Using Historical Volatility for Stock Option Expensing under SFAS 123R: Improving forecasting performance with long memory and comovements[1]

George J. Jiang
Finance Department
Eller College of Business
University of Arizona
Tucson, Arizona 85721-0108
Phone: 520-621-3373
Email: gjiang@eller.arizona.edu

Yisong S. Tian[2]
Schulich School of Business
York University
4700 Keele Street
Toronto, ON M3J 1P3
Phone: 416-736-2100 ext. 77943
Email: ytian@schulich.yorku.ca

July, 2007

[1]We would like to thank Melanie Cao, Mark Kamstra, (Arizona colleagues???) and seminar participants at the 2006 *Northern Finance Association* annual meetings and *York University* for helpful comments and suggestions. The financial support of the *Social Sciences and Humanity Research Council of Canada* is gratefully acknowledged.

[2]Corresponding author.

Using Historical Volatility for Stock Option Expensing under SFAS 123R: Improving forecasting performance with long memory and comovements

Abstract

The *Financial Accounting Standards Board* recognizes historical volatility as an acceptable volatility forecasting method for option expensing purposes. In this paper, we empirically investigate the performance of using historical volatility to forecast long-term stock return volatility in comparison with a number of alternative forecasting methods. Analyzing forecasting errors and their impact on reported income due to option expensing, we find that historical volatility is a poor forecast for long-term volatility and the shrinkage adjustment toward the comparable-firm volatility only marginally improves its performance. Forecasting performance can be improved substantially by incorporating both the long-memory property and comovements with common market factors. If this improved forecasting method becomes the new industry standard for volatility forecasting, the potential range of managerial discretion over expected volatility estimation could be reduced substantially.

Key words: Volatility forecasting; SFAS 123R; Option expensing; Historical volatility; Shrinkage forecast; Long memory; Fractional integration; Comovements; Vector autoregressive.

JEL classifications: G13, G18, G38, M52, E37.

I. Introduction

U.S. companies are now required by the *Financial Accounting Standards Board* (FASB) to recognize, as opposed to disclosing it in footnotes, the full cost of employee stock options (ESOs) granted to their top executives and other employees. While the FASB's final statement on option expensing (*Statement of Financial Accounting Standards* No. 123R) provides fairly detailed guidlines, firms still have much leeway in their estimation of option fair value. In particular, management may have substantial discretion in estimating inputs to option pricing models such as expected stock volatility and expected option life. A number of recent studies (e.g., Balsam, Mozes and Newman (2003), Bartov, Mohanram and Nissim (2004), Aboody, Barth and Kasznik (2006), Johnston (2006), and Hodder, Mayew, McAnally and Weaver (2006)) have examined such managerial discretion and the possible motives. These studies find that a large proportion of firms appear to be motived by opportunistic incentives and exercise discretion over option pricing model inputs to reduce reported option value (relative to the level predicted by historical and/or industry experiences). They do this primarily by underestimating expected option life and expected stock volatility. At the same time, there is also some evidence that many other firms are motivated by informational incentives and exercise discretion to increase reported option value. These firms appear to incorporate private information about their future prospect and try to improve the ex post accuracy of reported option value.

In this paper, we do not examine the broader issue of managerial incentives to manipulate the reported stock option expense. Instead, we focus on the appropriateness and impact of using historical volatility to forecast expected stock volatility for option expensing purposes. Although SFAS 123R recognizes historical volatility as an acceptable volatility forecasting method for option expensing purposes, there is some evidence (e.g., Alford and Boatsman (1995)) that historical volatility is a poor forecast for future realized volatility. As prior research suggests, most firms appear to use historical volatility as the starting point in volatility forecasting and then make discretionary adjustments to finalize their reported volatility forecasts. If historical volatility is indeed a rather poor forecast for expected volatility, management may enherently have too much discretion in the estimation of expected volatility and can easily justify even fairly large deviations from historical experience. We empirically examine the performance of using historical volatility to forecast long-term stock return volatility in comparison with a number of alternative volatility forecasts. Analyzing forecasting errors and their impact on option value for a large sample of U.S. firms, we find that historical volatility is a poor forecast for long-term stock return volatility and suggest an alternative approach that substantially improves its forecasting performance. By providing a more reliable volatility forecasting method, our study may help reduce the level of potential discretion management has over reported volatility.

Given the possible impact of volatility forecasting on option expensing, it is surprising that there has been little prior research devoted to forecasting *long-term* stock return volatility. Although there is a vast literature on volatility estimation and forecasting,[1] virtually all previous research concentrates on forecasting *short-term* volatility (varying from one-day to one-year horizons). In comparison, the valuation of ESOs requires the forecasting of stock return volatility over three- to ten-year horizons. We are aware of only one previous study that examines this important issue for a large sample of U.S. firms. Alford and Boatsman (1995) empirically evaluate the performance of historical forecasts in predicting long-term stock return volatility over the sample period from 1965 to 1987. Their results indicate that the historical forecast is on average 17 to 19% off the target (based on median absolute percentage errors). This level of forecasting errors can translate into sizeable differences when option value is calculated using the estimated volatility.[2] A better approach to forecasting long-term volatility is thus needed.

Since stock return volatilities typically exhibit persistence over time and contemporaneous comovements with market-wide common factors (e.g., Ding, Granger, and Engle (1993), Bollerslev and Mikkelsen (1996, 1999), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), and Pong, Shackleton, Taylor and Xu (2004)), we propose an alternative volatility forecasting method based on the *Long-Memory Vector AutoRegressive* (LM-VAR) model using a system of multivariate

[1] See the survey papers by Poon and Granger (2003) and Andersen, Bollerslev, Christoffersen, and Diebold (2005) for a comprehensive review of the related literature.

[2] For an at-the-money option with five years remaining to maturity, a 17% forecasting error in volatility would translate into a 10% difference in option value (estimated using the Black-Scholes-Merton (1973) model assuming the firm's true volatility is 30%, the risk-free rate is 5% and the firm does not pay any dividends). Such differences are clearly economically too large to be ignored.

volatility time series. The LM-VAR forecasting method incorporates both long-memory (LM) properties and comovements (VAR) in stock return volatility and is still relatively simple to implement.[3] In particular, the LM-VAR volatility forecasts have several advantages over historical volatility based forecasts. First, the LM-VAR model separates short-term or transitional fluctuations from the long-term or permanent trend in the volatility time series and formulates volatility forecast using the latter instead of the former. Secondly, the LM-VAR model recognizes comovements of stock return volatility with common market factors and incorporates long-term trends in the latter to construct stock return volatility forecasts. Finally, the LM-VAR forecasts utilize all available historical data instead of truncating it at an arbitrarily chosen date.

For a sample of more than 2,000 U.S. firms, we empirically evaluate the performance of a number of volatility forecasts including three historical volatility based forecasts (i.e., historical, comparable-firm and shrinkage forecasts), volatility forecasts based on commonly used time series models (e.g., AR(1) and ARMA(1,1)), and those based on the LM-VAR model. For each volatility forecast, we evaluate its out-of-sample forecasting performance by comparing the volatility forecast with the actual stock return volatility realized over the forecasting period. The statistical significance of forecasting errors as well as their impact on option expensing are then analyzed. To ensure the robustness of our findings, we evaluate forecasting performance across different firm-size groups, forecasting horizons and performance measures.

The main findings from our empirical analysis are summarized as follows. First, historical volatility is a poor forecast for long-term stock return volatility. When historical volatility is used to predict future volatility over a five-year horizon, the median absolute percentage error is 22.8%, 20.7% and 22.0% for large, medium and small firms, respectively. These figures are comparable with but slightly higher than the corresponding figures (17 to 19%) previously reported in Alford and Boatsman (1995). The slightly elevated level of forecasting errors in our study is consistent with the differences in market conditions during the sample periods used in the two studies. Alford and Boatsman (1995) use an earlier sample period (prior to

[3]It is important to emphasize that for a volatility forecasting method to be acceptable for option expensing purposes, it must be relatively simple to implement. As we demonstrate subsequently, this is indeed the case with the LM-VAR volatility forecast. This important practical consideration has played a role in our choice of the LM-VAR model as a key alternative forecasting method to historical volatility.

1987) during which stock returns are generally less variable than they are in our sample period (from 1990 to 2004).

Secondly, the LM-VAR forecast is consistently more accurate than other volatility forecasts, with significantly smaller forecasting errors (both statistically and economically), across all firm-size groups, forecasting horizons and performance measures. For example, the LM-VAR forecast is on average 14 to 30% more accurate than the historical forecast in predicting long-term volatility (based on median absolute percentage errors). Such differences in forecasting errors can have a substantial impact on the valuation and expensing of stock options due to the sensitivity of option value to changes in volatility. Evaluating annual stock option grants awarded to CEOs in our sample, we find that the valuation errors are on average reduced by 47 to 73% if the LM-VAR forecast is used instead of the historical forecast.[4] Not surprisingly, the choice of volatility forecasts has a substantial impact on option expensing as well. We find that option valuation errors can on average impact reported income by 3.7%, 3.4%, and 8.2% for large, medium and small firms in our sample, respectively, if historical forecasts are used. The impact of option valuation errors on reported income is roughly cut in half if the LM-VAR forecasts are used instead. This level of differences highlights the importance of volatility forecasting and lends strong support for the LM-VAR model as an appropriate estimation and forecasting method for long-term stock return volatility.

In addition, volatility forecasts based on commonly used time series models (e.g., AR(1) and ARMA(1,1)) perform as poorly as historical volatility in predicting long-term volatility. Some of these forecasts exhibit marginal improvement while others perform even more poorly than historical volatility. Intuitively, these volatility forecasts are constructed to place more weight on the short-term variations in volatility than on the long-term trends or dynamics. Although they may be reasonable methods for forecasting short-term volatility, they are poor candidate for forecasting long-term volatility.

[4]We use the Black-Scholes-Merton (1973) model, adjusted for the expected life of the option due to premature exercise or forfeiture, to determine the option value because it is the primary valuation method recommended by the FASB. As the Black-Scholes-Merton model assumes constant volatility, there are potential inconsistencies when we use volatility forecasts from time-series based models (such as the LM-VAR model) as an input to the Black-Scholes-Merton model. A different option pricing model is required when stock returns are no longer normally distributed. Over longer horizons (e.g., five years), however, the impact of stochastic volatility and/or jump risk tend to average out over time and stock returns are roughly normally distributed. It is thus not unreasonable to use the Black-Scholes-Merton model in our study.

Finally, shrinkage forecasts, constructed by adjusting the historical volatility toward the comparable-firm volatility, at best perform only marginally better than historical volatility in predicting long-term volatility. Though generally consistent with the findings in prior research, our results suggest that the shrinkage adjustment makes little difference in the forecasting performance of historical volatility. In comparison, the LM-VAR forecasts perform substantially better than the historical forecasts (as well as the shrinkage forecasts) across all firm-size groups, forecasting horizons and performance measures.

Overall, our findings have important implications for forecasting long-term volatility and option expensing. Historical volatility is a poor forecast for future stock return volatility and can lead to economically large errors if it is used to estimate the fair value of ESOs. More importantly, we identify two key features of stock return volatilities – long memory (LM) and comovements (VAR) – that are particularly helpful in predicting long-term volatility. We also construct a relatively simple volatility forecast incorporating these two features and demonstrate its superior performance in forecasting accuracy and consistency. By providing a more reliable and more accurate forecasting method, our research helps reduce the level of potential discretion firms have over reported volatility.

The rest of the paper proceeds as follows. The next section examines the problem of forecasting long-term volatility and provides a detailed description of the LM-VAR forecast. Section III describes the data, sample selection and summary statistics of stock return volatilities. Section IV empirically investigates the performance of alternative volatility forecasts. Section V examines the economic significance of forecasting errors and the implications for option expensing. Section VI evaluates the performance of comparable-firm and shrinkage forecasts. Robustness analysis and further discussions are in Section VII. The final section concludes.

II. The LM-VAR Volatility Model

Suppose we are interested in predicting stock return volatility for firm i during period t. The volatility forecast is constructed from a time series of historical volatilities, $\sigma_{i,j}$ (for $j = t-1, t-2, ...$) where $\sigma_{i,j}$ is estimated from higher frequency stock returns during each time period. For example, a monthly volatility series can be constructed from daily stock returns. In general, this type of volatility forecasts can be written as:

$$E_{t-1}(\sigma_{i,t}) = f(\sigma_{i,t-1}, \sigma_{i,t-2}, ...). \tag{1}$$

The historical forecast used in Alford and Boatsman (1995) is a special case of this more general volatility model when $E_{t-1}(\sigma_{i,t}) = \sigma_{i,t-1}$. It essentially assumes that volatility is a random walk and past volatility is an unbiased forecast for future volatility.

Prior research (e.g., Ding, Granger and Engle (1993) and Andersen and Bollerslev (1997)) has established the importance of persistence in volatility time series and demonstrated that volatility persistence can be conveniently captured by a long-memory or fractionally-integrated (FI) process. Although transitory shocks to volatility (such as jumps) may play a role in forecasting volatility over short horizons, they are generally less informative about long-term trend in volatility movements. This is because jumps are rare events and their impact tends to be averaged out or much reduced over longer horizons. The long-memory feature captures the long-term trend in a volatility time series and is thus more important for forecasting long-term volatility.

Following prior research, we consider the ARFIMA(p, d, q) specification of equation (1):

$$\Phi(B)(1-B)^d \sigma_{i,t} = \Theta(B)\epsilon_{i,t}, \tag{2}$$

where B denotes the lag operator, $\Phi(B)$ the AutoRegressive (AR) polynomial $1 - \sum_{k=1}^{p} \phi_k B^k$, and $\Theta(B)$ the Moving Average (MA) polynomial $1 + \sum_{k=1}^{q} \theta_k B^k$, $(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k B^k$, and $\epsilon_{i,t}$ is white noise. The *AutoRegressive Fractionally Integrated Moving Average* (ARFIMA) process is stationary with long memory if $0 < d < 0.5$ and non-stationary if $d > 0.5$. The specification in equation (2) will be referred to as the LM model subsequently.

Another important dimension in forecasting long-term volatility is the correlation or comovement between stock return series. Over long horizons, the economy and the stock market tend to experience cyclical patterns. These market wide movements are likely to influence individual stock returns and should be incorporated into volatility forecasts. Following Ross (1976, 1977), we hypothesize that stock returns are generated by a multivariate APT model where the common factors may include broadly-based or industry-based stock market indexes. Within this general framework, stock return volatility is decomposed into systematic and idiosyncratic components. While the idiosyncratic component is firm specific, the systematic components capture the comovement of stock returns with the common factors. This well-known decomposition motivates the modelling of the joint dynamics of stock return volatility and volatilities of the common factors using a Vector-AutoRegressive (VAR) approach. The correlation or comovement between stock returns and the common factors should be helpful in forecasting the systematic components of the stock return volatility. Incorporating the common factors in the volatility model in equation (2), we propose a more general volatility forecasting method based on the following LM-VAR model:

$$\Phi(B)(1-B)^d X_{i,t} = \Theta(B)\epsilon_{i,t}, \tag{3}$$

where $X_{i,t}$ is the vector of stock return volatility ($\sigma_{i,t}$) for firm i and the volatilities of the common factors ($\delta_{k,t}$) in period t. In our subsequent empirical tests, we use monthly periods and model the joint dynamics of correlated monthly volatility time series.

Note that the proposed LM-VAR approach is quite general that it includes many standard volatility forecasts as special cases. In a univariate setting, the random walk, AR(p), and ARMA(p, q) models are all special cases of the LM-VAR model. For example, the historical forecast is a special case of the univariate LM-VAR forecast when $p = 1$ and $d = q = 0$. By embedding these commonly used volatility forecasts within our LM-VAR approach, it is straightforward to set up a *horse race* and evaluate the forecasting performance of competing volatility models. In addition, the LM-VAR approach also has some obvious advantages over the shrinkage forecast. The shrinkage forecast is constructed as a weighted average of historical and comparable-firm forecasts. As there are no theoretical guidelines for the appropriate weights to use, the weighting scheme is mostly chosen arbitrarily (with equal weights being the most common). In the LM-VAR approach, the weighting scheme is determined by the estimation process and optimized to provide the best fit to the volatility dynamics.

III. Data and Descriptive Statistics

To select the largest possible sample of U.S. firms for our empirical tests, we begin with all firms covered by the daily stock return database from the Center for Research in Security Prices (CRSP) over the time period from January 1, 1962 to December 31, 2004. To be included in our sample, we require firms to have return data available for the entire 15-year period from January 1, 1990 to December 31, 2004. The last ten years (from 1995 to 2004) of the sample period are used for out-of-sample testing in the subsequent empirical study while the preceding time period is used for volatility estimation. This inclusion criterion results in a final sample of 2,066 firms including 448 large firms, 550 medium firms and 1,068 small firms, based on the NYSE firm-size breakpoints in January 1990. Among the 2,066 firms in our sample, data availability varies widely from firm to firm. In each of the three size groups, there are some firms that have stock returns available for the entire 43-year sample period. In comparison, there are also firms in each group that barely have enough data to meet the minimum 15-year requirement. For each firm in the sample, we construct a monthly time series of stock return volatility. Following previous research (e.g., Merton (1980), Poterba and Summers (1986), French, Schwert, and Stambaugh (1987), and Schwert (1989)), we calculate the monthly volatility as the realized volatility during each calendar month using daily stock returns.

Table 1 reports descriptive statistics of the monthly volatility time series for firms in our sample. We calculate the mean, median, standard deviation, skewness, kurtosis, and the first five-order autocorrelations of the monthly volatility series for each firm in our sample. As these statistics (e.g., skewness) vary widely across firms, we report the average and median value of the statistics in each firm-size group as well as various percentiles to illustrate variations across firms. Results for large, medium and small firms are presented in Panels A-C, respectively. All reported volatilities and their statistics are annualized.

As expected, stock return volatility generally decreases with firm size. Take the firm-level average monthly volatility for example. The median value of this firm-level statistic is 27.4%, 32.0% and 53.0% for large, medium and small firms, respectively. In comparison, the corresponding average volatility on the

S&P 500 index is much lower at only 13.1%. It is even lower than the fifth percentile of the stock return volatility in each of the three size groups. This is not surprising due to the diversification effect across 500 stocks. In addition, monthly volatility time series for all firms are right skewed with fat tails as suggested by the positive skewness and excess kurtosis. This is consistent with the notion that volatilities usually vary within a normal range but can have occasional spikes or shifts as illustrated in Figure 1 which provides time series plots of the average monthly volatility for large, medium and small firms as well as the S&P 500 index.

More relevant to volatility forecasting are the dynamic properties of the volatility time series. The volatility time series of the S&P 500 index exhibits stronger autocorrelation than most individual firms do. As reported in the last column in Table 1, the first-order autocorrelation of the S&P 500 index volatility is 57.1%. In comparison, the average first-order autocorrelation is 47.4%, 45.6% and 48.7% for large, medium and small firms, respectively. These high levels of autocorrelation suggest volatility persistence for firms in our sample and particularly for the S&P 500 index. More importantly, the level of autocorrelation decays slowly as the number of lags increases (see the first five order autocorrelations in Table 1). This pattern of slow decay is further illustrated in Figure 3 which plots autocorrelations for up to 24 lags (or two years).

We also perform diagnostic analysis of the volatility time series which not only provides a better understanding of the dynamic properties of the volatility time series but also useful guidance for model specification. We fit the commonly used AR(1) model to the volatility time series of each individual stock and compute the diagnostic Ljung-Box statistics using 20 lags. The mean, median and other percentiles of the Ljung-Box statistics across firms in each size group are reported in Table 2. As the p-values for the Ljung-Box statistics indicate, the AR(1) is severely misspecified for virtually all firms in our sample.

Table 2 also reports the estimated degree of fractional integration (d), obtained using the Geweke and Porter-Hudak (1983) log-periodogram regression estimator as formally developed by Robinson (1995). The average estimated value of d is quite similar across the three firm-size groups, varying from 0.362, 0.352 to 0.375 for large, medium and small firms in our sample. In comparison, the corresponding estimate for the

S&P 500 volatility series is 0.425. As indicated by the 5th and 95th percentiles reported in Table 2, more than 90 percent of the estimated d values are within the range between 0.195 and 0.485 for each of the three firm-size groups, suggesting a stationary volatility process with long memory for these firms. Figure 2 plots the estimated values of d separately for large, medium and small firms in Panels A-C, respectively. It is clear that the overwhelming majority of estimated d values are between 0.2 and 0.5. There is no estimated value of d near or below zero. Although there is a small fraction of estimated d values that are slightly above 0.5, they are rarely statistically significantly above 0.5, judging by the standard error of the asymptotic normal distribution. These results provide strong evidence that the monthly volatility series is stationary with long memory for the S&P 500 index and nearly all firms in our sample.[5]

Figure 3 provides further visual illustration of the long-memory properties. It displays the sample autocorrelations of the volatility time series for up to 24 lags (or two years). The signature slow hyperbolic autocorrelation decay of long memory processes is evident in the plots. Figure 3 also shows the autocorrelations of the fractionally differenced volatility series which are obtained by applying the filter $(1-B)^d$ to the raw monthly volatilities. It is evident that the fractional differencing operator eliminates the bulk of the serial dependence in each of the monthly volatility series. The autocorrelation coefficients of the fractionally differenced volatility series are close to zero for the S&P 500 index and all three size-based groups. This finding indicates that fractional differencing based on the long-memory property of the raw volatility series effectively identifies all predictive components of the time series. In other words, the residuals from fractionally differenced volatility series are essentially random noise with no predictive component. This result has direct implications for volatility forecasting and confirms the importance of the long-memory property in the volatility time series.

Finally, the last row of each panel in Table 2 also reports the correlation between the monthly return volatility of individual stocks and that of the S&P 500 index. The average correlation is 45.4%, 31.3% and 18.4% for large, medium and small firms in our sample, respectively. The correlation in volatilities

[5]We also re-estimate the value of d for the monthly series of logarithmic volatilities. Untabulated results indicate that the logarithmic volatility series are also stationary with long memory, supporting the results in Table 2.

between large firms and the S&P 500 index is thus quite high. Not surprisingly, the correlation in volatility between firms and the S&P 500 index weakens as firm size declines. The S&P 500 index is primarily a stock market index for large firms. The correlation in volatilities between an individual firm and the S&P 500 index is also plotted in Figure 4, separately for large, medium and small firms. It is clear that the correlations may vary widely across firms and between size groups. For example, the 5th percentile for large firms is 23.8% compared to the 95th percentile of 67.3%. For small firms, the corresponding numbers are 7.1% and 45.9%, respectively. Nonetheless, it is important to recognize the correlation in volatilities between firms and the S&P 500 index. Such level of correlation suggests that the S&P 500 volatility is likely helpful in forecasting stock return volatility for these firms (more so for large firms) in a multivariate (VAR) framework. It provides direct support for our subsequent construction of a benchmark volatility forecast based on the LM-VAR model of the joint volatility time series of the firm and the S&P 500 index.

IV. Forecasting Performance

In this section, we empirically examine the performance of a number of volatility forecasts including historical forecasts, volatility forecasts based on commonly used time series models (e.g., AR(1) and ARMA(1,1)), and those based on the LM-VAR model. For the LM-VAR forecasts, we consider two additional variations that are based on either the LM model or the VAR model. Both volatility models are similar to the LM-VAR model except that one feature (LM or VAR) is dropped in order to focus on the contribution of the other feature. These two variations allow us to isolate each feature (LM or VAR) and determine its marginal contribution to the improvement in forecasting performance. Note that we do not include the GARCH model and its variations even though they are among the most commonly used volatility models in prior research. As documented in the literature, the GARCH models mainly capture volatility dynamics over relatively high frequencies (e.g., daily) and its volatility forecasts perform better over short horizons than over long horizons.[6]

[6] Although the long-memory property can be incorporated into the GARCH framework such as the FIGARCH(1,d,0) models proposed by Baillie, Bollerslev, and Mikkelsen (1996), we do not include these models in our analysis. The estimation and implementation of the multivariate FIGARCH(1,d,0) models are more complex and time consuming, particularly in the context of large-scale empirical tests such as ours.

To construct the volatility forecast for a given future horizon, the underlying volatility model must be first estimated using data available at the time of the forecast. For the historical forecasts, such estimation is rather trivial since they are simply calculated as the historical volatility over a most recent period prior to the forecasting date. For the remaining five volatility forecasts, model estimation is necessary. On each forecasting date, the volatility model is estimated using the entire volatility time series available to date. As the forecasting date moves forward, we re-estimate the model by including additional volatility observations available since the previous forecasting date. The AR(1) and ARMA(1,1) models are estimated using the standard approach. For the LM model, we set the autoregressive lag to 1 based on the diagnosis from the Akaike and Schwarz criteria.**Added**[7] For the VAR model, we implement a bivariate version of the model by including the volatility time series of the S&P 500 index and estimate the joint dynamics of the two volatility time series for the firm and the S&P 500 index. Finally, the LM-VAR model is estimated by applying an OLS estimation equation by equation under the normalization $\Phi(0) = I$. For justification of this approach and further details, see Andersen, Bollerslev, Diebold, and Labys (2003). **added** With the forward rolling volatility forecast in each subperiod, stock return volatility over a given forecast horizon, such as 5-year, is aggregated using the square-root of sum of sub-period volatility forecast.[8]

To determine the appropriate forecasting horizon for our empirical tests, we follow the FASB's guidelines for determining the fair value of stock options. Firms are required to use the Black-Scholes-Merton model or its binomial variations, modified to take into account the option's expected life or early exercise, to determine the fair value of a stock option. Although nearly all options are granted with a ten-year maturity, most are exercised between four and six years after the grant date (see Huddart and Lang (1996, 2003), Hemmer, Matsunaga and Shevlin (1996), Carpenter (1998), Heath, Huddart and Lang (1999), and Carpenter and Remmers (2001)). We thus primarily focus on a five-year forecasting horizon. Other forecasting horizons are also included for comparison purposes.

[7]As suggested by Corsi (2004), heterogeneous autoregressions can also pick up long memory properties in the data. As shown in Andersen, Bollerslev and Diebold (2006), this approach can further simplify practical implementation of long memory models.

[8]Given the fact that we are using historical data to forecast volatility over multi-horizons, e.g., 1-, 3- and 5-year horizons, a mixed frequency model such as the MIDAS model suggested by Ghysels, Santa-Clara and Valkanov (2006a, 2006b) can further improve the forecasting efficiency. We leave this for future research.

To evaluate the performance of each volatility model, we compare the volatility forecast with the realized volatility over the forecasting horizon. As in Alford and Boatsman (1995), the forecasting error is calculated as the difference between the volatility forecast and the realized volatility. Realized volatility is simply the sample standard deviation of daily stock returns over the specified forecasting period. The performance of a volatility forecast is then evaluated separately for each firm-size group and forecasting horizon, using various commonly used performance measures (or loss functions) including mean absolute errors, median absolute errors, mean absolute percentage errors and median absolute percentage errors. Model estimation and volatility forecasting are performed at the end of June and December in each calendar year during the period from 1995 to 2003. The results are summarized in Tables 3-5 for large, medium and small firms, respectively.

The main findings from our analysis of forecasting errors are as follows. First, historical volatility is a poor forecast for long-term volatility. When the five-year historical volatility is used to predict future volatility over the next five years, the median absolute percentage error is 22.8%, 20.7% and 22.0% for large, medium and small firms, respectively. These figures are comparable with but slightly higher than the corresponding figures (17 to 19%) previously reported in Alford and Boatsman (1995). The slightly elevated level of forecasting errors in our study is consistent with the differences in market conditions during the sample periods used in the two studies. Alford and Boatsman (1995) use an earlier sample period (from 1965 to 1987 (but excluding the October stock market crash in 1987)) during which stock returns are generally less variable than they are in our sample period (from 1990 to 2004). The differences between the two sample periods are clearly visible in the time series plot of historical volatilities in Figure 1. Interestingly, we also find that historical forecast should be estimated using past returns over a matching horizon as suggested by Alford and Boatsman (1995). The horizon-matching historical forecast appears to be more accurate than other historical forecasts. For example, the one-year historical volatility is a more accurate forecast for the one-year forecasting horizon than the five-year historical volatility is. Likewise, the five-year historical volatility provides a more accurate forecast for the five-year forecasting horizon than the one-year historical volatility does. Not surprisingly, the most recent historical volatility with a comparable horizon is a key factor that the FASB requires U.S. companies to consider in their volatility estimation process.[9]

Secondly, volatility forecasts based on commonly used time series models such as AR(1) and ARMA(1,1) perform as poorly as the historical forecast in predicting long-term volatility. Volatility forecasts based on the AR(1) model are generally more accurate than historical forecasts at longer forecasting horizons (e.g., the five-year horizon) but less accurate at shorter forecasting horizons (e.g., the one-year horizon), although the difference in forecasting accuracy appears to be quite minor in both cases. Intuitively, the AR(1) forecast is based on an adjustment to the mean using the intertemporal persistence of the volatility process. This is likely more beneficial at longer horizons than at shorter ones. The performance of the ARMA(1,1) model is similar except that it seems to perform more poorly than the AR(1) model at longer horizons (e.g., the five-year horizon). There appears to be a typical trade-off between in-sample fitting and out-of-sample forecasting performance. While the ARMA(1,1) fits the volatility process better in sample as a result of a more flexible model specification, the out-of-sample forecasting performance does not necessarily improve. This finding also suggests that over longer forecasting horizons, the autoregressive (or persistence) property is more useful than the moving-average property in forecasting volatility, providing further support for our long-memory based volatility forecasting approach.

More importantly, results in Tables 3-5 indicate that the LM-VAR model consistently provides the most accurate forecast for long-term volatility across firm-size groups, forecasting horizons and performance measures. Measured by the median absolute (percentage) error, the LM-VAR forecast is 14.9% to 31.3% (14.0% to 30.4%) more accurate than the historical volatility across firm-size groups and forecasting horizons. Consider the results for the large-firm sample with the five-year forecasting horizon (Panel C of Table 3) for example. The five-year historical volatility (HIS_5) has a median absolute (percentage) error of 0.066 (0.228). In comparison, the LM-VAR forecast has a median absolute (percentage) error of 0.054 (0.185) or a reduction of 18.2% (18.9%). More importantly, the advantage of the LM-VAR forecast over the historical

[9]See Section A32 of the FASB's final statement on share-based payment, Statement of Financial Accounting Standards No. 123R.

forecast is substantial across all firm-size groups and forecasting horizons. In addition, the LM-VAR forecast also performs better than either the LM or VAR forecast, highlighting the importance of both features in predicting long-term volatility. It is thus important to capture both long-memory and comovement features of the volatility dynamics in order to provide a more accurate forecast for long-term stock return volatility.

An interesting question is whether the improved performance of the LM-VAR forecast is statistically significant. Although it is difficult to find a universally acceptable loss function for the ex-post evaluation and comparison of the forecasting performance of nonlinear models, several statistical procedures have been used for assessing the quality of such forecasts as discussed by Andersen, Bollerslev, and Lange (1999) and Christoffersen and Diebold (1999). Among such procedures, the simplest test is a t- or F-test of the difference in mean absolute error between two volatility forecasts. For the results reported in Tables 3-5, untabulated t-test results indicate that the LM-VAR forecast is significantly different from other volatility forecasts at either the 5% or 10% level. The improvement by the LM-VAR over competing models is thus statistically significant.

There are also two additional attractive features of the volatility forecasts from the LM-VAR model. One such feature is that the forecasting errors of the LM-VAR model are not only smaller than those from other models but also less variable. In Tables 3-5, we also report the 75th and 95th percentiles for the absolute errors and absolute percentage errors. These percentiles are smaller for the LM-VAR model than those for any other model across all size groups and forecasting horizons. This level of consistency in forecasting performance makes the LM-VAR model very appealing. Another attractive feature of the LM-VAR model is that its forecasting performance is consistent over time. As our empirical analysis covers the ten-year period from 1995 to 2004, it is interesting to see whether the superior performance of the LM-VAR model is observed consistently over time. This is an important issue as market volatility may vary significantly over time and experience occasional spikes or shifts as the plot in Figure 1 indicates. We thus divide the ten-year period into two five-year subperiods and re-evaluate the forecasting performance of different volatility models. Untabulated results indicate that the LM-VAR model consistently outperform all other methods in both subperiods.

In summary, historical volatility is a poor forecast for long-term volatility while the LM-VAR forecast is substantially more accurate on a consistent basis. It also appears that both LM and VAR features of the LM-VAR model play an important role in improving forecasting performance. The long-memory feature of the model captures the persistence property of the volatility dynamics. By focusing on and extracting the long-run components of the volatility series, it provides a more accurate volatility forecast. The VAR setup also takes advantage of the comovement between volatilities of the target firm and the common market factors (such as the S&P 500 index). By capturing the long-term trend in the volatilities of the common factors and its potential impact on the firm's stock return volatility, it further improves the forecasting performance.

V. The Impact of Forecasting Errors on Option Expensing

In this section, we investigate the economic impact of volatility forecasting errors on the expensing of stock options. Specifically, we wish to find out how much impact the forecasting errors can have on the fair value of stock options and the firm's reported income due to option expensing . Before we empirically investigate this important issue, a simple numerical example provides an intuitive understanding of the potential impact on option valuation and expensing. Consider a hypothetical stock option with the following terms: the option is at the money, the stock return volatility is 30%, the risk-free rate is 5%, the option maturity varies from one to seven years, and the stock does not pay any dividends. These terms are roughly consistent with those of stock options granted to executives by a typical firm in our medium-size sample. In Figure 5, we plot the percentage pricing error of the above hypothetical stock option if the true volatility is overestimated or underestimated by 20% (or 6 percentage points). This level of estimation error is roughly consistent with the average forecasting error from using the historical forecast as reported in our Tables 3-5 and in Alford and Boatsman (1995). The Black-Scholes-Merton model is used to calculate option values. As shown in Figure 5, the volatility forecasting error can have a rather large impact on option value, ranging from 10% to 16%.

To further analyze the economic significance of volatility forecasting errors, we empirically investigate their impact on option values. We assess the impact of forecasting errors by calculating option values using both the realized volatility and the various volatility forecasts and then comparing and analyzing the differences. To do so, we need stock option data for firms in our sample. We thus cross reference firms in our sample with the Standard and Poors' ExecuComp database. We only keep firms that are covered by the ExecuComp and grant stock options to their top executives. Of the 448 large firms in the original sample, 390 firms still remain after cross referencing with the ExecuComp. As the firm size declines, the number of firms remaining drops off significantly. While 316 of the original 550 medium firms remain after cross referencing with the ExecuComp, only 176 of the original 1,068 small firms are left. The sharp drop-off, especially for small firms, is expected as the ExecuComp only include firms covered by the S&P 1500 index which tend to be larger than firms covered by the CRSP. Nonetheless, we still have a sizeable sample of remaining firms with a total of 804 firms.

For these 804 remaining firms in the sample, we evaluate the economic significance of forecasting errors by calculating the value of all stock option grants received by the CEO during the year the volatility forecast is constructed. We choose the CEO as he or she is the highest ranked executive in the firm and usually receives the most option grants. Only new option grants are included in our analysis because the ExecuComp provides details of the option terms (such as strike price, maturity date, and number of options in each grant) for new options only. We also make several simplifying assumptions: the Black-Scholes-Merton model is a reasonable model for estimating option value, all options have a five-year maturity, the risk-free rate is 5%, and the expected dividend yield over the next five years is identical to the actual dividend yield over the previous five years. These assumptions are reasonable since our goal is to examine the impact of volatility forecasting errors across different volatility models and these assumptions are unlikely to change the relative performance between volatility models in any biased manner. A potential inconsistency does arise when volatility forecasts from time-series based models (such as the LM-VAR model) are used as an input to the Black-Scholes-Merton model. While stock returns are assumed to be normally distributed with constant volatility in the latter, they are not in the former. A different option pricing model is required when stock returns are no longer normally distributed. Over longer horizons (e.g., five years), however, the impact of stochastic volatility and/or jump risk tend to average out over time and stock returns are roughly normally distributed. It is thus not unreasonable to use the Black-Scholes-Merton model this way in our study.

Tables 6-8 present the impact of volatility forecasting errors on option value and two incentive measures. In each table, the results for large, medium and small firms are shown in Panels A-C, respectively. For each volatility forecast, we first calculate the value of the CEO's option grants and two incentive measures, all in millions of dollars, using the volatility forecast. We repeat the calculation using the realized volatility over the expected life of the option (i.e., the five-year period after the grant date). The absolute error and absolute percentage error are then calculated for each volatility forecast. In addition to option values estimated using the Black-Scholes-Merton model, we also calculate two incentive measures that are commonly used in the executive compensation literature – the pay-performance sensitivity and the risk incentive of the CEO's option grants. The pay-performance sensitivity is Hall and Liebman's (1998) "dollars-on-percentage" measure, calculated as the change in the value of the option grants, in millions of dollars, if the firm value (or equivalently, the stock price) increases by one percent. The risk incentive is a variant of option vega, calculated as the change in the value of the option grants, in millions of dollars, if the stock return volatility increases by one percentage point.

Overall, the results in Tables 6-8 are consistent with the level of volatility forecasting errors reported in Tables 3-5 and the potential impact of volatility forecasting errors illustrated in Figure 5. As expected, a more accurate volatility forecast translates into improved estimates for option value and incentive measures. The estimation error is the smallest if the LM-VAR forecast is used. In particular, the errors from the LM-VAR forecast are generally less than half the size of the errors from the five-year historical volatility across all firm-size groups and alternative measures of forecasting errors. This level of improvement is clearly economically significant.

Consider first the impact of forecasting errors on option value reported in Table 6. For large firms (Panel A), the mean absolute (percentage) error in option value is \$1.084 million (21.6%) for the five-year historical volatility forecast. The corresponding figure for the LM-VAR forecast is \$0.479 million (9.1%) which is 56% (58%) smaller than the error for the historical forecast. Similar or even stronger improvement is observed when we examine the median, 75th and 95th percentiles of the absolute errors. The results for medium and small firms (Panels B and C) are qualitatively similar.

In addition, results in Tables 7 and 8 also suggest even stronger gains in accuracy when incentive measures are calculated using the LM-VAR forecast instead of the other volatility forecasts. Consider the impact of forecasting errors on incentive measures for large firms (Panel A in Tables 7 and 8). For the pay-performance sensitivity and risk incentive measures, the mean absolute error of the LM-VAR forecast is 65% and 71% smaller than the corresponding error of the historical forecast, respectively. The results for medium and small firms (Panels B and C) are qualitatively similar. It is therefore even more important to use the right volatility forecast when evaluating the incentive effects of stock options.

To further illustrate the economic significance of volatility forecasting errors, we assess the impact of forecasting errors on the firm's reported income due to option expensing. In order to do so, we calculate the total value of all options granted by each firm (to all employees including the CEO) during the year, using first the volatility forecast and then the realized volatility. We then determine the error in the estimated option value by calculating the percentage difference in the two option values. Finally, we translate both the option value and absolute error as a fraction of the firm's net income, in order to capture the impact of option expensing and volatility forecasting errors on reported income. The results are summarized in Table 9. Take the large-firm sample (Panel A) for instance. As reported in the top half of the panel, the mean (median) ratio of total option value to net income varies from 23.6% to 27% (from 5.2% to 6.4%) depending on the choice of volatility forecasts. These numbers indicate that mandatory option expensing would result in significant downward revisions to reported income for a large fraction of these firms. The choice of volatility forecasting methods also matters given the range of variations across volatility forecasts. As for the impact of forecasting errors on reported income (the bottom half of the panel), there are also wide variations across different volatility forecasts. Using the five-year historical volatility forecast, the mean (median) absolute error in total option value is 3.7% (0.8%) of net income. In comparison, the corresponding figure is 1.8% (0.3%) if the LM-VAR forecast is used. As a result, the LM-VAR forecast cuts the impact of forecasting errors on reported income by about half relative to that of the historical forecast. For option valuation and expensing, the choice of volatility forecasts is thus quite important.

For medium and small firms in our sample (Panels B and C, respectively), the impact of forecasting errors on option expensing is even stronger. For medium firms, the mean (median) ratio of total option value to net income varies from 46.4% to 49.4% (from 6.0% to 7.3%) depending on the choice of volatility forecasts. The corresponding figures for small firms are from 118.0% to 123.5% (from 15.1% to 16.0%). Not surprisingly, the impact of forecasting errors on reported income is also larger for these firms. For example, the mean (median) absolute error in total option value from using the five-year historical volatility is 3.4% (0.7%) of net income for medium firms and 8.2% (1.3%) for small firms. For the LM-VAR volatility method, the corresponding figure is 1.9% (0.4%) for medium firms and 3.6% (0.7%) for small firms. Again, the LM-VAR forecast reduces the impact of forecasting errors on reported income by about half relative to that of the historical forecast.

VI. Comparable-Firm and Shrinkage Forecasts

Previous research (e.g., Alford and Boatsman (1995)) suggests that a shrinkage adjustment toward comparable-firm volatilities can improve the forecasting performance of historical volatility. The basic idea is to combine the historical volatility of the target firm (i.e., whose volatility we wish to predict) with the historical volatilities of comparable firms. Although a firm's stock return volatility fluctuates over time, it tends to exhibit strong comovements with volatilities of comparable firms. The median or average volatility of comparable firms is thus useful in establishing the long-term trend in volatility movement and provides a useful benchmark for forecasting the target firm's future volatility. The shrinkage forecast is constructed by combining the target firm's historical volatility with the median or average historical volatility of the comparable firms. Although equal weights are typically used, other weights (e.g., 1/3 and 2/3) can be used as well.

Although the shrinkage forecast is expected to be more accurate than the historical forecast, it is unclear how its performance compares with that of the LM-VAR forecast. In this section, we empirically evaluate the performance of the shrinkage forecast in comparison with the LM-VAR forecast as well as the historical and comparable-firm forecasts. In order to do so, we need to match each firm in our sample with a number of comparable firms. The comparable-firm forecast is then constructed as the median historical volatility of the selected comparable firms while the shrinkage forecast is a weighted average of the historical and comparable-firm forecasts.[10] Once all volatility forecasts are constructed, we evaluate their forecasting performance by comparing them with the realized volatility over the forecasting period.

To select appropriate comparable firms, we follow prior research (e.g., Lev (1983), Christie (1982) and Karolyi (1993)) by considering firms in the same industry with similar size and leverage. We begin with the full sample of firms selected in Section III. For each firm (referred to as the *target firm*) in that sample, we search for ten comparable firms based on industry, firm size and leverage. We use the first two digit of the SIC to identify industries. Firm size is proxied by market capitalization while leverage is defined as the ratio of total debt to total assets. To qualify as a comparable firm, it must have at least three years of historical daily stock return data on the forecasting date, be in the same industry, and have both its size and leverage within 50% of the target firm. If there are less than seven firms meeting the requirement, the target firm is not included in the out-of-sample test on that particular forecasting date. Although the number of such firms varies across forecasting dates, they represent on average only 6% of the sample firms. If there are more than ten firms meeting the requirement, we select the top ten firms based on how closely they match up with the target firm. The quality of the match is proxied by a score based on the sum of squared percentage deviation in firm size and squared percentage deviation in leverage. This score summarizes how the comparable firm deviates from the target firm in both size and leverage and can vary from a minimum of zero to a maximum of 50 (%). We select the ten firms with the lowest scores as our choice of comparable firms for the target firm. Since firm size and leverage change over time, a different set of comparable firms are likely matched

[10]We follow Alford and Boatsman (1995) and construct the comparable-firm forecast as the *median* historical volatility of the comparable firms. We also consider the *average* historical volatility of the comparable firms as robustness check. The results are not materially different.

for the same target firm on different dates.

Table 10 reports summary statistics of the monthly volatility series for both the target firms (T.F.) and comparable firms (C.F.). For comparable firms, the monthly volatility is the median volatility for the ten comparable firms. Results for large, medium and small firms are separately reported in Panels A-C. Overall, these statistics illustrate similarities between the volatilities of the target and comparable firms. The two volatility series are matched quite well with comparable mean, median and other percentiles for all three size-based groups. More importantly, the correlation (ρ) between the volatilities of the target and comparable firms indicates substantial comovements between the two volatility series. As expected, the level of correlation is higher for large firms (with a mean of 0.291) than for small firms (with a mean of 0.220). In addition, the degree of fractional integration (d) reported in Table 10 suggests that the target and comparable firms share similar dynamic properties. Both volatility series are stationary with long memory. Table 10 also reports the correlation of the stock return volatility with the S&P 500 index volatility (denoted ρ(SPX) in the table) for both target and comparable firms. These correlations are very closely matched between target and comparable firms in every firm-size group.

We next empirically examine the performance of the shrinkage forecast in comparison with a number of alternative forecasts. For each target firm, we evaluate eight volatility forecasts including three historical volatility based forecasts (i.e., historical, comparable-firm and shrinkage forecasts) and five additional forecasts that incorporate either the LM, VAR or both properties. The historical forecast (denoted HIS) is constructed by calculating the target firm's historical volatility using daily stock returns over the most recent time period matching the length of the forecasting horizon. The comparable-firm forecast (COM) is the median historical volatility for the ten comparable firms, again using daily stock returns over the most recent time period matching the length of the forecasting horizon. The shrinkage forecast (SHR) is the equally weighted average of the historical and comparable-firm forecasts. We do consider other weighting schemes in subsequent robustness analysis and evaluate their impact on forecasting errors. The VAR2 forecast is constructed from a two-dimensional VAR model that combines the volatility time series of the target and comparable firms. The VAR3 forecast is constructed from a three-dimensional VAR model that combines the volatility time series of the target firm, the comparable firms, and the S&P 500 index. The LM forecast is constructed from a single volatility time series of the target firm, incorporating the LM feature. The LM-VAR2 forecast is similar to the VAR2 forecast except that the LM feature is incorporated in both volatility time series. Finally, the LM-VAR3 forecast is constructed similarly as the LM-VAR2 forecast except that the volatility time series of the S&P 500 index is added as another common factor.

Note that volatility forecasts based on the VAR feature alone are likely to perform more poorly than the corresponding forecasts based on both LM and VAR features. We include these purely VAR-based forecasts in the comparison because they have some similarities with the shrinkage forecasts (SHR). In particular, the VAR2 forecast is constructed using the same two volatility time series (of the target and comparable firms) as the shrinkage forecast is and imposes a relatively minimal model structure. Intuitively, the VAR2 forecast can be considered as a generalized form of the shrinkage forecast. The advantage of the VAR2 forecast is that the weights between historical and comparable-firm volatilities are determined by how the two volatility series correlate instead of the ad hoc scheme used in the shrinkage forecast. It is thus interesting to see how the VAR2 forecast compares with the shrinkage forecast in predicting long-term volatility.

Tables 11-13 summarize the forecasting errors for the eight volatility estimators for large, medium and small firms, respectively. As in Section III, we apply several performance measures to evaluate the forecasting errors including the mean absolute error, the median absolute error, the mean absolute percentage error and the median absolute percentage error. Overall, the results suggest that the three historical volatility based forecasts (i.e., historical, comparable-firm and shrinkage forecasts) are all poor candidate for forecasting long-term volatility, with the shrinkage (comparable-firm) forecast slightly more (less) accurate than the historical forecast. The two LM-VAR based forecasts (LM-VAR2 and LM-VAR3) provide the most accurate forecast and are substantially more accurate than the three historical volatility based forecasts.

To focus on the shrinkage forecast, we further examine its forecasting performance in comparison with the LM-VAR based forecasts. First, the shrinkage forecast is substantially less accurate than either the

LM-VAR2 forecast or the LM-VAR3 forecast in predicting long-term volatility. Take the sample of large firms with the five-year forecasting horizon (Panel C in Table 11) for instance. The shrinkage forecast has a median absolute (percentage) error of 0.064 (0.253). In comparison, the corresponding errors from the LM-VAR2 and LM-VAR3 forecasts are 0.057 and 0.053 (0.194 and 0.179), respectively. The forecasting error of the shrinkage forecast is, respectively, 12% and 21% (30% and 41%) larger than the corresponding errors of the LM-VAR2 and LM-VAR3 forecasts. A similar level of differences in forecasting performance is also observed for other firm-size groups, forecasting horizons, and performance measures. As expected, the LM-VAR3 forecast is the most accurate among all volatility forecasts.

Secondly, the shrinkage forecast is consistently less accurate than the VAR2 forecast across all firm-size groups, forecasting horizons and performance measures. Take the sample of medium firms with the five-year forecasting horizon (Panel C in Table 12) for example. The VAR2 forecast has a median absolute (percentage) error of 0.057 (0.183). In comparison, the corresponding error from the shrinkage forecast is 0.064 (0.217) or 12% (19%) larger. Since both volatility forecasts are constructed using the same two volatility time series, the reported differences in forecasting performance highlight the importance of incorporating comovement between volatility series. In comparison, the shrinkage forecast is constructed using arbitrary weights between the historical and comparable-firm forecasts regardless how the two volatility series are correlated.

Finally, the shrinkage forecast is substantially less accurate than the LM forecast across all firm-size groups, forecasting horizons and performance measures. Take the sample of medium firms with the five-year forecasting horizon (Panel C in Table 12) for instance. The LM forecast has a median absolute (percentage) error of 0.056 (0.178). In comparison, the corresponding error from the shrinkage forecast is 0.064 (0.217) or 14% (22%) larger. This is an interesting finding as the LM forecast is constructed using the target firm's volatility time series alone while the shrinkage forecast is constructed using the volatility time series of both the target and comparable firms. The shrinkage forecast is thus much less effective than the LM forecast in extracting relevant information from historical data to project future stock return volatility. The poor performance of the shrinkage forecast is likely due to a couple of inherent problems in its construction. One such inherent problem is that the shrinkage forecast is based on historical volatilities. As we have already demonstrated, historical volatility is a poor forecast for future realized volatility. It fails to incorporate the long-term trend or dynamics in the volatility time series as the LM forecast does. Another inherent problem of the shrinkage forecast is the ad hoc weighting scheme it applies to the historical volatilities of the target and comparable firms. The arbitrarily chosen weights do not reflect the correlation or comovement between the two volatility time series at all, which can add further variation in the performance of the shrinkage forecast.

VII. Robustness and Further Analysis

A key implementation issue for the LM-VAR forecast is the choice of common factors in the model. Aside from capturing the market-wide stock return movements, the selected common factors must also have two other important time series properties – stability and persistence. It should be less volatile than the stock returns of the firms whose volatility we are trying to forecast. A volatile common factor brings to the table both useful information on market movements and noise. An excessive level of the latter can substantially reduce the forecasting efficiency. The common factor should also be persistent over time, ideally exhibiting a high degree of long memory. Such common factors are more predictable and thus more helpful in identifying long term trends. We have focused on the S&P 500 index as the primary common factor in our empirical tests thus far. It is the most commonly used stock market index in both academic research and practical applications. The VIX volatility index, a widely-followed stock market volatility index constructed by the *Chicago Board Options Exchange*, is also based on the S&P 500 index. As we demonstrated previously (see Table 2), the S&P 500 index exhibits both required time series properties of a proper common factor for the LM-VAR model. The superior forecasting performance of the LM-VAR model reported previously supports the use of the S&P 500 index as a common factor.

Nevertheless, we have also considered other common factors such as the S&P 400 (mid-cap) index, the S&P 600 (small-cap) index, and the Fama-French factors in a robustness analysis. Unreported empirical tests indicate that the S&P 500 index is a far better common factor for the LM-VAR model than any of the alternative common factors. The LM-VAR forecasts constructed using the alternative common factors are usually less accurate than the corresponding LM-VAR forecasts constructed using the S&P 500 index. This is because these alternative common factors are either more volatile or less persistent than the S&P 500 index or both. Even for medium and small firms, it is better to use the S&P 500 index as the common factor than the S&P mid-cap and small-cap indexes in the LM-VAR model. The increased volatility and reduced persistence make these two indexes less attractive for volatility forecasting purposes. Similarly, the Fama-French factors have very little persistence in their volatility time series and are not suitable choices for the LM-VAR model. The choice of the common factor is thus quite important for ensuring the forecasting performance of the LM-VAR model.

In addition, we have considered volatility forecasts based on other time series models such as the AR(2), GARCH(1,1), GARCH(1,2) and GARCH(2,1) models and investigated how our LM-VAR forecast performs relative to them. Our prior expectation is that although these time series models may be reasonable choices for forecasting stock return volatilities over short horizons (e.g., one-day to one-month horizons), they are generally less suitable for forecasting long-term volatilities (e.g., one-year or longer horizons). Untabulated results indicate that our expectation is correct and these alternative time series models perform similarly as the AR(1) and ARMA(1,1) models in forecasting long-term stock return volatility. The LM-VAR model remains to be a far better choice for long-term volatility forecasting.

Another implementation issue is the selection of comparable firms in the shrinkage forecast. Results from Alford and Boatsman (1995) suggest that the forecasting performance can vary widely if different matching criteria are used to select the comparable firms. They find that the best forecasting performance is obtained when comparable firms are matched based on industry, firm size and leverage. Our empirical results are consistent with their findings. Untabulated results indicate that the best results are obtained when comparable firms are selected by matching all three factors (i.e., industry, firm size and leverage). The forecasting performance deteriorates slightly if either firm size or leverage is not matched. We also investigate the impact of weighting schemes on the performance of the shrinkage forecast. All reported results for the shrinkage forecasts are produced using equal weights between the historical volatilities of the target and comparable firms. In a robustness check, we reproduced the forecasting errors of the shrinkage forecast using either the [2/3, 1/3] or the [1/3, 2/3] weighting scheme between the two historical volatilities. Unreported results show that the two shrinkage forecasts with unequal weights perform slightly worse than the shrinkage forecast with equal weights. All our inferences still hold if we have used either of these unequal weighting schemes.

A related issue is how many comparable firms should be included in constructing the comparable-firm or shrinkage forecast. Following Alford and Boatsman (1995), we use ten comparable firms for each target firm. Is the forecasting performance affected by the number of comparable firms used? The answer is yes, although no previous study has actually addressed this intriguing question. To provide some definitive answers, we re-examine the empirical tests in Tables 11-13 but now vary the number of matching firms from one to 15. Untabulated results show that the forecasting performance of the shrinkage forecast improves as the number of comparable firms increases. The improvement tails off sharply after ten comparable firms are used. More importantly, the shrinkage forecast performs more poorly than the target firm's own historical volatility if less than seven comparable firms are used. The main reason for the improvement in forecasting performance (as more comparable firms are used) is the significant reduction in the comparable firms' volatility due to the diversification effect. As the noise in the comparable firms' volatility is reduced, the forecasting efficiency improves. In comparison, the performance of the LM-VAR forecast is only slightly influenced by the number of comparable firms. Even if only one comparable firm is used, the LM-VAR model still provides a more accurate forecast than the target firm's own historical volatility. This is because the VAR setup has the inherent ability to filter out much of the noise in the comparable firms' volatility. As a result, the forecasting performance is only slightly influenced by the number of comparable firms used.

Finally, it is prudent to point out some potential limitations of the LM-VAR model in forecasting long-term volatility. The LM-VAR model imposes nonlinear dynamics on the underlying volatility time series and allows for comovements between volatility series. Its implementation is thus more demanding in model estimation and data requirement than historical-volatility based forecasts do. These technical complications are not insurmountable, however. As suggested by Andersen, Bollerslev, Diebold, and Labys (2003), the LM-VAR model can be estimated by applying an OLS estimation equation by equation under the normalization $\Phi(0) = I$. We use this simplified estimation method in all our empirical tests involving the LM-VAR model. Our results indicate that the simplified estimation method performs quite well and is indeed a reasonable method for estimating the LM-VAR model. This rather simple estimation method makes the LM-VAR forecast accessible to a much wider audience in academia and the business community. Another potential limitation is the data requirement in the estimation of the LM-VAR model. Unlike historical volatility based forecasts, the estimation of the LM-VAR model requires longer time series of data for volatilities of the target firm and the selected common factors. Data limitation can be a problem if the target firm is a recent IPO firm. With little or no past history of stock returns, it is difficult to estimate the LM-VAR model. One possible solution is to boost sample size through bootstrapping. This is reasonable if the available data, though sparse, are representative of the true underlying distribution. If that is not the case or there is simply too little or no data available (e.g., a recent IPO firm), we may have to rely on stock returns from comparable firms to estimate the LM-VAR model. This problem is beyond the scope of the present study and will be left for future research.

VIII. Conclusions

In this paper, we empirically investigate the forecasting performance of historical forecasts in comparison with a number of alternative volatility forecasts. We focus on long-term volatility forecasting and its impact on valuing and expensing stock options. Although previous research has examined the performance of historical forecasts in this context (e.g., Alford and Boatsman (1995)), the focus was merely the comparison of three historical volatility based forecasts (i.e., historical, comparable-firm, and shrinkage forecasts). A key innovation in the present research is to re-evaluate historical volatility based forecasts in a horse race with a number of alternative volatility forecasts that incorporates both the long-memory property of volatility time series and comovements with common factors.

Using a large sample of over 2,000 large, medium and small U.S. firms, we find that historical volatility is a poor forecast for long-term stock return volatility and the shrinkage adjustment only marginally improves its performance. Forecasting performance can be improved substantially, however, if the volatility forecast incorporates both the long-memory property of the volatility time series and comovements with common factors (such as the S&P 500 index volatilities). Our results indicate that the impact of forecasting errors on option expensing is reduced by approximately 50% if the LM-VAR forecast is used instead of the historical forecast. Although volatility forecasting remains a challenging task, our research provides insight on the key factors influencing forecasting performance and how volatility forecasts should be constructed.

Although our research does not directly address managerial incentives to manipulate a firm's estimate of expected stock volatility, we contribute to the literature by providing a more reliable, more accurate volatility forecasting method relative to the commonly used historical volatility method. Since historical volatility is a rather poor forecast for expected stock volatility, it is perhaps too easy or convenient for firms to justify even fairly large deviations from historical forecasts. Given the documented evidence on the opportunistic use of managerial discretion to understate expected volatility, historical forecasts are probably providing too much discretion for management over volatility forecasts. If the LM-VAR forecasting method becomes the new industry standard for volatility forecasting, managers are more likely to refrain from the opportunistic use of discretion and focus more on forecasting accuracy instead.

References

Aboody, D., M.E. Barth, and R. Kasznik. "Do Firms Understate Stock Option-Based Compensation Expense Disclosed under SFAS 123?" *Review of Accounting Studies*, 11 (2006), 429-461.

Alford, A.W., and J.R. Boatsman. "Predicting Long-Term Stock Return Volatility: Implications for Accounting and Valuation of Equity Derivatives." *Accounting Review*,70 (1995), 599-618.

Andersen, T.G., and T. Bollerslev. "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns." *Journal of Finance*, 52 (1997), 975-1005.

Andersen, T.G., T. Bollerslev, P.F. Christoffersen, and F.X. Diebold. "Volatility Forecasting." Forthcoming *Handbook of Economic Forecasting*, Eds: G. Elliot, C.W.J. Granger, and A. Timmermann, Amsterdam: North Holland (2005).

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics*, 61 (2001), 43-76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. "The Distribution of Realized Exchange Rate Volatility." *Journal of American Statistical Association*, 96 (2001), 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. "Modeling and Forecasting Realized Volatility." *Econometrica*, 71 (2003), 579-625.

Andersen, T.G., T. Bollerslev, and S. Lange. "Forecasting Financial Market Volatility: Sample Frequency vis-á-vis Forecast Horizon." *Journal of Empirical Finance*, 6 (1999), 457-477.

Baillie, R., T. Bollerslev, and H.O. Mikkelsen. "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 74 (1996), 3-30.

Balsam, S., H.A. Mozes, and H.A. Newman. "Managing Pro Forma Stock Option Expense under SFAS No.123." *Accounting Horizon*, 17 (2003), 31-45.

Bartov, E., P.S. Mohanram, and D. Nissim. "Stock Option Expense, Forward-Looking Information, and Implied Volatilities of Traded Options." Working paper, *Columbia University* and *New York University* (2004).

Black, F., and M. Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, 81 (1973), 637-659.

Bollerslev, T., and H.O. Mikkelsen. "Modelling and Pricing Long-Memory in Stock Market Volatility." *Journal of Econometrics*, 73 (1996), 151-184.

Bollerslev, T., and H.O. Mikkelsen. "Long-Term Equity Anticipation Securities and Stock Market Volatility Dynamics." *Journal of Econometrics*, 92 (1999), 75-99.

Carpenter, J.N. "The Exercise and Valuation of Executive Stock Options." *Journal of Financial Economics*, 48 (1998), 127-158.

Carpenter, J.N., and B. Remmers. “Executive Stock Option Exercises and Inside Information.” *Journal of Business*, 74 (2001), 513-534.

Christie, A.A. “The Stochastic Behavior of Common Stock Variance: Value, Leverage, and Interest Rate Effects.” *Journal of Financial Economics*, 10 (1982), 407-432.

Christoffersen, P.F., and F.X. Diebold. “How Relevant is Volatility Forecasting for Financial Risk Management.” *Review of Economics and Statistics*, 82 (1999), 12–22.

Ding, Z., C.W.J. Granger, and R.F. Engle. “A Long Memory Property of Stock Returns and a New Model.” *Journal of Empirical Finance*, 1 (1993), 83-106.

French, K., G.W. Schwert, and R.F. Stambaugh. “Expected Stock Returns and Volatility.” *Journal of Financial Economics*, 19 (1987), 3-29.

Geweke, J., and S. Porter-Hudak. “The Estimation and Application of Long Memory Time Series Models.” *Journal of Time Series Analysis*, 4 (1983), 221–238.

Hall, B.J., and J.B. Liebman. “Are CEOs Really Paid Like Bureaucrats?” *Quarterly Journal of Economics*, 113 (1998), 653-691.

Heath, C., S. Huddart, and M. Lang. “Psychological Factors and Stock Option Exercise.” *Quarterly Journal of Economics*, 114 (1999), 601-627.

Hemmer, T., S. Matsunaga, and S. Shevlin. “Optimal Exercise and the Cost of Granting Employee Stock Options with a Reload Provision.” *Journal of Accounting Research*, 36 (1998), 231-255.

Hodder, L., W. Mayew, M.L. McAnally, and C.D. Weaver. “Employee Stock Option Fair-Value Estimates: Do Managerial Discretion and Incentives Explain Accuracy?” Working paper, *Texas A&M University* (2006).

Huddart, S., and M. Lang. “Employee Stock Option Exercises: An Empirical Analysis.” *Journal of Accounting and Economics*, 21 (1996), 5-43.

Huddart, S., and M. Lang. “Information Distribution within Firms: Evidence from Stock Option Exercises.” *Journal of Accounting and Economics*, 34 (2003), 3-31.

Johnston, D. “Managing Stock Option Expense: The Manipulation of Option-Pricing Model Assumptions.” *Contemporary Accounting Research*, 23 (2006), 395-425.

Karolyi, A. “A Bayesian Approach to Modeling Stock Return Volatility for Option Valuation.” *Journal of Financial and Quantitative Analysis*, 28 (1993), 579-594.

Lev, B. “Some Economic Determinants of Time-Series Properties of Earnings.” *Journal of Accounting and Economics*, 5 (1983), 31-48.

Merton, R.C. “The Theory of Rational Option Pricing.” *Bell Journal of Economics and Management Science*, 4 (1973), 141-183.

Merton, R.C. “On Estimating the Expected Return on the Market: An Exploratory Investigation.” *Journal of Financial Economics*, 8 (1980), 323–361.

Pong, S.-H., M.B. Shackleton, S.J. Taylor, and X. Xu. “Forecasting Currency Volatility: A Comparison of Implied Volatilities and AF(FI)MA Models.” *Journal of Banking and Finance*, 28 (2004), 2541-2563.

Poon, S.-H., and C.W.J. Granger. “Forecasting Volatility in Financial Markets: A Review.” *Journal of Economic Literature*, 61 (2003), 478-539.

Poterba, J.M., and L.H. Summers. “The Persistence of Volatility and Stock Market Fluctuations.” *American Economic Review*, 76 (1986), 1142-1151.

Robinson, P.M. “Log-Periodogram Regression of Time Series with Long-Range Dependence.” *Annals of Statistics*, 23 (1995), 1048–1072.

Ross, S.A. “Arbitrage Theory of Capital Asset Pricing.” *Journal of Economic Theory*, 13 (1976), 341-360.

Ross, S.A. “Risk, Return, and Arbitrage.” In *Risk and Return in Finance: I*, eds I. Friend and J.L. Bicksler, Cambridge, MA: Ballinger (1977).

Schwert, G.W. “Why Does Stock Market Volatility Change Over Time?” *Journal of Finance*, 44 (1989), 1115-1153.

Eric Ghysels, Predicting volatility: How to get most out of returns data sampled at different frequencies , (with P.Santa-Clara and R. Valkanov) Journal of Econometrics (2006a) forthcoming)

There is a Risk-Return Tradeoff After All , (with P.Santa-Clara and R. Valkanov) - Journal of Financial Economics (2006b) forthcoming)

A Simple Long Memory Model of Realized Volatility Fulvio Corsi 10th July 2004

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2006), ”Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility,” Review of Economics and Statistics, forthcoming.