Forecasting Volatility

Louis H. Ederington
University of Oklahoma

Wei Guan
University of South Florida St. Petersburg

January 2004
Initial Draft: June 1998

Contact Info: Louis Ederington: Finance Division, Michael F. Price College of Business, University of Oklahoma, 205A Adams Hall, Norman, OK 73019, USA. Phone: (405) 325-5591
Wei Guan: College of Business, University of South Florida St. Petersburg, 140 Seventh Avenue, St. Petersburg, FL 33701, USA. Phone: (727)-553-4945

Comments may be sent to the authors at lederington@ou.edu or wguan@stpt.usf.edu. We have benefitted from comments on earlier versions of this paper by Tim Bollerslev, Steve Figlewski, Michael Fleming, Charles Jones, Jose Lopez, and seminar participants at the Federal Reserve Bank of New York, Queensland University, and Australian National University. Remaining errors are of course our responsibility.

ABSTRACT

Forecasting Volatility

The paper explores the forecasting ability of popular volatility forecasting models focusing on four issues: 1) the relative weighting of recent versus older observations, 2) the emphasis placed on large shocks, 3) the estimation criteria, and 4) the trade-off in terms of forecasting error between simple and complex models. Like previous studies we find evidence that financial markets have longer memories that reflected in GARCH(1,1) model estimates. However, we find that the major problem is not with the exponential model but with the GARCH estimation procedure in that regression estimates of the GARCH model parameters imply considerably longer memories. While more complex models which allow a more flexible weighting pattern than the exponential model forecast better on an in-sample basis, due to the additional estimation error introduced by additional parameters, they forecast poorly out-of-sample. We find that models based on absolute return deviations generally forecast volatility better than otherwise equivalent models based on squared return deviations - though not for GARCH models. Among the most popular time series models, we find that GARCH(1,1) generally yields better forecasts than the historical standard deviation and exponentially weighted moving average models but between GARCH and EGARCH there is no clear favorite. However, in terms of forecast accuracy, all are dominated by a simple non-linear least squares model (developed here) based on historical absolute return deviations.

Forecasting Volatility

Accurate volatility forecasts are important to traders, investors, and risk managers, as well as researchers seeking to understand market dynamics. For example, estimates of future volatility are critical inputs in both option pricing models and value-at-risk models. Such volatility forecasts may be obtained from either time-series models or implied volatilities calculated from observed option prices. Although theoretically implied volatilities should reflect all available information, including time-series information, evidence is mixed on which of the two forecasts better. Moreover implied volatilities cannot simultaneously be used to price the derivative assets from whose prices they are calculated and are only available for specific time horizons for a limited set of assets. Consequently time-series models are the major source of volatility forecasts.

This paper examines the forecasting ability of popular time-series volatility forecasting models and suggests alternatives. The econometrics literature is replete with studies comparing the forecasting ability of various time-series models.[1] For instance, Poon and Granger (2003) list 39 studies comparing the out-of-sample forecasting abilities of the GARCH(1,1) model and the historical variance.[2] However, our approach differs from most previous studies in that we seek to determine why some models forecast better than others exploring four issues: 1) the relative weighting of recent versus older observations, 2) the emphasis placed on large shocks, 3) the importance of the estimation criteria, and 4) the trade-off in terms of forecasting error between simple, but possibly incomplete, forecasting models and more complex models which may be more realistic but add estimation error. While the first issue, the proper weighting of recent versus older observations, has received considerable attention in the literature, the other three have not.

We choose models, data series, and forecast horizons which are in common usage. For instance, while many extant studies compare the ability of various models to forecast volatility in the only the next period, e.g. the next day if the model was estimated using daily data, most uses of volatility forecasts, such as for option pricing and value-at-risk models, are for much longer horizons. Accordingly, we compare the ability of the models to forecast volatility over horizons of 10, 20, 40, 80, and 120 trading days. We also restrict our focus to commonly used volatility forecasting models, such as the historical standard deviation, the GARCH(1,1) model, Riskmetrics's exponentially weighted moving average model, and on alternatives which can be easily implemented using standard statistics software. We base our forecasts on daily data since those are the data sets available to most forecasters.[3] To ensure generality, we compare forecasting ability across a variety of markets: the S&P 500 Index, the Deutschmark/Dollar exchange rate, the 3-month Eurodollar rate, the 10-year Treasury Bond rate, and five equities: Boeing, Exxon, International Paper, 3M, and McDonald's.

One issue explored in the paper is the proper weighting of recent versus older observations. Like others, we find evidence that GARCH(1,1) puts too much weight on recent observations relative to older observations. However, despite the attention this issue has received, we find that out-of-sample forecast accuracy is fairly insensitive to the weighting scheme.

A second issue is which estimation criterion yields the best forecast. For instance, we construct a non-linear least squares regression model which is structurally identical to a GARCH(1,1) model but is estimated using a two stage least squares procedure rather than maximum likelihood. We find that the parameters estimated using least squares imply considerably longer memories that the GARCH model and generate better forecasts in-sample. However, out-of-sample results are mixed.

Third, we explore the trade-off between model complexity and forecast error. A number of models, including GARCH(1,1), impose a functional form in which the weights attached to squared return observations decline exponentially as one moves back in time. We find that a model in which the depreciation rate slows as one moves back in time improves in-sample forecast accuracy.

However, because this more flexible model involves additional parameters, estimation error is increased and the more complicated model does not generally forecast as well out-of-sample.

A fourth issue is whether the popular models attach too much weight to large shocks which may or may not reoccur. For instance, since both the historical standard deviation and the GARCH(1,1) forecasts are functions of the squared surprise returns, a single large return deviation has a large impact on the volatility forecast. Except within the GARCH family, we find that models based on absolute return innovations generally forecast better than otherwise equivalent models based on squared deviations.

While no one model forecasts best in all markets in all circumstances, we find that a simple non-linear least squares model based on past absolute return innovations, which we develop, forecasts best in most markets at most horizons than the methods currently in use.

The remainder of the paper is organized as follows. In the next section, we explore what we term linear squared deviation (or LSD) models focusing on the three most widely used volatility forecasting models: the historical standard deviation, the exponentially weighted moving average model and the GARCH(1,1) model. We also develop a non-linear least squares models which is structurally identical to the GARCH(1,1) model but estimated differently. Our data sets and estimation procedures are described in section II. In section III, we explore the question of the optimal weighting of more recent as opposed to older observations and the appropriateness of the exponential weights structure common most popular models. Out-of-sample forecasting ability is compared in IV. In section V, we turn to our fourth issue and compare models based on absolute return deviations to those based on squared deviations. Section VI concludes the paper.

I. Linear Squared Deviation Models of Financial Market Volatility

While countless time-series volatility forecasting models have been proposed by econometricians, in terms of usage by market professionals and textbook attention, three clearly dominate: (1) the sample variance or standard deviation of returns calculated over some recent period, (2) an exponentially weighted moving average of the squared surprise returns and (3) Bollerslev's GARCH(1,1) model. All three belong to what might be called the linear squared deviation (LSD) class of estimators in that the forecast variance is a linear combination of the squared deviations of recent returns from their expected value. In the case of the historical variance, each squared deviation (or observation) back to a chosen cutoff date is weighted equally while observations prior to the chosen cutoff receive a zero weight. In the exponentially weighted moving average and GARCH (1,1) models, the weights decline exponentially. In other words, the weight attached to observation t-(j+1) is a fixed proportion, β, of the weight attached to observation t-j. We consider the three in turn.

A. The Historical Variance.

Letting $R_t = \ln(P_t/P_{t-1})$ represent daily returns on a financial asset,[4] the simplest forecast of the volatility of R_t over the future period from time t+1 through t+s is the sample standard deviation or variance of returns from the recent past. Derivatives textbooks commonly recommend this procedure. We calculate the historical variance, VAR(n), over the n day historical period:

$$\mathrm{VAR}(n)_t = \frac{1}{n}\sum_{j=0}^{n-1} r_{t-j}^2 \qquad \textbf{(1)}$$

where $r_{t-j} = R_{t-j} - \mu$ and μ is the expected return.[5] This estimator assigns each squared return deviation, r_{t-j}^2, after time t-n a weight of 1/n while observations before t-n receive a weight of zero.

An obvious issue in applying this procedure is choosing the cutoff date n. While 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 much longer period. Accordingly, we consider a variety of sample period lengths.

B. Exponentially Weighted Moving Average.

The exponentially weighted moving average (EWMA) forecasts volatility for the next day as

$$v_{t+1} = \frac{1}{\Gamma} \sum_{j=0}^{J} \beta^{j} r_{t-j}^{2} \tag{2}$$

where $\Gamma = \sum_{J=0}^{J} \beta^{j}$. An obvious issue is the value for exponentially declining weight β. By far the most well known user of the EWMA is Riskmetrics which uses it for its Value-at-risk model. Since Riskmetrics sets β=.94 (see Riskmetrics, 1996), we use this parameter for our estimations. It is easily shown by successive substitution in equation 2 that the implied forecast for a future time period t+k is identical to that for period t+1, i.e., $v_{t+k} = v_{t+1}$ so the EWMA forecast of average volatility over an interval from t to t+s, which we label $EWMA_t$, is

$$EWMA_{t} = \frac{1}{\Gamma} \sum_{j=0}^{J} \beta^{j} r_{t-j}^{2} \tag{3}$$

C. The GARCH(1,1) model.

The GARCH(1,1) model is similar to the EWMA model but adds a mean reversion term. The GARCH(1,1) model assumes that the log-return at time t, R_t, is normally distributed with mean μ and variance v_t and that v_t follows the process:

$$v_{t+1} = \alpha_0 + \alpha_1 r_t^2 + \beta v_t \quad \textbf{(4)}$$

Since $v_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta v_{t-1}$, $v_{t+1} = (\alpha_0 + \beta\alpha_0) + \alpha_1 r_t^2 + \beta\alpha_1 r_{t-1}^2 + \beta^2 v_{t-1}$, successive substitution back to time t-j yields the alternative expression of the GARCH(1,1) model:

$$v_{t+1} = \alpha_0' + \alpha_1 \sum_{j=0}^{J} \beta^j r_{t-j}^2 \quad \textbf{(5)}$$

where $\alpha_0' = \alpha_0 \sum_{j=0}^{J} \beta^j + \beta^{J+1} v_{t-J}$. As equation 5 makes clear, in the GARCH(1,1) forecast of the variance at time t+1, the squared return deviation at time t receives the weight, α_1, the squared deviation at time t-1 receives a weight of $\alpha_1\beta$, and (assuming $\beta<1$) the weights decline exponentially. Since $E_t(r_{t+1}^2) = v_{t+1}$, successive forward substitution yields the expression for the expected volatility at a future time t+k based on the information available at time t:

$$v_{t+k} = \alpha_0 \sum_{j=0}^{k-2} (\alpha_1+\beta)^j + (\alpha_1+\beta)^{k-1} v_{t+1} \quad \textbf{(6)}$$

If $(\alpha_1 + \beta)<1$, the k-step ahead volatility forecast declines from v_{t+1} toward the unconditional variance at the rate $(\alpha_1+\beta)$ as k increases while the weight attached to past observations declines at the rate β. The forecast volatility over the future period from t+1 through t+s, which we label "$GARCH_t$", is an average of the volatility expected each day from t to t+s hence:

$$GARCH_t = (1/s)\sum_{k=1}^{s} v_{t+k} = \alpha + \lambda \sum_{j=0}^{J} \beta^j r_{t-j}^2 \quad \textbf{(7)}$$

where $\alpha = (1/s)\sum_{k=1}^{s}\left[\alpha_0 \sum_{j=0}^{k-2}(\alpha_1+\beta)^j + \alpha_0'(\alpha_1+\beta)^{k-1}\right]$ and $\lambda = (\alpha_1/s)\sum_{k=1}^{s}(\alpha_1+\beta)^{k-1}$. Note that, in the GARCH volatility forecast for the interval from t+1 to t+s, the weights attached to successive past observations decline exponentially.

D. Previous Comparisons.

While numerous studies have compared the forecasting abilities of the historical variance and GARCH models, no clear winner has emerged. In an thorough review of 39 such studies, Poon and Granger (2003) find that 22 find that historical volatility forecasts implied volatility better out-of-sample while 17 find that GARCH models (not necessarily GARCH(1,1)) forecast better.

E. The Restricted Least Squares Estimator and the Estimation Criterion.

Several past studies, such as Pagan and Schwert (1990), West and Cho (1995) and Lopez (2001), have considered the forecasting ability of linear regression models of the form: $v_{t+1} = \alpha_0 + \sum_{j=0}^{J} \alpha_j r_{t-j}^2$. These models have several problems: (1) the estimated coefficients, α_j usually don't decline in an orderly fashion as j increases,[6] (2) they only incorporate a few recent observations (commonly 8 to 15), and (3) (and relatedly) they don't forecast out-of-sample very well. We specify a non-linear OLS model which avoids these problems.

While the GARCH(1,1) estimates of the parameters α, λ, and β in equation 7 are obtained by first estimating the parameters α_0, α_1, and β in equation 4 using maximum likelihood, it is also possible to estimate the parameters in equation 7 using non-linear least squares. Letting $AV(s)_t$ represent the actual realized variance over the period from t+1 through t+s. i.e., $AV(s)_t = (1/s)\sum_{i=1}^{s} r_{t+i}^2$, α, λ, and β can be estimated by applying least squares to the equation pair:

$$AV(s)_t = \alpha + \lambda Z_t \quad \text{where} \quad Z_t = \sum_{j=0}^{J} \beta^j r_{t-j}^2 \qquad \textbf{(8)}$$

We label the resulting estimates of equation 8 the "restricted least squares" (or RLS) model. It is "restricted" in that the coefficients of r^2_{t-j} are forced to decline exponentially as j increases and non-linear due to the β^j term . Once the parameters α, β, and λ have been estimated, the RLS forecast of volatility from t+1 to t+s is generated as:

$$RLS_t = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{J} \hat{\beta}^j r_{t-j}^2 \quad \textbf{(9)}$$

Note that equations 7 and 9 are structurally identical; that is the GARCH and RLS forecasts only differ in that they employ different estimates for α, β, and λ. While GARCH chooses the parameter estimates which maximize the likelihood of observing r_t, RLS chooses parameters which minimize the sum of squared errors of $AV(s)_t$. Consequently, the question arises whether the estimated parameters differ substantially and, if so, which forecast better.

II. Data and Procedures.

We compare the forecasting ability of these four models for four financial series: the S&P 500 Index, the Deutschmark/Dollar exchange rate, the 3-month Eurodollar rate, the 10-year Treasury Bond rate, and five equities chosen from those in the Dow-Jones Index: Boeing, Exxon, International Paper, 3M, and McDonald's. Daily price data for the five equities for the period 7/2/62 to 12/30/94 were obtained from CRSP tapes as were prices for the S&P 500 index (7/2/62-12/29/95). Daily interest rate and exchange rate data were obtained from Federal Reserve Board files for the periods: 1/2/62-6/13/97 for the 10 year bond rate, 1/1/73-6/20/97 for the Eurodollar rate, and 1/1/71-6/30/97 for the Deutschemark/Dollar exchange rate. Daily log returns are defined as $R_t = \ln(P_t/P_{t-1})$ and daily return deviations or innovations are defined as $r_t = R_t - \mu$ where μ is measured as the mean of R_t over the entire data period.

Our primary measure of forecasting ability is the root mean squared forecast error (RMSFE) measured in terms of the difference between actual and forecast annualized standard deviation of returns. For each forecast period beginning on day t+1 and extending through day t+s, we calculate the actual annualized standard deviation of returns as $AS(s)_t = [252\sum_{j=1}^{s} \frac{r_{t+j}^2}{s}]^{1/2}$. The root mean squared forecast error is then measured as

$$RMSFE = [\,(1/M)\sum_{m=1}^{M} (AS(s)_m - FSTD(s)_m)^2\,]^{1/2} \quad \textbf{(10)}$$

where $FSTD(s)_m$ is the forecast standard deviation (also annualized) for an s day horizon beginning on day m using one of the four forecasting procedures outlined in section I. M represents the number of forecast periods as reported in the tables below.[7] One issue addressed by Poon and Granger (2003) is whether volatility forecast errors are best measured in terms of the standard deviation or variance. As they point out, when the RMSFE is measured in terms of the variance, a few outliers tend to dominate the results. In addition, derivative prices are roughly proportional to the standard deviation. Consequently we define the RMSFE (and mean absolute deviation) in terms of the standard deviation. In Appendix B, we also report comparisons based on mean absolute forecast errors. Both metrics yield similar conclusions.

III. Estimation and Weighting Issues.

A. In-sample Comparisons of Popular Models.

To explore why the models forecast differently it is helpful to first consider in-sample results before turning to the more important out-of-sample forecasts. In-sample RMSFEs for the historical standard deviation, EWMA, GARCH(1,1), and RLS models are reported in the first eight rows in Table 1 for a forecast horizon of 40 trading days.[8] For comparison, the annualized standard deviation of the ex-post standard deviation, AS(s), which represents the RMSFE for a naive forecast equal to mean volatility over the entire data period, is shown in the last row of Table 1. istorical standard deviations over past periods of 10, 20, 40, 80, and 120 days, are reported in the first rows as STD(10), STD(20), ...,STD(120) respectively. For each data series, the STD(n) with the lowest RMSFE is shown in bold. Note that in most markets, the RMSFE for a forecast horizon of 40 days is minimized using the standard deviation over the last 120 days.

Comparing RMSFEs for (1) GARCH(1,1) model, (2) the lowest of the five RMSFEs for the historical standard deviations, and (3) EWMA we find that the GARCH forecast has the smallest in-sample RMSFE in eight of our nine markets, the single exception being the volatility of the 3-month Eurodollar rate. On average GARCH's RMSFE is 5.8% lower than that of STD(120).

Results for the RLS model are reported in row 8. To facilitate comparison, the cell in each column (market) with the lowest RMSFE among all the models in Table 1 is shaded. As this shading makes clear, in all nine markets, the RLS model has the lowest in-sample RMSFE among the four (eight if we consider the STD models separately) models we examine. This result holds for all other forecasting horizons except the 10 day when GARCH is best in one market and holds if the mean absolute forecast error is used as the comparison metric instead of the root mean squared forecast error.

B. The Estimation Question.

Since the RLS model is structurally identical to the GARCH(1,1) model, the question arises as to why it should consistently out-forecast the GARCH model on an in-sample basis. The answer is that the different estimation methods yield quite different parameter estimates. Specifically, the GARCH model chooses parameters which maximize the likelihood of observing the observed sample of returns while the RLS chooses parameters which minimize the variance of the forecast error.[9]

How and how much the parameters estimates differ is shown in Table 2 where we report the two models' estimates of the parameters β, α, and λ parameters of equation 7. For GARCH, we first report the standard GARCH(1,1) parameter estimates, α_0, α_1, and β of equation 4 and then the implied parameters α, λ, (and β) of equation 7. The differences between the GARCH and RLS estimates of α, λ, and β are substantial and strikingly consistent. In all nine markets, $\beta_{RLS} > \beta_{GARCH}$, and $\lambda_{RLS} < \lambda_{GARCH}$. This means that as compared with the GARCH model (also the EWMA model as used by Riskmetrics), the restricted least squares model places considerably less weight on the most recent observations and more weight on observations in the more distant past. For instance, in T-Bond market, as illustrated in Figure 1, the RLS model weights observations on days t-32 through t less than GARCH and weighs observations on day t-33 and earlier more heavily. Other implied "cross-over" days are t-32 in the Eurodollar market, day t-39 in the S&P 500 index market, and day t-18 in Dollar/Deutschmark market. The implication is that the GARCH(1,1) procedure (also the EWMA model with the Riskmetrics parameters) yields parameter estimates which over-weight recent observations and underweight older observations. In other words, in-sample forecast accuracy is increased by reducing the weights GARCH attaches to the most recent observations and increasing those attached to older observations.

Our finding that the GARCH model puts too much weight on recent observations relative to those in the past is consistent with prior evidence showing that asset market volatility has a long memory. Consequently, the question arises whether the problem is solely with the GARCH(1,1) parameter estimates or with the exponential model itself. To test the appropriateness of the exponential model, we estimate a more flexible model which nests exponential weights as a special case. As compared with switching to a completely different model, examining a more general model which nests the exponential model, allows us to explore how and how much the weight pattern that maximizes forecast accuracy differs from the exponentially declining weights model.

One distributed lag form meeting these requirements is an adaptation of Schmidt's (1974) model, which combines the Koyck and Almon lag forms. Our Schmidt-type model of the variance over a future period s is:

$$GEN_t = \alpha'' + \sum_{i=0}^{I} \lambda_i Z_{it} \quad \text{where} \quad Z_{it} = \sum_{j=0}^{J} \beta^j j^i r_{t-j}^2 \tag{11}$$

where I=2.[10] Like the RLS model, this model, which we label GEN due to its flexible general form, is estimated using non-linear least squares by regressing the ex-post variance on Z_{it}'s defined for various values of β. Note that $Z_{0t} = \sum_{j=0}^{J} \beta^j r_{t-j}^2$ so if $\lambda_i=0$ for i>0, the GEN model is identical to the RLS model. Consequently, we can test the appropriateness of the exponential weights assumption by testing whether $\lambda_i=0$ for all i>0.

As reported in Table 3, the null hypothesis that $\lambda_1=\lambda_2=0$ (i.e. that the weights decline exponentially) is clearly rejected at the .01 level in all nine markets. Moreover, the pattern is remarkably consistent across all nine markets in that $\lambda_1<0$ and $\lambda_2>0$ (and in that the λ_0 estimated for the GEN model exceeds the RLS λ). The implication is that an exponential lag structure (in which the weights depreciate at the constant rate β), forces the weights to deprecate too slowly across the most recent observations and too rapidly later.[11] This is illustrated for two of our markets in Figures 2a and 2b where the implied coefficients of r_{t-j}^2 are plotted for j=0 to 200 according to both the RLS and GEN models.

However, as illustrated in Figures 2a and 2b, the difference in the coefficient lag structure between the RLS and GEN model estimates (that is between the more restrictive and flexible regression forms), is less than that between the GARCH and RLS estimates of the same model. In all nine markets, the GARCH model's parameter estimates put much more weight on the most recent observations and less on older observations than either RLS or GEN.[12]

In summary, we like others find evidence that the GARCH(1,1) model puts too much weight on the most recent observations and not enough on observations in the more distant past. However, in contrast to previous studies which fault the exponential model, we find that much of the fault lies with the estimation procedure. We find that a regression estimation of the exponential model results in parameters which put more weight on the more distant observations. Also, it should be noted that despite the considerable difference in the coefficient lag structure, the differences in forecast accuracy are not that great. Comparing GEN with GARCH, the improvement in the RMSFE ranges from .4% (3M) to 9.2% (Eurodollars) and averages only 3.1%.

IV. Out-of-Sample Forecasting Ability.

Next, we compare the five models' out-of-sample forecasting ability.[13] To generate out-of-sample forecasts, each model is estimated using 1260 daily return observations - approximately five years of daily data. To limit the computational burden, the models are re-estimated every 40 days.[14] The resulting root mean squared forecast errors for a 40 day forecasting horizon are reported in Table 4 while results for other time horizons are reported in the Appendix.

When we compared forecasting ability in-sample in the previous section we observed that GARCH consistently dominated the historical standard deviation, RLS consistently dominated GARCH, and GEN dominated RLS. As shown in Table 4, these relationships no longer hold when out-of-sample forecasts are compared. The RLS and GARCH procedures forecast somewhat better than the other three models in most markets but the RMSFE differences are generally small and insignificant. Between RLS and GARCH there is no clear winner. In the four macro markets, RLS has the lowest RMSFE in both the T-Bond and Deutschmark markets, GARCH's RMSFE is lowest in the S&P 500 market, while the RMSFE of the historical standard deviation based on observations over the last 120 days is lowest in the Eurodollar market. None of the RMSFE differences are significant at a .05 level. In the five equity markets, GARCH has the lowest RMSFE in four (two of which are significant at the .05 level) and RLS in one (Boeing). In summary, on an out-of-sample basis, no one procedure clearly dominates although GARCH and RLS generally forecast better than the STD, EWMA, and GEN models. As shown in the Appendix A, the forecast horizon makes little difference though there appears to be a tendency for STD to forecast better and for RLS to do less well as the horizon increases.

The GARCH(1,1) model was not always well-behaved in our estimations. In six of our nine markets, $\hat{\beta} = 0$ (the lower bound) in some sub-samples. In other words, replacing only 40 out of 1260 observations would occasionally cause $\hat{\beta}$ to suddenly drop from .8 or .9 to zero and replacing a different 40 observations would cause $\hat{\beta}$ to return to .8 or .9. In most markets, this only happened occasionally, e.g., in 2 out of 190 sub-samples in the T-Bond market and 2 out of 126 in the Eurodollar market, but there was at least one such occurrence in six markets and for 3M there were 16 (out of 173). In the individual equity estimations, there were also numerous times when $\hat{\beta} < .4$ implying that only the most recent observations have any information content. In contrast, the RLS parameter estimates were much more stable across samples. In only one sub-sample in one market (Eurodollar) was the estimated β at our lower bound of .505.

While the GEN model consistently dominated the others in terms of in-sample forecast accuracy, this is certainly not the case when out-of-sample forecasting ability is compared. In all nine markets, both RLS and GARCH have lower out-of-sample RMSFEs. The inability of the GEN model to forecast very well out-of-sample illustrates the cost of added complexity underscoring the argument of Dimson and Marsh (1990). As econometrics texts commonly point out, if the true model is $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + u$ but one instead estimates $Y = \beta'_0 + \beta'_1 Z_1 + u'$, $\hat{\beta}'_1$ will be a biased estimator of β_1 (unless Z_1 and Z_2 are orthogonal) but will have a smaller variance than $\hat{\beta}_1$. Since the variance of the forecast error is a function of both, if Z_2 is a relatively minor determinant of Y, better forecasts may often be obtained with the simpler, though biased, model. . While the GEN model corrects the tendency for the RLS model to under-weigh recent and older observations relative to those in the middle, by adding two additional parameters, more estimation error is introduced leading to a deterioration in out-of-sample forecasting ability.

In summary, we find that the exponential model using parameters estimated using either the GARCH or regression procedures (RLS) forecasts better than either the historical standard deviation, the exponentially weighted moving average model, or a more general model that nests the exponential model. The RLS estimation of the exponential model results in parameter estimates which place less weight on the most recent observations and more on older observations than the GARCH estimation procedure and also results in more stable parameter estimates. Nonetheless, the forecasting record of the RLS and GARCH(1,1) models is roughly the same. A more flexible model that nests the exponential model forecasts somewhat better than the RLS and GARCH(1,1) models on an in-sample basis indicating the exponential form may not be the most appropriate. However, since this model involves more parameters, estimation error is increased and it forecasts consistently worse out-of-sample.

V. The Impact of Large Return Surprises.

To this point, all the models which we have considered belong to what we have termed LSD models, that is, volatility forecasts based on linear combinations of past squared return deviations. The fact that all are based on squared return deviations means that the forecasts are quite sensitive to big outliers. An obvious question, but one which to our knowledge has not been explored heretofore, is whether better forecasts can be obtained using models which are less sensitive to a single highly volatile day, such as models based on absolute return deviations.

A. A Least Squares Exponential Model Based on Absolute Return Deviations.

To explore whether models based on absolute return observations forecast better than equivalent forms based on squared return deviations, we first develop an exponential model similar to that of the RLS and GARCH models but specified in terms of absolute, rather than squared, past return observations. Unlike the models based on squared return deviations those based on absolute returns require a return distribution assumption. If log returns R_t=ln(P/P_{t-1}) are normally distributed with mean μ, then $E(|r_t|) = \sigma\sqrt{2/\pi}$ where $r_t = R_t$-, and $E\left(\sqrt{\pi/2}\sum_{j=0}^{n-1} W_j\,|r_{t-j}|\right) = \sigma$ where $\sum W_j$ =0. Based on this we define a regression model with exponentially declining weights analogous to RLS but defined in terms of absolute, rather that squared, return deviations:

$$AS(s)_t = \alpha + \lambda W_t \quad \text{where} \quad W_t = \sqrt{\pi/2}\sum_{j=0}^{J} \beta^j\,|r_{t-j}| \qquad \textbf{(12)}$$

where $AS(s)_t$ is the realized standard deviation of returns over the s day period following day t. We refer to this as the absolute restricted least squares model or A-RLS Note the similarities and differences between equations 12 and 8. In estimating the RLS model of equation 8, we regressed the ex post variance on Z_t's defined using squared return innovations. In 12, we regress the ex post standard deviation on W_t's defined in terms of absolute return innovations. In both the weights decline exponentially. Again, we first generate the series $W(\beta)_t$ using values of β from .500 through 1.000 in increments of .005, then regress $AS(s)_t$ on $W(\beta)_t$ using OLS, repeat the regression for all values of β, and choose the values of β, α, and λ for the regression resulting in the lowest residual sum of squares. The resulting parameter estimates for each of the nine markets are reported in Table 6. Once the parameters are determined, we then generate the forecasts:

$$A\text{-}RLS_t = \hat{\alpha} + \hat{\lambda}\sqrt{\pi/2}\sum_{j=0}^{J} \hat{\beta}^j\,|r_{t-j}| \qquad \textbf{(13)}$$

where J=200. Out-of-sample RMSFEs for this forecasting model are reported in Table 5 where the RMSFEs for the RLS model are repeated for comparison. The lower of the two is in boldface. In eight of the nine markets , A-RLS's RMSFE is lower than that of the RLS model; Exxon where the two RMSFEs are virtually identical is the exception. As reported in Panel B, in five of the nine markets, the null that RLS and A-RLS yield equally accurate out-of-sample forecasts is rejected at the .05 level. A-RLS's out-of-sample RMSFE averages about 6.7% below RLS's across our nine markets - a fairly substantial improvement. As shown in the Appendix, A-RLS's dominance over RLS holds at all forecasting horizons.

B. AGARCH and EGARCH Models.

Next attention is turned to two models in the GARCH family: AGARCH since our analysis to this point indicates that models based on absolute return deviations out-perform similar models based on squared return deviations and EGARCH, which we consider because it is second to the GARCH(1,1) model in popularity. Analogous to the GARCH(1,1) model's assumption in terms of the variance, the AGARCH model assumes the standard deviation, S, of log-normally distributed returns evolves following the process:

$$S_{t+1} = \alpha_0 + \alpha_1 [\sqrt{\pi/2}\,|r_t|] + \beta S_t \qquad \textbf{(14)}$$

Successive backward substitution yields the expression, $S_{t+1} = \alpha_0' + \alpha_1 [\sqrt{\pi/2} \sum_{j=0}^{J} \beta^j |r_{t-j}|]$, while forward substitution yields, $S_{t+k} = \alpha_0 \sum_{j=0}^{k-2} (\alpha_1 + \beta)^j + (\alpha_1 + \beta)^{k-1} S_{t+1}$. The AGARCH forecast of the standard deviation of returns over the period from t+1 to t+s is then:

$$AGARCH_t = \sqrt{(1/s) \sum_{k=1}^{s} E(S_{t+k}^2)} \qquad \textbf{(15)}$$

If daily returns are independent, the variance of returns over the s day period is the sum of the variance of returns each day. Hence, while the GARCH forecast could be expressed as a linear function of past squared return innovations, the expression for the AGARCH forecast in equation 15 cannot be rewritten as a linear function of past absolute return innovations. This also means that while the RLS and GARCH forecasts were structurally identical, A-RLS and AGARCH are not.

The final model which we consider is the EGARCH model of Nelson (1991) which assumes that the variance of the log-return, v_t, follows the process:

$$\ln(v_{t+1}) = \omega + \beta \ln(v_t) + \gamma \left[\frac{|r_t|}{\sqrt{v_t}} - \left(\frac{2}{\pi} \right)^{.5} \right] + \gamma\theta \frac{r_t}{\sqrt{v_t}} \quad \textbf{(16)}$$

The most well-known property of the EGARCH model is that (through the last right-hand side term) it, unlike GARCH, allows equal positive and negative shocks to have different impacts on the conditional variance. Equally important for our purposes is the fact that the conditional variance is modeled in terms of absolute, instead of squared, return innovations. However, since the left land side variable is the log of the conditional variance, EGARCH's properties are quite different from AGARCH's. While extreme observations have less impact on the volatility forecast in AGARCH than in GARCH, EGARCH tends to place more weight on extreme observations.

Note that $E_t[\ln(v_{t+2})] = \omega + \beta E_t[\ln(v_{t+1})]$ since the step-head values of both of the last two terms in equation 16 are zero. So for k>1,

$$E_t[\ln(v_{t+k})] = \omega \sum_{j=0}^{k-2} \beta^j + \beta^{k-1} E_t[\ln(v_{t+1})] \quad \textbf{(17)}$$

To estimate volatility over the next s days, we first estimate equation 16 using maximum likelihood. Then, we forecast the variance every day from day t+1 through day t+s using equations 16 and 17 and average the v_{t+k}'s for all s days to obtain the EGARCH forecast. In the S&P 500 and individual equity estimations $\hat{\theta}<0$, confirming the common finding that in equity markets negative shocks tend to increase the conditional variance more than equivalent positive shocks but (while significant) its explanatory value is low. The estimated theta is also negative in the DMark market but positive in the two interest rate markets.

As shown in Table 5, where the lowest of the three GARCH model RMSFE's is in boldface, in terms of out-of-sample forecasting accuracy, it is difficult to choose between GARCH and EGARCH but both tend to dominate AGARCH. EGARCH has the lowest RMSFE in five markets and GARCH(1,1) in four.[15] Pairwise, GARCH dominates AGARCH in every market and EGARCH dominates AGARCH in all save International Paper. While we found that the regression model based on absolute return deviations forecast better than equivalent model based on squared deviations, that clearly does not hold true for GARCH type models. Interestingly, the results in Appendix A indicate that GARCH forecasts somewhat better at the longer horizons and EGARCH at the shorter horizons.[16]

As with GARCH, the EGARCH and AGARCH estimations were not always well-behaved in the 1260 observation sub-samples. For the EGARCH model, in seven of the nine markets, $\hat{\beta}<0$ in some sub-samples and in six markets $\hat{\beta}>1$ in some sub-samples. In nine sub-samples in the T-Bond market, $\hat{\gamma}<0$ implying that higher than normal volatility tends to be followed by lower than normal volatility. Similar problems were observed with AGARCH.

C. And the Winner Is ...

In each column in Table 5, we shade the cell with the lowest RMSFE for that market among all the forecasting models which we consider. The results are fairly dramatic. In six of the nine markets, the RMSFE of the A-RLS model is the lowest of the ten forecasting models. GARCH's RMSFE is lowest in two equity markets and STD(120) in one. Across the nine markets, A-RLS's RMSFE averages about 3.8% below GARCH's and 3.8% below EGARCH's. It beats GARCH in seven of nine markets and in four of those the null that the A-RLS and GARCH out-of-sample forecasts are equally accurate is rejected at the .05 level. A-RLS's out-of-sample RMSFE is lower than EGARCH's in eight of the nine markets and in four the null of equal accuracy is rejected at the .05 level. As reported in the Appendix, when forecast errors are measured in terms of mean absolute error, A-RLS's forecasting dominance is even more complete. Interestingly, as shown in the Appendix, at the shorter horizons of 10 and 20 days, A-RLS consistently dominates the other procedures in terms of out-of-sample RMSFE but as the length of the forecast horizon expands, this dominance declines and at the longest horizons no one procedure is clearly best.

D. Weights reconsidered.

Estimated parameters for the RLS and A-RLS models are reported in Tables 2 and 6 respectively. In every market $\lambda_{A\text{-}RLS} > \lambda_{RLS}$ and $\beta_{A\_RLS} < \beta_{RLS}$. In other words, the A-RLS model consistently assigns greater weight to the most recent observations and less weight to older observations in forming its forecasts than the RLS model. Apparently, once the impact of outliers on the volatility forecast is reduced by switching from squared to absolute return deviations, the model no longer needs to spread the weights over a large number of past observations. It should be noted however that even though A-RLS attaches more weight to the most recent observations and less to older observations than RLS, it still attaches less weight to the most recent observations and more to the older than GARCH(1,1).

Given the apparent forecasting superiority of the A-RLS model over the other models and the difference in its parameters versus RLS, the question again arises whether the exponentially declining weight structure in A-RLS is appropriate. To explore this issue, we again estimate a more general model which nests the A-RLS model as a special case. To be specific, we define the terms $W(\beta)_{it} = \sqrt{\pi/2}\sum_{j=0}^{J} j^{i}\beta^{j}\,|r_{t-j}|$ for i=0,1,and 2. The ex post standard deviation, $AS(s)_t$ is then regressed on $W(\beta)_{0t}$, $W(\beta)_{1t}$, and $W(\beta)_{2t}$. If the coefficients of $W(\beta)_{1t}$, and $W(\beta)_{2t}$ (again designated as λ_1 and λ_2) equal zero, the model is equivalent to A-RLS. Parameter estimates for both the A-RLS and A-GEN models are reported in Table 6. The parameter pattern for A-GEN is consistent across all nine markets and identical to the pattern observed in Table 3 for the estimators based on squared return deviations in that $\hat{\lambda}_1<0$, and $\hat{\lambda}_2>0$. In eight markets $\hat{\lambda}_0$ exceeds the λ estimated for the A-RLS model. The null that $\lambda_1=\lambda_2=0$ is rejected at the .01 level in all nine markets. This coefficient pattern implies that the geometric lag structure embodied in the A-RLS model tends to under-weight the youngest and oldest observations and over-weight those in-between. However the difference in the in-sample RMSFEs between A-RLS and A-GEN is small and the null that the two models are equally accurate cannot be rejected in any market. On the other hand, the out-of-sample RMSFEs for the A-GEN model consistently exceed those of the A-RLS model sometimes substantially. In summary, for both forecasting models based on squared return deviations and those based on absolute return deviations, our evidence indicates that an exponentially declining weight structure tends to under-weight the youngest and oldest observations and over-weight those in-between. However, by adding parameters, more flexible lag forms introduce additional estimation error which severely impacts their out-of-sample forecasting ability.

VI. Conclusions.

We have compared the forecasting ability of several volatility models - the historical standard deviation, Riskmetric's exponentially weighted moving average, GARCH(1,1), AGARCH, EGARCH, and two regression models developed here - focusing on four issues: (1) the proper weighting of older versus recent observations, (2) the relevance of the parameter estimation procedure, (3) the tradeoff between model flexibility and estimation error, and (4) the proper weighting of large return surprises. As regards the first issue, our evidence indicates that the

GARCH(1,1) model puts too much weight on the most recent observations and not enough on older observations. However we find that out-of-sample forecast accuracy is relatively insensitive to this parameter choice. As regards the second, we find that different parameter estimation procedures result in quite different parameter estimates for the same model. In particular, we find that regression estimates of the exponential model differ substantially and consistently from those estimated using the GARCH procedure. The regression parameter estimates put more weight on older observations and are also more stable. Again however out-of-sample forecast accuracy appears relatively insensitive to the parameter choice.

Turning to the third issue, our evidence underscores the cost of added flexibility in terms of forecast accuracy. While more complex and flexible models forecast better than simple models on an in-sample basis, by adding parameters they increase the scope for estimation error and forecast consistently worse out-of-sample. Perhaps our strongest results relate to the fourth issue. Apparently because extreme observations have less impact on the forecasts when absolute return deviations are used, we find that an exponential model based on absolute return deviations forecasts considerably better than one based on squared return deviations.

While no one model dominates at all horizons in all nine markets, one certainly stands out from the others. In most markets and at most horizons, the forecasting model with the lowest root mean squared forecast error or mean absolute forecast error among the models we consider is the least squares regression model developed here (A-RLS) in which the forecast volatility for the coming period is a weighted average of recent absolute return deviations with exponentially declining weights. GARCH or EGARCH often come in second but they are dominated by the ARLS model in most markets.

REFERENCES

Akgiray, V., 1989, Conditional heteroskedasticity in time series of stock returns: evidence and forecasts, *Journal of Business* 62, 55-80.

Andersen, Torben, and Tim Bollerslev, 1998, Answering the skeptics: Yes standard volatility models do provide accurate forecasts, *International Economic Review* 39, #4, 885-905.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, #2, 579-625.

Andersen, T.G., T. Bollerslev, and S. Lange, 1999, Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon, *Journal of Empirical Finance* 6, 457-477.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3-30

Bollerslev, Tim, and Hans Ole Mikkelsen, 1996, Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* 73, 151-184.

Brailsford, Tim, and Robert Faff, 1996, An evaluation of volatility forecasting techniques, *Journal of Banking and Finance* 20, 419-438.

Brooks, Chris, 1998, Predicting stock index volatility: can market volume help, *Journal of Forecasting* 17, 59-80.

Diebold, Francis X., and Roberto S. Mariano, 1995, *Journal of Business and Economic Statistics* 13, 253-263.

Dimson, E., and P. Marsh, 1990, Volatility forecasting without data snooping, *Journal of Banking and Finance* 14, 399-421.

Figlewski, Stephen, 1997, *Forecasting volatility*, Financial Markets, Institutions, and Instruments vol. 6 #1, Stern School of Business, (Blackwell Publishers: Boston).

Hansen, Peter R., and Asger Lunde, 2001, A forecast comparison of volatility models: does anything beat a GARCH(1,1)? working paper #01-04, Brown University Department of Economics,

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

Lopez, Jose, 2001, Evaluating the predictive accuracy of volatility models, *Journal of Forecasting* 20, 87-109.

Loudon, G.F., W.H. Watt, and P.K. Yadav, 2000, An empirical analysis of alternative parametric ARCH models, *Journal of Applied Econometrics* 15, 117-136.

Mitchell, Douglas, and Paul Speaker, 1986, A simple, flexible distributed lag technique: the polynomial inverse lag, *Journal of Econometrics* 31, 329-340.

Pagan, A. R. and G. W. Schwert, 1990, Alternative models of conditional stock volatilities, *Journal of Econometrics* 45, 267-290.

Poon, Ser-Huang and Clive Granger, 2003, Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature* 41, #2, 478-539.

Schmidt, Peter, 1974, A modification of the Almon distributed lag, *Journal of the American Statistical Association* 69, 679-681.

Tse, Y. K., 1991, Stock returns volatility in the Tokyo Stock Exchange, *Japan and the World Economy* 3, 285-298.

West, Kenneth, and Dogchul Cho, 1995, The predictive ability of several models o exchange rate volitility, *Journal of Econometrics* 69, 367-391.

Figure 1
Comparison of Implied Coefficients of Lagged Squared Return Deviations for GARCH and the Restricted Least Squares Models
10-year T-Bond Market

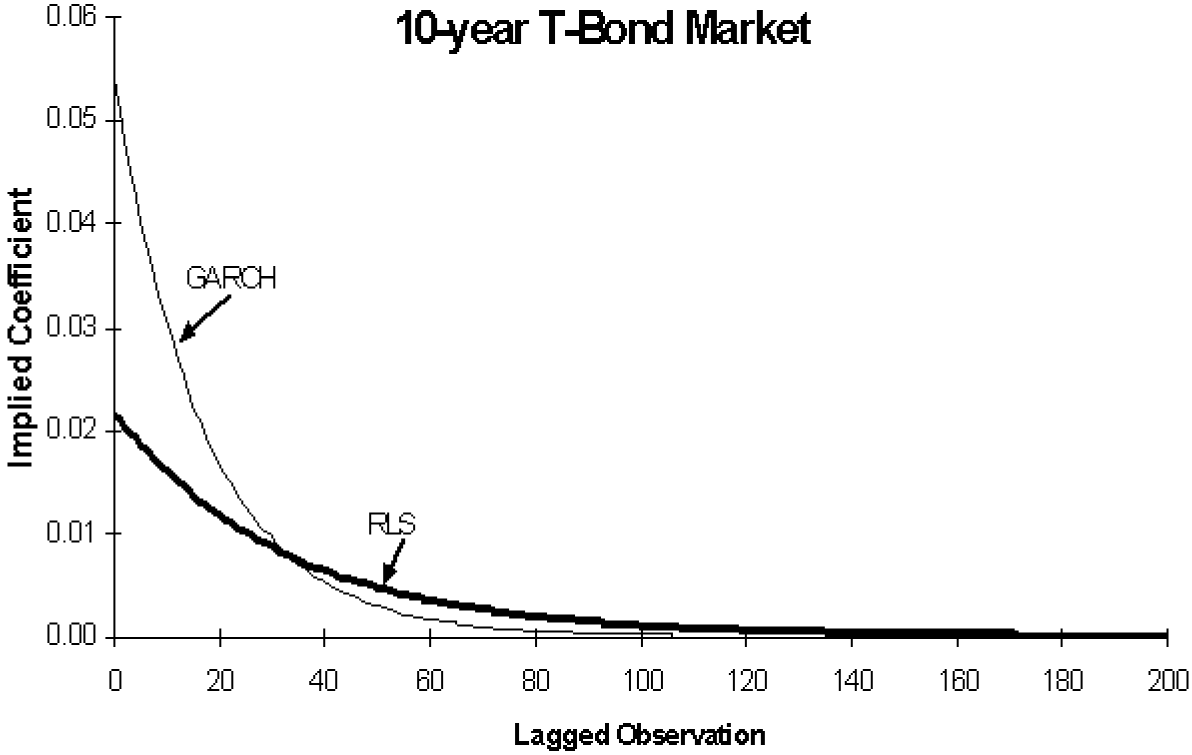

Figure 2a
Comparison of Implied Coefficients of Lagged Squared Return Deviations for the GARCH, RLS, and GEN Models
S&P 500 Index Market

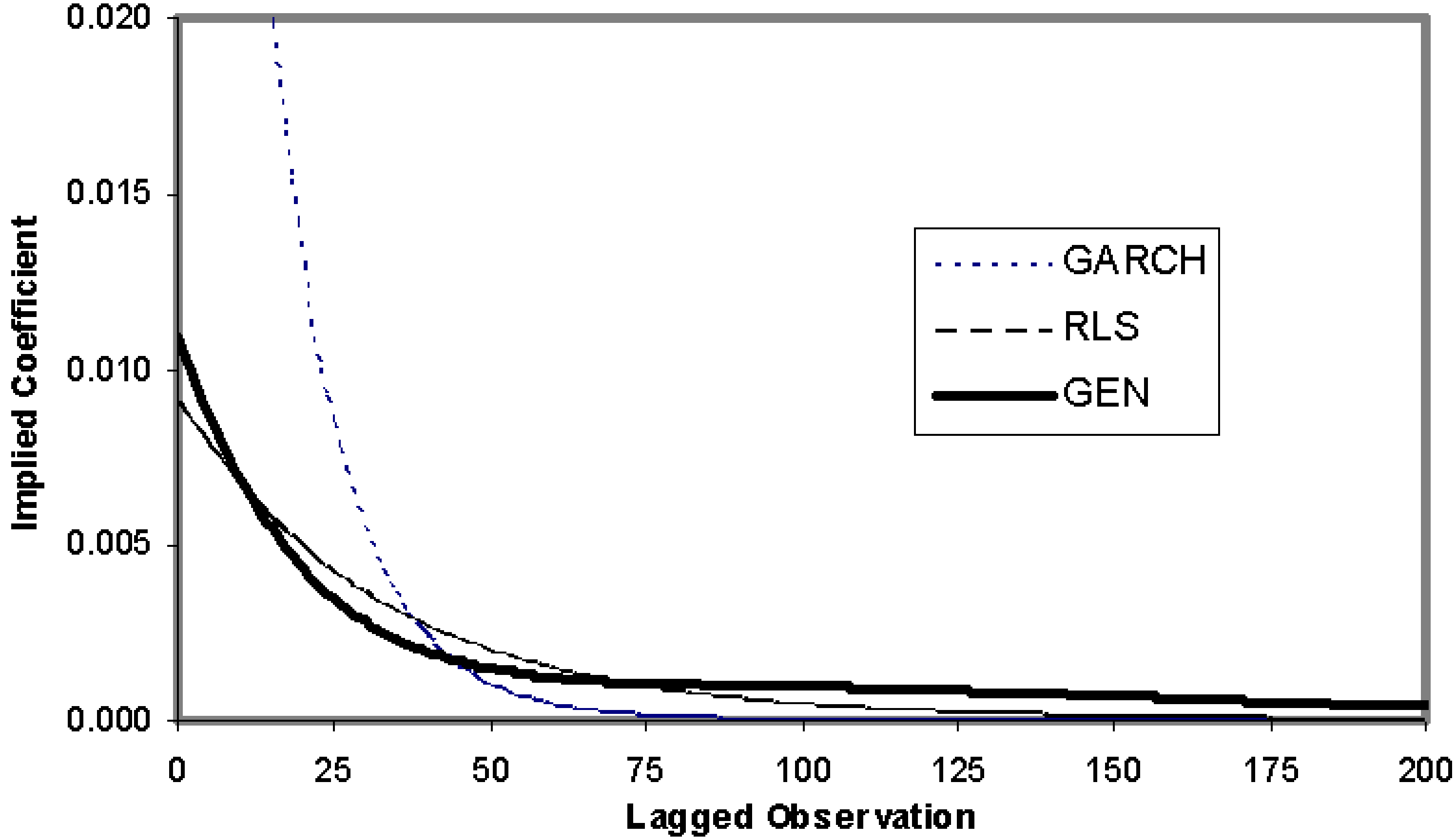

Figure 2b
Comparison of Implied Coefficients of Lagged Squared Return Deviations for the GARCH, RLS and GEN Models
T-Bond Futures Market

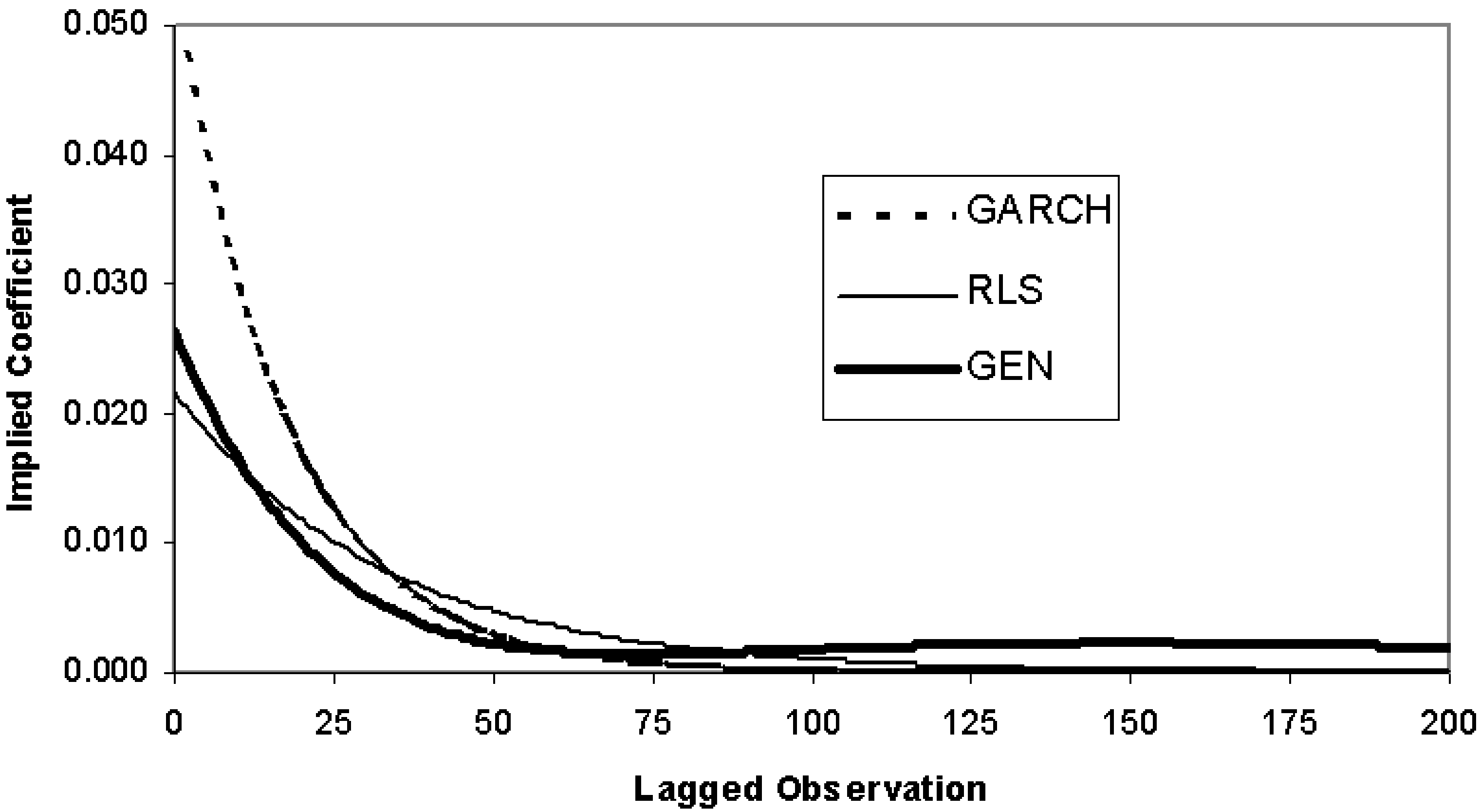

Table 1 - In-Sample Root Mean Squared Forecast Errors of LSD Class Volatility Models

The procedures listed in column 1 are used to forecast the annualized standard deviation of daily returns over the next 40 trading days. Root mean squared forecast errors for these forecasts are reported in Panel A. STD(n) denotes the historical standard deviation over the last n days. EWMA denotes an exponentially weighted moving average model using the Riskmetrics depreciation parameter. GARCH denotes the forecast derived from a GARCH(1,1) model. RLS is structurally identical to the GARCH model but the parameters are estimated using OLS. The lowest RMSFE among the STD(n)'s is shown in bold face, the lowest among the STD(n), GARCH, EWMA and RLS models is shaded. For comparison, we also report the mean and standard deviation of the ex post standard deviation, AS at the bottom of the table.

Forecasting Model		Markets								
		S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
Panel A - Root Mean Squared Forecast Errors										
STD(10)		0.06814	0.04719	0.11156	0.04841	0.12488	0.09028	0.10561	0.11283	0.09293
STD(20)		0.06467	0.04225	0.09697	0.04476	0.10564	0.08151	0.09250	0.10067	0.08224
STD(40)		0.06256	0.03977	0.08688	0.04242	0.09172	0.07555	0.08261	**0.09405**	0.07580
STD(80)		0.06337	**0.03929**	0.08465	0.03988	0.08774	0.07497	0.08039	0.09635	**0.07563**
STD(120)		**0.06235**	0.03944	**0.08254**	**0.03848**	**0.08613**	**0.07326**	**0.07951**	0.09693	0.07574
EWMA		0.06123	0.03870	0.08658	0.04105	0.09315	0.07536	0.08350	0.09273	0.07559
GARCH		0.05969	0.03771	0.08554	0.03754	0.08103	0.06659	0.07287	0.08723	0.06708
RLS		0.05802	0.03737	0.07859	0.03581	0.08078	0.06578	0.07191	0.08558	0.06683
AS	Mean	0.12324	0.10987	0.24595	0.09790	0.31747	0.18604	0.25034	0.28340	0.20613
	STD	0.06353	0.05369	0.12714	0.03946	0.10314	0.07229	0.08135	0.11674	0.07630
OBS		8111	8523	5965	6322	7859	7859	7859	6849	7859

Table 2 - Comparison of Parameters for the GARCH(1,1) and Restricted Least Squares Models

For both the restricted least squares (RLS) and GARCH models, we report parameter estimates for the forecasting model $FV_t = \alpha + \lambda\sum_j \beta^j r^2_{t-j}$ where FV_t is the forecast of the variance over the period from day t+1 through t+40 and J=200. The RLS parameters are obtained by regressing the ex post variance over 40-day periods on the lagged r^2's using non-linear least squares (where the possible β are in increments of .005). The GARCH parameters are obtained by first estimating the GARCH(1,1) model $v_{t+1} = \alpha_0 + \alpha_1 r_t^2 + \beta v_t$ (whose parameter estimates are also reported) where v_t is the conditional variance on day t and then solving for the implied parameters α and λ in the 40-day forecasting model as outlined in equations 5, 6, and 7 in the text.

Market	GARCH(1,1)					Restricted Least Squares		
	α_0	α_1	β	α	λ	β	α	λ
S&P 500	5.110e-7	0.07702	0.91867	1.522e-5	0.07089	0.970	5.322e-5	0.00910
10 year T-bond	8.889e-8	0.05534	0.94334	3.234e-6	0.05392	0.970	1.683e-5	0.02162
Eurodollar	9.660e-7	0.06195	0.93655	3.327e-5	0.06017	0.990	3.173e-5	0.01008
Dollar/DMark	1.190e-6	0.10861	0.86820	2.356e-5	0.07128	0.985	2.186e-5	0.00799
Boeing Co.	1.985e-6	0.02551	0.96988	9.750e-5	0.02334	0.980	1.355e-4	0.01410
Exxon Corp.	5.115e-6	0.08093	0.88487	9.181e-5	0.04445	0.970	1.048e-4	0.01015
Int'l Paper Co.	3.870e-6	0.04050	0.94577	1.191e-4	0.03133	0.970	1.563e-4	0.01299
McDonald's	2.261e-6	0.04652	0.94842	8.118e-5	0.04221	0.965	1.135e-4	0.02404
3M	4.190e-6	0.05122	0.92700	1.017e-4	0.03443	0.970	1.079e-4	0.01313

Table 3 - Parameter Estimates for the GEN Model

Parameter estimates for the model $AV(40)_t = \alpha + \lambda_0 Z_{0t} + \lambda_1 Z_{1t} + \lambda_2 Z_{2t} + \varepsilon_t$ are reported where $Z_{it} = \sum \beta^j j^i r_{t-j}^2$ for j = 0, 1, ... , 200, r_t is the return on day t expressed in deviation (from its mean) form, and $AV(40)_t$ is the ex post variance of returns over the period from t+1 through t+40. The parameters are estimated using non-linear least squares over the entire data period. The F statistic for a test of the null hypothesis that $\lambda_1 = \lambda_2 = 0$ is also reported.

Market	Parameter Estimates					F ($\lambda_1=\lambda_2=0$)	RMSFE
	β	α	λ_0	λ_1	λ_2		
S&P 500	0.975	4.975e-5	0.01097	-0.00024	2.525e-6	11.898	.05758
10 year T-bond	0.980	1.289e-5	0.02633	-0.00069	5.525e-6	165.45	.03634
Eurodollar	0.970	3.049e-5	0.02444	-0.00099	1.665e-5	95.268	.07764
Dollar-DMark	0.965	2.072e-5	0.02014	-0.00106	1.836e-5	76.266	.03534
Boeing Co.	0.970	1.277e-4	0.01711	-0.00021	5.501e-6	14.016	.08075
Exxon Corp.	0.880	1.107e-4	0.02223	-0.00432	3.962e-4	23.165	.06585
Int'l Paper Co.	0.965	1.482e-4	0.01638	-0.00034	5.986e-6	12.343	.07188
McDonald's Co.	0.980	1.000e-4	0.02558	-0.00050	3.068e-6	22.391	.08520
3M	0.900	1.141e-4	0.01898	-0.00179	2.199e-4	12.741	.06681

Table 4 - Out-of-Sample Root Mean Squared Forecast Errors of LSD Class Volatility Models.

The procedures listed in column 1 are used to forecast the annualized standard deviation of daily returns over the next 40 trading days. Root mean squared forecast errors for these forecasts are reported in each cell. The number of periods over which the RMSFE is calculated is reported in the last row. STD(n) denotes the historical standard deviation over the last n days. GARCH denotes the forecast derived from a GARCH(1,1) model. RLS is structurally identical to the GARCH model but the parameters are estimated using OLS. GEN is a more flexible regression model defined in equations 11 in the text. Parameters of all models except STD(n) and EWMA are estimated using data over the past 1260 market days and the models are re-estimated every forty days. The lowest RMSFE among the STD(n)'s is shown in bold face, the lowest among the STD(n), EWMA, GARCH, RLS, and GEN models is shaded. For comparison, we also report the mean and standard deviation of the ex post standard deviation, AS. Under the null that two forecasts are equally accurate, the Diebold-Mariano S1 statistics in Panel B are normally distributed (0,1) in large samples.

Forecasting Model		Markets								
		S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
Panel A - Roor Mean Squared Forecast Errors										
STD(10)		0.07221	0.04971	0.10520	0.04299	0.12918	0.09556	0.10733	0.11081	0.09385
STD(20)		0.06852	0.04444	0.09368	0.03948	0.10959	0.08660	0.09438	0.10046	0.08384
STD(40)		0.06621	0.04182	0.08617	0.03750	0.09459	0.08031	0.08519	**0.09652**	0.07723
STD(80)		0.06721	**0.04151**	0.08490	0.03677	0.09061	0.07961	0.08313	0.09897	**0.07716**
STD(120)		**0.06614**	0.04167	**0.08398**	**0.03593**	**0.08901**	**0.07785**	**0.08272**	0.09811	0.07725
EWMA		0.06492	0.04072	0.08447	0.03618	0.09638	0.08011	0.08563	0.09397	0.07700
GARCH		0.06128	0.04093	0.08908	0.03420	0.08914	0.07346	0.07916	0.08791	0.07459
RLS		0.06499	0.04056	0.08926	0.03361	0.08706	0.07786	0.08895	0.09414	0.07800
GEN		0.07428	0.04390	0.09317	0.03461	0.08996	0.07928	0.10300	0.10286	0.08174
Panel B - Diebold-Marino Tests for Differences in Forecast Accuracy										
RLS vs. STD(120)		0.207	0.607	-0.922	1.541	0.540	-0.001	-1.355	0.635	-0.155
GARCH vs. STD(120)		0.693	0.368	-0.891	0.931	-0.035	0.608	0.622	1.501	0.481
RLS vs. GARCH		-1.793	0.276	-0.057	0.518	1.413	-2.266	-1.776	-2.440	-1.946
AS	Mean	0.13133	0.11862	0.21370	0.10238	0.32072	0.19376	0.25505	0.26650	0.20611
	STD	0.06469	0.05134	0.12045	0.03503	0.10412	0.07456	0.08397	0.11606	0.07954
OBS		6812	7224	4666	5023	6560	6560	6560	5550	6560

Table 5 - Out-of-Sample Root Mean Squared Forecast Errors for non-LSD Models.									
The procedures listed in column 1 are used to forecast the annualized standard deviation of daily returns over the next 40 trading days and the resulting root mean squared errors are reported in Panel A. A-RLS represents the forecast from the restricted least squares model based on absolute deviations as defined in equations 12 and 13 in the text. Parameters of the A-RLS, AGARCH, and EGARCH models are calculated using daily data over the last 1260 days and the model is re-estimated every 40 days. Within each group, the lowest RMSFE is shown in bold face. The lowest RMSFE among the STDk(n), EWMA, GARCH, AGARCH, EGARCH, A-RLS and RLS models is shaded. Under the null that two forecasts are equally accurate, the Diebold-Mariano S1 statistics in Panel B are normally distributed (0,1) in large samples.									
Forecasting Model	Markets								
	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
Panel A - Root Mean Squared Forecast Errors									
STD(20)	0.06852	0.04444	0.09368	0.03948	0.10959	0.08660	0.09438	0.10046	0.08384
STD(40)	0.06621	0.04182	0.08617	0.03750	0.09459	0.08031	0.08519	**0.09652**	0.07723
STD(80)	0.06721	**0.04151**	0.08490	0.03677	0.09061	0.07961	0.08313	0.09897	**0.07716**
STD(120)	**0.06614**	0.04167	**0.08398**	**0.03593**	**0.08901**	**0.07785**	**0.08272**	0.09811	0.07725
EWMA	0.06492	0.04072	0.08447	0.03618	0.09638	0.08011	0.08563	0.09397	0.07700
A-RLS	**0.05634**	**0.03797**	**0.08625**	**0.03165**	**0.08473**	0.07817	**0.07781**	**0.08922**	**0.06946**
RLS	0.06499	0.04056	0.08926	0.03361	0.08706	**0.07786**	0.08895	0.09414	0.07800
AGARCH	0.06167	0.04195	0.09760	0.03407	0.08978	0.07538	0.08304	0.09083	0.07593
EGARCH	**0.05978**	**0.03997**	0.09346	**0.03265**	**0.08770**	0.07378	0.08376	0.09045	**0.07273**
GARCH	0.06128	0.04093	**0.08908**	0.03420	0.08914	**0.07346**	**0.07916**	**0.08791**	0.07459
Panel B - Diebold-Marino Tests for Differences in Forecast Accuracy									
A-RLS vs RLS	3.094	3.030	1.188	2.737	1.287	-0.074	1.860	2.584	2.591
GARCH vs. EGARCH	-1.091	-0.864	1.919	-1.435	-0.454	0.592	1.997	0.618	NA
A-RLS vs. GARCH	1.546	2.743	0.823	2.342	2.307	-0.861	0.570	-0.468	2.776
OBS	6812	7224	4666	5023	6560	6560	6560	5550	6560

Table 6 - Parameter Estimates for A-RLS and A-GEN Models									
We report parameter estimates for the equations $AS(40)_t = \alpha + \lambda W_{0t} + \epsilon_t$ and $AS(40)_t = \alpha'' + \lambda_0 W_{0t} + \lambda_1 W_{1t} + \lambda_2 W_{2t} + \varepsilon_t$ where $W_{it} = \sum \beta^j j^i \lvert r_{t-j} \rvert$ for j=1,...200, r_t is the daily return on day t in deviation form, and $AS(40)_t$ is the ex post standard deviation of returns over the period from t+1 through t+40. The models are estimated using non-linear least squares over the data periods reported in Table 1. The F statistic for a test of the null hypothesis that $\lambda_1 = \lambda_2 = 0$ is also reported. Values of β from .5 to 1.0 in increments of .005 are considered.									
Market	A-RLS model $AS(40)_t = \alpha + \lambda W_{0t}$			A-GEN model $AS(40)_t = \alpha'' + \lambda_0 W_{0t} + \lambda_1 W_{1t} + \lambda_2 W_{2t}$					
	β	α	λ	β	α	λ_0	λ_1	λ_2	F
S&P 500	0.940	0.00270	0.05086	0.970	0.00189	0.05218	-0.00192	1.926e-5	92.974
10 year T-bond	0.965	0.00161	0.03586	0.975	0.00134	0.04381	-0.00120	1.045e-5	127.77
Eurodollar	0.980	0.00400	0.02098	0.970	0.00380	0.03624	-0.00124	1.725e-5	89.112
Dollar/DMark	0.955	0.00297	0.03110	0.955	0.00259	0.04807	-0.00268	4.749e-5	84.781
Boeing Co.	0.975	0.00503	0.02465	0.975	0.00442	0.02904	-0.00046	5.170e-6	29.114
Exxon Corp.	0.960	0.00513	0.02915	0.970	0.00442	0.03492	-0.00103	1.100e-5	36.470
Int'l Paper Co.	0.970	0.00534	0.02601	0.940	0.00527	0.03242	-0.00081	6.410e-5	7.357
McDonald's Co.	0.955	0.00447	0.04338	0.970	0.00372	0.05015	-0.00143	1.297e-5	45.799
3M	0.955	0.00485	0.03693	0.890	0.00499	0.04440	-0.00203	4.369e-4	7.102

Appendix A.1
Out-of-Sample Root Mean Squared Forecast Errors For a 10 Trading Day Forecasting Horizon

	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
STD(10)	0.07359	0.05452	0.11359	0.04523	0.15135	0.10337	0.11853	0.11968	0.10441
STD(20)	0.07184	0.05068	0.10815	0.04402	0.13811	0.09730	0.11054	0.11171	0.09732
STD(40)	**0.07023**	**0.04902**	**0.10413**	**0.04305**	0.12933	0.09257	0.10578	**0.10846**	0.09258
STD(80)	0.07124	0.05016	0.10565	0.04403	0.12549	**0.09087**	**0.10390**	0.11229	**0.09219**
STD(120)	0.07204	0.05107	0.10702	0.04418	**0.12517**	0.09105	0.10457	0.11593	0.09279
EWMA	0.06775	0.04772	0.10063	0.04114	0.12849	0.09073	0.10322	0.10531	0.09118
GARCH	0.06558	0.04841	0.10492	0.04046	0.12518	0.08622	0.10082	0.10401	0.08859
RLS	0.06877	0.04871	0.10857	0.04024	0.12466	0.09036	0.10209	0.10582	0.09113
A-RLS	0.06207	0.04625	0.10198	0.03850	0.11915	0.08663	0.09795	0.10262	0.08531
AGARCH	0.06406	0.04806	0.10575	0.03972	0.12333	0.08775	0.10027	0.10513	0.08846
EGARCH	0.06502	0.04752	0.10596	0.03906	0.12281	0.08704	0.10170	0.10419	0.08829
OBS	6732	7144	4586	4943	6480	6480	6480	5470	6480

Note: The lowest RMSFE among the five STD(n)'s is shown in bold. In each column, the cell with the lowest RMSFE is shaded.

Appendix A.2
Out-of-Sample Root Mean Squared Forecast Errors For a 20 Trading Day Forecasting Horizon

	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
STD(10)	0.07276	0.05094	0.10849	0.04394	0.13804	0.09792	0.11106	0.11256	0.09828
STD(20)	0.06948	0.04585	0.09935	0.04133	0.12160	0.08964	0.10035	0.10187	0.08913
STD(40)	**0.06744**	**0.04404**	**0.09288**	**0.03957**	0.10987	0.08472	0.09328	**0.09910**	0.08326
STD(80)	0.06864	0.04459	0.09381	0.03999	0.10561	0.08354	**0.09114**	0.10244	**0.08266**
STD(120)	0.06855	0.04526	0.09403	0.03977	**0.10455**	**0.08293**	0.09120	0.10524	0.08301
EWMA	0.06571	0.04268	0.09072	0.03813	0.10993	0.08351	0.09230	0.09586	0.08251
GARCH	0.06280	0.04335	0.09521	0.03698	0.10445	0.07906	0.08792	0.09245	0.07954
RLS	0.06697	0.04328	0.09815	0.03652	0.10413	0.08221	0.09324	0.09619	0.08270
A-RLS	0.05831	0.04079	0.09398	0.03482	0.08324	0.08048	0.08516	0.09212	0.07549
AGARCH	0.06146	0.04322	0.09901	0.03590	0.10327	0.08059	0.08834	0.09391	0.07958
EGARCH	0.06181	0.04221	0.09777	0.03511	0.10218	0.07958	0.08977	0.09386	0.07853
OBS	6732	7144	4586	4943	6480	6480	6480	5470	6480

Note: The lowest RMSFE among the five STD(n)'s is shown in bold. In each column, the cell with the lowest RMSFE is shaded.

Appendix A.3
Out-of-Sample Root Mean Squared Forecast Errors For a 80 Trading Day Forecasting Horizon

	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
STD(10)	0.07525	0.05072	0.10730	0.04443	0.12594	0.09609	0.10552	0.11564	0.09449
STD(20)	0.07173	0.04503	0.09494	0.04036	0.10566	0.08765	0.09275	0.10482	0.08459
STD(40)	0.06906	0.04190	0.08575	0.03720	0.09091	0.08208	0.08393	0.10039	0.07838
STD(80)	0.06700	0.04031	0.08127	0.03476	0.08404	0.07838	0.07995	0.09746	**0.07606**
STD(120)	**0.06561**	**0.03965**	**0.07855**	**0.03302**	**0.08231**	**0.07638**	**0.07941**	**0.09481**	0.07616
EWMA	0.06791	0.04116	0.08488	0.03649	0.09205	0.08143	0.08405	0.09801	0.07787
GARCH	0.06170	0.04086	0.08769	0.03080	0.08132	0.06988	0.07665	0.08810	0.07538
RLS	0.06914	0.03971	0.08653	0.03063	0.07756	0.07839	0.08870	0.10269	0.07793
A-RLS	0.05852	0.03738	0.08424	0.03001	0.07767	0.07443	0.07393	0.09433	0.06905
AGARCH	0.06716	0.04521	0.10247	0.03106	0.08548	0.07256	0.09009	0.09483	0.08114
EGARCH	0.05953	0.04049	0.09372	0.02903	0.08186	0.07006	0.08830	0.09080	0.07203
OBS	6732	7144	4586	4943	6480	6480	6480	5470	6480

Note: The lowest RMSFE among the five STD(n)'s is shown in bold. In each column, the cell with the lowest RMSFE is shaded.

Appendix A.4
Out-of-Sample Root Mean Squared Forecast Errors For a 120 Trading Day Forecasting Horizon

	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
STD(10)	0.07569	0.05139	0.10721	0.04454	0.12553	0.09642	0.10617	0.11833	0.09561
STD(20)	0.07140	0.04556	0.09425	0.04007	0.10443	0.08748	0.09274	0.10680	0.08557
STD(40)	0.06813	0.04203	0.08416	0.03632	0.08913	0.08113	0.08364	0.09951	0.07898
STD(80)	0.06608	0.03971	0.07820	0.03308	0.08225	0.07735	0.07968	0.09505	0.07671
STD(120)	**0.06476**	**0.03852**	**0.07554**	**0.03106**	**0.07990**	**0.07576**	**0.07937**	**0.09272**	**0.07651**
EWMA	0.06760	0.04156	0.08388	0.03592	0.09068	0.08118	0.08418	0.09881	0.07893
GARCH	0.06083	0.04071	0.08834	0.02897	0.07774	0.06744	0.07646	0.08856	0.07677
RLS	0.07076	0.03952	0.08502	0.02906	0.07541	0.07381	0.08418	0.10244	0.07800
A-RLS	0.05809	0.03795	0.08397	0.02801	0.07578	0.07078	0.07335	0.09210	0.06952
AGARCH	0.07328	0.04884	0.10682	0.03032	0.08522	0.07057	0.10404	0.09952	0.08766
EGARCH	0.05840	0.04087	0.09485	0.02672	0.08058	0.06741	0.09925	0.09033	0.07238
OBS	6732	7144	4586	4943	6480	6480	6480	5470	6480

Note: The lowest RMSFE among the five STD(n)'s is shown in bold. In each column, the cell with the lowest RMSFE is shaded.

Appendix A.5
Out-of-sample Mean Absolute Forecast Errors For a 40 Trading Day Forecasting Horizon

	S&P 500	10-year T-Bond	90-day Eurodollar	Dollar/ DMark	Boeing	Exxon	Int'l Paper	McDonalds	3M
STD(10)	0.03700	0.03642	0.07521	0.03231	0.09456	0.05684	0.07311	0.07366	0.06085
STD(20)	0.03445	0.03276	0.06800	0.02967	0.07945	0.04957	0.06241	0.06432	0.05292
STD(40)	**0.03415**	**0.03125**	0.06387	0.02904	0.06975	0.04434	0.05476	**0.06016**	0.04910
STD(80)	0.03522	0.03195	**0.06321**	0.02912	0.06863	0.04285	**0.05271**	0.06120	**0.04755**
STD(120)	0.03556	0.03196	0.06331	**0.02856**	**0.06821**	**0.04196**	0.05446	0.06253	0.04778
EWMA	0.03278	0.03030	0.06238	0.02743	0.07044	0.04466	0.05520	0.05931	0.04791
GARCH	0.03366	0.03366	0.03155	0.07149	0.02705	0.07214	0.04206	0.05177	0.059020.0
RLS	0.03764	0.03146	0.07020	0.02710	0.06955	0.04641	0.05813	0.06502	0.05153
A-RLS	0.02980	0.02868	0.06586	0.02476	0.06696	0.04237	0.05039	0.05782	0.04284
AGARCH	0.03601	0.03366	0.08112	0.02734	0.07395	0.04439	0.05587	0.06288	0.05100
EGARCH	0.03283	0.03148	0.07637	0.02630	0.07101	0.04242	0.05475	0.06074	0.04676
OBS	6812	7224	4666	5023	6560	6560	6560	5550	6560

Note: The lowest MAFE among the five STD(n)'s is shown in bold. In each column, the cell with the lowest MAFE is shaded.

ENDNOTES

1. For an excellent review of existing studies in this area see Poon and Granger (2003). Notable examples include Akgiray (1989), Pagan and Schwert (1990), Tse (1991), Jorion (1995), West and Cho (1995), Brailsford and Faff (1996), Figlewski (1997), Brooks (1998), Loudon et al (2000), Lopez (2001), Hansen and Lunde (2001), and Anderson et al (2003).

2. In their excellent survey Poon and Granger (2003) survey 93 studies with out-of-sample comparisons but about half of these compare time-series and implied volatility forecasts. Of the 39 comparing GARCH(1,1) with the historical variance, 17 find that GARCH forecasts better while 22 find in favor of the historical variance.

3. In a series of papers Andersen and Bollerslev and others (e.g., Andersen et al, 2003) have shown that (at least for fairly short forecast horizons such as a day or week or two) models based on high frequency intra-day data forecast better than those based on daily data but these high frequency models have not yet gained acceptance in practice.

4. In the case of dividend paying stocks, P_t is adjusted for any dividends.

5. Often μ is replaced by the sample mean and accordingly the r^2 are divided by n-1, rather than n in equation 1. 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 VAR(n). Given the low auto-correlation in returns there is no justification for this assumption and Figlewski (1997) shows that better forecasts are obtained by setting μ=0. In our calculations, we set μ equal to the average daily return over the entire data period.

6. For instance, when Lopez (2001) estimates an AR(10), in every market the largest coefficient is for a lag of 6 days or longer.

7. These periods overlap. In other words, we calculate $AS(s)_m$ and $FSTD(s)_m$ for the s-day period beginning on day m, move up one day to day m+1 and recalculate.

8. Results of other forecast horizons are quite similar and are available from the authors on request.

9. We follow an iterative estimation procedure to estimate the parameters in equation 9. Specifically, for values of β from .500 to 1.000 in increments of .005, we form the variables $Z(\beta)_t$ as defined in equation 8, then regress the realized or ex post variance over the period from t through t+40, $AV(40)_t$, on $Z(\beta)_t$ using OLS. This regression is repeated for each value of β from .500 to 1.000. The regression with the lowest error-sum-of-squares yields our estimates of β, α, and λ. Using these parameter estimates, the observed values of r^2_{t-j} from j=200 through j=0, and equation 9, we obtain forecast values of the standard deviations over the interval from t to t+40.In obtaining the RLS forecasts we set J=200 in equations 8 and 9, that is we use only the 200 most recent squared return deviations to forecast the future standard deviation. This saves considerable computing time for all our models in all our markets, and observations before t-200 have an inconsequential effect on the forecast. For instance for β=.97, the implied coefficient of r^2_{t-200} is

only 0.23% of the coefficient of r_t^2. Note that RLS minimizes the variance of the volatility forecast error where this error is defined in terms of the variance. This is not quite the same as minimizing the in-sample RMSFE since the latter was defined in terms of the standard deviation but comes quite close

10. As Schmidt (1974) points out, for $\beta<1$, the term β^j eventually dominates the j^i terms as the length of the lag j increases so the weights on distant observations eventually decline toward zero. However, for values of β close to one, this decline may not occur until the lags are quite long. Consequently, we also estimated a least squares model in which the lag coefficients taper off more quickly based on both Schmidt's model and the polynomial inverse lag model of Mitchell and Speaker (1986). Results for this model are not presented here since they are very similar to the SCH model.

11. A similar tendency has been observed within the GARCH type estimators by Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996). The FIGARCH model developed in those papers provides a form to correct this tendency.

12. In order to show clearly implied coefficient differences at longer lags, the implied GARCH coefficients at the shortest lags are off the charts for the S&P500 in Figure 2a and Eurodollars in Figure 2c.

13. Of course, the historical standard deviation forecasts, STD(n), and EWMA forecasts reported in Table 1 are out-of-sample since the weights are fixed.

14. For example, the RLS and GEN models are first estimated using observations 201 (since J=200 in the RLS model) through 1460. The estimated model and r^2 observations before 1500 are then used to forecast volatility over the 40-day period from day 1501 to day 1540. The same parameter estimates but r^2 observations up through day 1501 are used to generate the volatility forecasts for the interval from day 1502 to day 1541 and this procedure is repeated for the next 38 days using unchanged parameters. However, when the time comes to forecast volatility over the interval from day 1541 to 1580, the models are re-estimated using data from day 241 through day 1500, these new parameter estimates are used to generate volatility forecasts for the next 40 days and this process is repeated.

15. The null that the GARCH and EGARCH forecasts are equally accurate is rejected in two markets: Eurodollars and International Paper where the results favor GARCH. The Diebold-Marino test statistic could not be calculated in the 3M market because it involved taking the square root of a negative number.

16. As reported in the Appendix, GARCH also looks somewhat better versus EGARCH when the forecast error is measured in terms of the mean absolute error instead of the RMSFE.