08759**	**0.09101**	**0.09891**	**0.07625**
GARCH	0.06196	0.04502	0.09932	0.03675	0.10667	0.08642	0.09543	0.09277	0.07335
Panel B. Mean Absolute Forecast Errors (MAFE)									
STD(10)	0.03948	0.03904	**0.08016**	0.03176	0.10003	**0.07454**	0.08170	0.07786	0.07097
AMAD(10)	**0.03901**	**0.03870**	0.08111	**0.03095***	**0.09825***	0.07504	**0.08039**	**0.07768**	**0.07088**
STD(20)	0.03680	0.03541	0.07479	0.02847	0.08803	0.06714	0.07332	0.06656	0.06296
AMAD(20)	**0.03524***	**0.03448***	**0.07366**	**0.02714***	**0.08442***	**0.06608**	**0.07071***	**0.06483***	**0.06185**
STD(40)	0.03687	0.03433	0.07197	0.02675	0.08142	0.06171	0.06646	0.06289	0.05881
AMAD(40)	**0.03490***	**0.03287***	**0.06773***	**0.02533***	**0.07729***	**0.05972***	**0.06363***	**0.06073***	**0.05726***
STD(80)	0.03850	0.03511	0.07108	0.02641	0.08020	0.05937	0.06227	0.06331	0.05639
AMAD(80)	**0.03579***	**0.03349***	**0.06728**	**0.02516***	**0.07632***	**0.05678***	**0.05877***	**0.06164**	**0.05493***
GARCH	0.03510	0.03331	0.07509	0.02767	0.08146	0.05755	0.06383	0.06167	0.05467

standard deviation as a measure of historical volatility. Like the historical standard deviation, the adjusted mean absolute deviation estimator is easy for students to calculate with standard spreadsheet software. Yet, it forecasts actual future volatility better across a wide variety of markets and forecast horizons, enabling students to better price options. The inability of the historical standard deviation to forecast future volatility very well is apparently at least partially because, as it is a function of *squared* return deviations, a single outlier produces a big jump in the historical standard deviation, causing it to overestimate subsequent volatility. The adjusted mean absolute deviation model is less sensitive to such outliers because it is based on absolute, rather than squared, return deviations.

It may be surprising that this simple volatility estimator compares favorably with the more sophisticated GARCH(1,1) model. By one forecast accuracy criterion, it forecasts about as well as the GARCH(1,1) model and better by the other criterion. ■

Exhibit 5. Volatility Forecast Accuracy for Trading Day Horizons of 10, 40, and 80 days

Root mean squared forecast errors (RMSFE) and mean absolute forecast errors (MAFE) are reported when the procedures listed in column 1 are used to forecast the standard deviation of returns. STD denotes the historical standard deviation and AMAD denotes the adjusted mean absolute return deviation. Forty trading days are used to calculate STD and AMAD for the 10-day forecast horizon, and 80 days are used for the 40- and 80-day horizons. For each horizon, the lowest RMSFE or MAFE of the three models is shaded, and the lower of STD and AMAD is shown in bold. * Indicates that RMSFE or MAFE of AMAD is significantly lower than that of STD at the .05 level according to Diebold and Mariano's (1995) S1 statistic.

	Markets								
Forecasting Model	S&P 500	10-Year T-Bond	90-Day Eurodollar	Yen/Dollar	Boeing	GM	Int'l Paper	McDonalds	Merck
Panel A. Root Mean Squared Forecast Errors (RMSFE)									
10-Day Horizon: STD	0.07055	0.05244	0.11238	0.04233	0.13321	0.10382	0.11368	0.10923	0.09056
AMAD	**0.06577**	**0.05113***	**0.10800***	**0.04113***	**0.12687***	**0.10011**	**0.10849**	**0.10658***	**0.08847***
GARCH	0.06535	0.05057	0.11140	0.04230	0.12762	0.09770	0.10776	0.10595	0.08665
40-Day Horizon: STD	0.06657	0.04441	**0.08858**	**0.03221**	0.09725	0.08340	0.08699	0.09539	0.07077
AMAD	**0.06135**	**0.04372**	0.09018	0.03233	**0.09400**	**0.08041**	**0.08037**	**0.09435**	**0.06916**
GARCH	0.06060	0.04214	0.09021	0.03423	0.09237	0.07848	0.08647	0.08643	0.06465
80-Day Horizon: STD	0.06568	0.04331	**0.08302**	**0.02951**	0.09027	0.08016	0.08199	0.09396	0.06847
AMAD	**0.05962**	**0.04287**	0.08607	0.03022	**0.08724**	**0.07689**	**0.07395**	**0.09209**	**0.06644**
GARCH	0.06033	0.04130	0.08304	0.03195	0.08338	0.07339	0.08354	0.08580	0.05911
Panel B. Mean Absolute Forecast Errors (MAFE)									
10-Day Horizon: STD	0.03944	0.03808	0.07925	0.03059	0.09453	0.07023	0.07733	0.07297	0.06724
AMAD	**0.03745***	**0.03646***	**0.07313***	**0.02898***	**0.08980***	**0.06816***	**0.07482***	**0.07056***	**0.06500***
GARCH	0.03773	0.03700	0.08169	0.03166	0.09523	0.06629	0.07461	0.07208	0.06492
40-Day Horizon: STD	0.03814	0.03338	0.06606	0.02390	0.07180	0.05476	0.05527	0.06111	0.05094
AMAD	**0.03539***	**0.03176***	**0.06414**	**0.02345**	**0.06892**	**0.05237**	**0.05129***	**0.05963**	**0.04940***
GARCH	0.03578	0.03179	0.07161	0.02579	0.07272	0.05224	0.05754	0.05921	0.04802
80-Day Horizon: STD	0.03874	0.03333	**0.06296**	**0.02235**	0.06927	0.05368	0.05318	0.06153	0.04929
AMAD	**0.03595***	**0.03228**	0.06444	0.02258	**0.06642**	**0.05128**	**0.04902***	**0.05994**	**0.04801**
GARCH	0.03676	0.03233	0.06718	0.02474	0.06687	0.05123	0.05682	0.06083	0.04377

References

Andersen, Torben and Tim Bollerslev, 1998, "Answering the Skeptics: Yes Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review* 39 (No. 4, November), 885-905.

Diebold, Francis X. and Roberto S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (No. 3, July), 253-263.

Ederington, Louis H. and Wei Guan, 2002, "Why Are Those Options Smiling," *Journal of Derivatives* 10 (No. 2, Winter), 9-34.

Ederington, Louis H. and Wei Guan, 2005, "Forecasting Volatility," *Journal of Futures Markets* 25 (No. 5, May), 465-490.

Ederington, Louis H. and Jae Ha Lee, 1993, "How Markets Process Information: News Releases and Volatility," *Journal of Finance* 48 (No. 4, September), 1161-1191.

Ederington, Louis H. and Jae Ha Lee, 2001, "Intraday Volatility in Interest Rate and Foreign Exchange Markets: ARCH, Announcement and Seasonality Effects," *Journal of Futures Markets* 21 (No. 6, June), 517-552.

Engle, Robert, 2004, "Risk and Volatility: Econometric Models and Financial Practice," *American Economic Review* 94 (No. 3, June), 405-420.

Figlewski, Stephen, 1997, *Forecasting Volatility, Financial Markets, Institutions, and Instruments* 6 (No. 1, January) Stern School of Business, Boston, MA, Blackwell Publishers.

Huber, Peter, 1996, *Robust Statistical Procedures*, 2nd ed., Philadelphia, PA, Society for Industrial and Applied Mathematics.

Poon, Ser-Huang and Clive Granger, 2003, "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature* 41 (No. 2, June), 478-539.

Staudte, Robert and Simon Sheather, 1990, *Robust Estimation and Testing*, New York, NY, John Wiley & Sons.

Stuart, Alan and Keith Ord, 1998, *Kendall's Advanced Theory of Statistics: Volume 1 Distribution Theory*, 6th ed., New York, NY, Oxford University Press.