FORECASTING VOLATILITY

LOUIS H. EDERINGTON*
WEI GUAN

The forecasting ability of the most popular volatility forecasting models is examined and an alternative model developed. Existing models are compared in terms of four attributes: (1) the relative weighting of recent versus older observations, (2) the estimation criterion, (3) the trade-off in terms of out-of-sample forecasting error between simple and complex models, and (4) the emphasis placed on large shocks. As in previous studies, we find that financial markets have longer memories than reflected in GARCH(1,1) model estimates, but find this has little impact on out-of-sample forecasting ability. 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. With the exception of GARCH models, we find that models based on absolute return deviations generally forecast volatility better than otherwise equivalent models based on squared return deviations. 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

We have benefited from comments on earlier versions of this paper by Tim Bollerslev, Steve Figlewski, Michael Fleming, Charles Jones, Jose Lopez, an anonymous referee, and seminar participants at the Federal Reserve Bank of New York, Queensland University, and Australian National University. Remaining errors are of course our responsibility.
*Correspondence author, Finance Division, Michael F. Price College of Business, University of Oklahoma, 205A Adams Hall, Norman, OK 73019, USA; e-mail: lederington@ou.edu

Received October 2003; Accepted September 2004

■ *Louis H. Ederington is the Michael F. Price Chair in Finance in the Division of Finance at the Price College of Business at the University of Oklahoma in Norman, Oklahoma.*

■ *Wei Guan is an Assistant Professor of Finance at the College of Business at the University of South Florida in St. Petersburg, Florida.*

The Journal of Futures Markets, Vol. 25, No. 5, 465–490 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fut.20146

average models, though between GARCH and EGARCH there is no clear favorite. However, in terms of forecast accuracy, all are dominated by a new, simple, nonlinear least squares model, based on historical absolute return deviations, that we develop and test here. Jrl Fut Mark 25:465–490, 2005

INTRODUCTION

Accurate volatility forecasts are important to traders, investors, and risk managers, as well as researchers seeking to understand market dynamics. Not only are estimates of future volatility necessary in order to derive option prices, they are also needed to derive deltas and therefore hedge ratios for derivatives portfolios and are critical inputs for value-at-risk models. The two main sources of volatility forecasts are time-series models and implied volatilities calculated from observed option prices. Although theoretically implied volatilities should reflect all available information, including time-series information, evidence is mixed on whether this is the case. Moreover implied volatilities cannot simultaneously be used to price the derivatives from whose prices they are calculated and are only available for specific time horizons for a limited set of assets. Consequently, time-series models remain the major source of volatility forecasts.

This paper examines the forecasting ability of popular time-series volatility forecasting models and develops a superior alternative. 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. However, our approach differs from most in that we focus on why some models forecast better than others, exploring four issues: (1) the relative weighting of recent versus older observations, (2) the importance of the estimation criteria, (3) 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, and (4) the emphasis placed on large shocks. While the first issue, the proper weighting of recent versus older observations, has received considerable attention in the literature, the other three have received much less scrutiny.

We choose models, data series, and forecast horizons that are in common usage. For instance, while many extant studies compare the ability of

[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).

various models to forecast volatility in 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 multi-day horizons. In our presentation, we focus on results for 40 trading days but have also compared the models for horizons 10, 20, 80, and 120 trading days with similar results. We restrict our focus to commonly used volatility forecasting models, such as the historical standard deviation, the GARCH(1,1) model, the exponentially weighted moving average model used by Riskmetrics, and to 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.[2] To ensure generality, we compare forecasting ability across a variety of markets: the S&P 500 index, the Japanese yen/dollar exchange rate, the three-month eurodollar rate, the 10-year treasury bond rate, and five equities—Boeing, GM, International Paper, McDonald's, and Merck.

One issue explored in the paper is the proper weighting of recent versus older observations. Like others,[3] we find that GARCH(1,1) puts too much weight on recent observations relative to older observations. However, despite the attention this issue has received, we found 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 constructed a nonlinear least squares regression model that is structurally identical to a GARCH(1,1) model, but is estimated using a two-stage least squares procedure rather than maximum likelihood. We found that the parameters estimated using least squares implied considerably longer memories than the GARCH model and generate better forecasts in-sample. Out-of-sample results are less clear-cut but the regression model forecasts better in six of the nine markets we examine.

Third, we explored the trade-off between model complexity and forecast error. Models, such as GARCH(1,1), impose a functional form in which the weights attached to squared return observations decline exponentially as the lag increases. We found that a more general model in

[2]In a series of papers, Andersen and Bollerslev and others (e.g., Anderson and Bollerslev, 1998; Anderson et al., 1999; and 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 intraday data forecast better than those based on daily data but these high frequency models have not yet gained acceptance in practice.

[3]Examples include Ding and Granger (1996), Anderson et al. (2003), and Zumbach (2003). See also the review by Poon and Granger (2003).

which the depreciation rate slows as one moves back in time improves in-sample forecast accuracy. However, because the more flexible model involves additional parameters, estimation error is increased and its out-of-sample forecasts are less accurate.

A fourth issue is whether the popular models attach too much weight to large shocks that may or may not reoccur. Since both the historical standard deviation and the GARCH(1,1) forecasts are functions of the *squared* surprise returns, a single large return deviation in the past 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.

Based on these findings, we propose a simple nonlinear least squares model based on past absolute return innovations. In all nine markets at all five forecast horizons that we considered, this regression model forecasts better both in- and out-of-sample than all extant models that we considered (including GARCH, EGARCH, and the historical standard deviation). It has the added advantage of requiring only standard regression software.

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 nonlinear least squares models that is structurally identical to the GARCH(1,1) model, but estimated differently. Our data sets and estimation procedures are described in the next section. After that, 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 to many popular models. Out-of-sample forecasting ability is compared next. We then turn to our fourth issue and compare models based on absolute return deviations to those based on squared deviations. There we develop our model that proves to out-forecast the extant models in all nine markets. Robustness issues are considered before the conclusions.

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 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 cut-off date is weighted equally while observations prior to the chosen cut-off 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.

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:

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

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.

[4]In the case of dividend paying stocks, P_t includes 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 autocorrelation in returns there is little justification for this assumption and Figlewski (1997) shows that better forecasts are obtained by setting $\mu = 0$. In our calculations, we set μ equal to the average daily return over the entire data period.

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 utilizes it for its value-at-risk modeling. Since Riskmetrics sets $\beta = .94$ (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

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

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:

$$v_{t+1} = \alpha_0 + \alpha_1 r_t^2 + \beta v_t \tag{4}$$

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:

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

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:

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

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:

$$\text{GARCH}_t = (1/s)\sum_{k=1}^{s} v_{t+k} = \alpha + \lambda \sum_{j=0}^{J} \beta^j r_{t-j}^2 \tag{7}$$

where $\alpha = (1/s)\sum_{k=1}^{s}[\alpha_0 \sum_{j=0}^{k-2}(\alpha_1 + \beta)^j + \alpha_0'(\alpha_1 + \beta)^{k-1}]$ 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.

Previous Comparisons

While numerous studies have compared the forecasting abilities of the historical variance and GARCH models, no clear winner has emerged. In a thorough review of 39 such studies, Poon and Granger (2003) report that 22 find that historical volatility forecasts future volatility better out-of-sample while 17 find that GARCH models forecast better.

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 nonlinear 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 nonlinear least squares. Letting $AV(s)_t$ represent the actual realized variance over the period from

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

$t + 1$ 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:

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

We label the resulting estimates of Equation (8) the "restricted least squares" (or RLS) model. It is "restricted" in that the coefficients of r_{t-j}^2 are forced to decline exponentially as j increases and are nonlinear 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:

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

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.

DATA AND PROCEDURES

We compare the forecasting ability of these four models for four financial series: the S&P 500 Index, the yen/dollar exchange rate, the three-month eurodollar rate, the 10-year treasury bond rate, and five equities chosen from those in the Dow-Jones Index: Boeing, GM, International Paper, McDonald's, and Merck. Daily price data for the five equities for the period 7/2/62 (7/5/66 for McDonald's) to 12/31/02 were obtained from CRSP tapes as were prices for the S&P 500 index (7/2/62–12/31/02). Daily interest rate and exchange rate data were obtained from Federal Reserve Board files for the periods: 1/2/62–12/31/03 for the 10-year bond rate, 1/4/71–12/31/03 for the eurodollar rate, and 1/4/71–12/31/03 for the yen/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

$$\text{RMSFE} = \left[(1/M) \sum_{m=1}^{M} (AS(s)_m - FSTD(s)_m)^2 \right]^{1/2} \qquad (10)$$

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 the last section. 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 RMSFE in terms of the standard deviation.[8]

ESTIMATION AND WEIGHTING ISSUES

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 I for a forecast horizon of 40 trading days.[9] 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 penultimate row of Table I. Historical 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

[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]RMSFE weights over- and under-forecasts of volatility equally. Holders of long derivative positions would hurt by surprise decreases in volatility and holders of short positions by surprise increases so each would likely have asymmetric loss functions. Due to (1) the general usage of RMSFE in the literature, (2) the difficulty of specifying loss functions, and (3) because underlying positions can either be short or long volatility, we use the symmetric measure.

[9]Results for forecast horizons of 10, 20, 80, and 120 days are quite similar.

TABLE I

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

Forecasting Model		*Markets*								
		S&P 500	*10-year T-Bond*	*90-day Eurodollar*	*Yen/Dollar*	*Boeing*	*GM*	*Int'l Paper*	*McDonald's*	*Merck*
STD(10)		0.06940	0.05063	0.12286	0.04869	0.13095	0.10174	0.11287	0.11241	0.08791
STD(20)		0.06534	0.04545	0.10875	0.04397	0.11183	0.09090	0.09924	0.09882	0.07610
STD(40)		0.06311	0.04289	0.09691	0.04129	0.09894	0.08292	0.08796	**0.09130**	0.06936
STD(80)		0.06383	**0.04232**	0.08952	0.03932	0.09499	0.07950	0.08404	0.09278	**0.06787**
STD(120)		**0.06258**	0.04269	**0.08831**	**0.03887**	**0.09294**	**0.07747**	**0.08238**	0.09321	0.06822
EWMA		0.06167	0.04162	0.09659	0.04050	0.09979	0.08256	0.08903	0.09069	0.06877
GARCH		0.06066	0.04048	0.09635	0.04257	0.08817	0.07730	0.08039	0.08489	0.06109
RLS		0.06003	0.04032	0.08430	0.03627	0.08675	0.07323	0.07773	0.08317	0.06099
AS	Mean	0.13311	0.11907	0.23344	0.09279	0.31888	0.24363	0.26532	0.28086	0.23483
	STD	0.06722	0.06129	0.13200	0.04086	0.10556	0.09155	0.09403	0.11101	0.07218
OBS		9,874	10,636	8,287	8,287	9,874	9,874	9,874	8,865	9,874

Notes. Root mean squared errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 40 trading days using the procedures listed in column 1 are reported. STD(*n*) denotes the historical standard deviation over the last *n* days. EWMA denotes an exponentially weighted moving average model. 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*.

RMSFE for a forecast horizon of 40 days is minimized using the standard deviation over the last 120 days.

Comparing RMSFEs for (1) the 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 seven of our nine markets, the exception being the three-month eurodollar rate and the yen/dollar exchange rate where STD(120) forecasts better.

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 I is shaded. As this shading makes clear, in all nine markets, the RLS model has the lowest in-sample RMSFE.

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.[10]

How and how much the parameters estimates differ is shown in Table II where we report the two models' estimates of the parameters β, α, and λ parameters of Equations (7) and (8). 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

[10]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_{t-j}^2 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 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 $\beta = .97$, the implied coefficient of r_{t-200}^2 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.

TABLE II

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

	GARCH(1,1)					*Restricted Least Squares*		
Market	α_0	α_1	β	α	λ	β	α	λ
S&P 500	5.305e-7	0.0780	0.9185	1.600e-5	0.07291	0.975	5.587e-5	0.00983
10-year T-bond	8.072e-8	0.0511	0.9478	3.066e-6	0.05002	0.975	1.528e-5	0.02026
Eurodollar	1.172e-6	0.0627	0.9360	4.027e-5	0.06114	0.995	3.837e-5	0.00662
Yen/Dollar	1.209e-6	0.1393	0.8448	2.527e-5	0.10366	0.980	1.960e-5	0.01058
Boeing	3.487e-6	0.0344	0.9582	1.346e-4	0.02987	0.980	1.648e-4	0.01304
GM	2.706e-6	0.0594	0.9315	8.052e-5	0.04998	0.990	8.083e-5	0.00845
Int'l Paper	3.836e-6	0.0477	0.9409	1.175e-4	0.03848	0.985	1.332e-4	0.00922
McDonald's	2.500e-6	0.0461	0.9483	8.890e-5	0.04141	0.965	1.201e-4	0.02359
Merck	3.164e-6	0.0405	0.9465	9.888e-5	0.03174	0.955	1.102e-4	0.02463

Notes. We report parameter estimates for the forecasting model $FV_t = \alpha + \lambda \sum \beta^j r_{t-j}^2$ where FV_t is the forecast of the variance over the period from day $t + 1$ through $t + 40$ and $j = 0, 1, \ldots, 200$ using both the restricted least squares (RLS) and GARCH procedures. 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.

markets, $\beta_{RLS} > \beta_{GARCH}$, and $\lambda_{RLS} < \lambda_{GARCH}$. This means that as compared with the GARCH model, 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 S&P 500 index market, as illustrated in Figure 1, the RLS model weights observations on days $t - 34$ through $t <$ GARCH and weighs observations on day $t - 35$ and earlier more heavily. Other implied "cross-over" days are $t - 32$ in the eurodollar market, day $t - 31$ in the T-bond market, and day $t - 16$ in dollar/yen market. The implication is that the GARCH(1,1) procedure yields parameter estimates which overweight 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, such as Ding and Granger (1996). 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

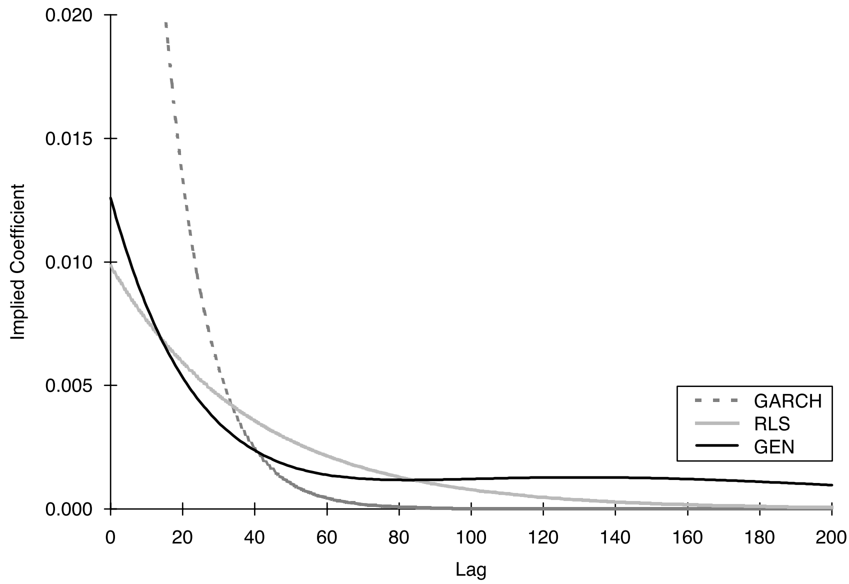

FIGURE 1
Implied coefficients of lagged squared return deviations for the GARCH, RLS, and GEN models for S&P 500 index market.

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:

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

where $I = 2$. Like the RLS model, this model, which we label GEN due to its flexible general form, is estimated using nonlinear 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 III, 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 most the λ_0 estimated for the GEN model exceeds the RLS λ). The implication is that,

TABLE III
Parameter Estimates for the GEN Model

Market	β	α	λ_0	λ_1	λ_2	$F(\lambda_1 = \lambda_2 = 0)$	*RMSFE*
	Parameter Estimates						
S&P 500	0.980	5.099e-5	0.01259	−0.00028	2.457e-6	38.097	.05919
10-year T-bond	0.985	8.740e-6	0.02723	−0.00077	5.621e-6	656.878	.03859
Eurodollar	0.975	3.771e-5	0.01353	−0.00046	9.206e-6	448.159	.08407
Yen/Dollar	0.950	1.991e-5	0.01755	−0.00073	2.974e-5	108.043	.03620
Boeing	0.985	1.457e-4	0.01528	−0.00025	1.791e-6	52.963	.08617
GM	0.995	7.598e-5	0.01258	−0.00022	1.135e-6	159.026	.07226
Int'l Paper	0.970	1.317e-4	0.01474	−0.00032	6.575e-6	79.357	.07766
McDonald's	0.980	1.077e-4	0.02478	−0.00047	2.827e-6	122.054	.08277
Merck	0.985	8.862e-5	0.02314	−0.00055	3.328e-6	308.976	.06058

Notes. 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 $j = 0, 1, \ldots, 200$, 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 nonlinear 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.

in the regression model, an exponential lag structure (in which the weights depreciate at the constant rate β), results in weights which are too low for observations in the recent and distant past and too high in-between.[11] This is illustrated in Figure 1 where the implied coefficients of r_{t-j}^2 for the GEN model are plotted for $j = 0$ to 200 along with those for RLS and GARCH. As also illustrated in Figure 1, the difference in the coefficient lag structure between the RLS and GEN model estimates (that is between the more restrictive and flexible regression forms), tends to be 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 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.

[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 in Figure 1.

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 IV.

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 IV, these relationships no longer hold when out-of-sample forecasts are compared. RLS forecasts best in a majority of markets but not in all. In the four macro markets, RLS has the lowest RMSFE in three: T-bonds, eurodollars, and yen, while GARCH's RMSFE is lowest in the S&P 500 market. RLS's RMSFE is significantly lower than GARCH's at the .05 level for yen and eurodollars, while GARCH's is significantly lower for the S&P 500 index. In the five equity markets, GARCH and RLS each have the lowest RMSFE in two markets but only in the case of Boeing (where RLS's RMSFE is significantly lower) is the difference significant. The historical standard deviation based on 120 observations has the lowest RMSFE in one market, International Paper, but the differences in forecast accuracy among the three models in that market are not significant. In summary, on an out-of-sample basis, GARCH and RLS generally forecast better than the STD, EWMA, and GEN models with RLS forecasting better in most, but not all, markets.

The GARCH(1,1) model was not always well behaved in our estimations. In seven of our nine markets, $\hat{\beta} = 0$ (the lower bound) in some subsamples. 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 9 out of 184

[13]Of course, the historical standard deviation forecasts, STD(n), and EWMA forecasts reported in Table 1 are already 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. 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.

TABLE IV

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

		Markets								
Forecasting Model		*S&P 500*	*10-year T-Bond*	*90-day Eurodollar*	*Yen/Dollar*	*Boeing*	*GM*	*Int'l Paper*	*McDonald's*	*Merck*
Panel A—Root Mean Squared Forecast Errors										
STD(10)		0.07278	0.05309	0.11397	0.04354	0.13496	0.10618	0.11517	0.11117	0.09075
STD(20)		0.06848	0.04764	0.10209	0.03815	0.11559	0.09504	0.10157	0.09865	0.07880
STD(40)		0.06607	0.04497	0.09268	0.03495	0.10202	0.08656	0.09058	**0.09297**	0.07218
STD(80)		0.06694	**0.04448**	0.08851	0.03257	0.09804	0.08316	0.08653	0.09445	**0.07093**
STD(120)		**0.06564**	0.04486	**0.08718**	**0.03199**	**0.09590**	**0.08116**	**0.08517**	0.09376	0.07154
EWMA		0.06466	0.04364	0.09206	0.03442	0.10307	0.08630	0.09138	0.09139	0.07142
GARCH		0.06076	0.04232	0.09036	0.03334	0.09323	0.07820	0.08636	0.08580	0.06493
RLS		0.06437	0.04178	0.08665	0.03150	0.08984	0.07526	0.08593	0.08712	0.06498
GEN		0.07568	0.04456	0.09768	0.03399	0.09575	0.08231	0.11063	0.09869	0.07283
Panel B—Diebold-Marino Tests for Differences in Forecast Accuracy										
RLS vs. STD(120)		0.180	1.978	0.242	0.073	1.902	1.544	−0.072	1.318	2.360
GARCH vs. STD(120)		0.835	1.598	0.769	−2.473	0.772	0.841	−0.135	1.481	2.062
RLS vs. GARCH		−2.452	0.505	1.710	4.033	2.198	1.442	0.179	−0.750	−0.052
AS	Mean	0.14104	0.12776	0.20678	0.10043	0.32166	0.25470	0.27127	0.26823	0.23906
	STD	0.06751	0.05906	0.12195	0.03391	0.10657	0.09178	0.09668	0.10975	0.07536
OBS		8575	9337	6988	6988	8575	8575	8575	7566	8575

Notes. Root mean squared errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 40 trading days using the procedures listed in column 1 are reported. *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 Equation 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 40 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 (1995) S1 statistics in Panel B are normally distributed (0,1) in large samples. The number of 40-day periods over which the RMSFE is calculated is reported in the last row.

subsamples in the eurodollar market and 6 out of 184 in the yen/dollar market, but there were 29 such occurrences for McDonalds (out of 199 subsamples). In contrast, the RLS parameter estimates were much more stable across subsamples. In only one subsample 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). While the GEN model corrects the RLS model's tendency to under-weigh the most recent and oldest observations relative to those in-between, by adding additional parameters to correct this tendency 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 (1) the historical standard deviation, (2) the exponentially weighted moving average model, and (3) a more general model that nests the exponential model. The RLS procedure forecasts better than GARCH in most, but not all, markets, and its parameter estimates are more stable. A more flexible model that nests the exponential model forecasts somewhat better than the RLS and GARCH models in-sample 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.

THE IMPACT OF LARGE RETURN SURPRISES

To this point, all the models that 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 their forecasts are quite sensitive to big outliers. An obvious question, but one that to our knowledge has not been explored heretofore, is whether better forecasts can be obtained using models that are less sensitive to a single highly volatile day, such as models based on *absolute* return deviations.

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 $r_t = R_t - \mu$, and $E(\sqrt{\pi/2}\,\sum_{j=0}^{n-1} W_j|r_{t-j}|) = \sigma$ where $\sum W_j = 1$. 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:

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

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. Once the parameters are determined, we then generate the forecasts:

$$\text{A-RLS}_t = \hat{\alpha} + \hat{\lambda}\sqrt{\pi/2}\sum_{j=0}^{J} \hat{\beta}^j |r_{t-j}| \tag{13}$$

where $J = 200$. Out-of-sample RMSFEs for this forecasting model are reported in Table V where the RMSFEs for the RLS model are repeated for comparison. In all nine markets, A-RLS's RMSFE is lower than that of the RLS model. As reported in panel B, in six 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 4.3% below RLS's across our nine markets.

TABLE V

Out-of-Sample Root Mean Squared Forecast Errors: All Models

	Markets								
Forecasting Model	*S&P 500*	*10-year T-Bond*	*90-day Eurodollar*	*Yen/Dollar*	*Boeing*	*GM*	*Int'l Paper*	*McDonald's*	*Merck*
Panel A—Root Mean Squared Forecast Errors									
STD(20)	0.06848	0.04764	0.10209	0.03815	0.11559	0.09504	0.10157	0.09865	0.07880
STD(40)	0.06607	0.04497	0.09268	0.03495	0.10202	0.08656	0.09058	0.09297	0.07218
STD(80)	0.06694	**0.04448**	0.08851	0.03257	0.09804	0.08316	0.08653	0.09445	**0.07093**
STD(120)	**0.06564**	0.04486	**0.08718**	**0.03199**	**0.09590**	**0.08116**	**0.08517**	**0.09376**	0.07154
EWMA	0.06466	0.04364	0.09206	0.03442	0.10307	0.08630	0.09138	0.09139	0.07142
A-RLS	**0.05736**	**0.04026**	**0.08332**	**0.03077**	**0.08804**	**0.07388**	**0.07858**	**0.08405**	**0.06371**
RLS	0.06437	0.04178	0.08665	0.03150	0.08984	0.07526	0.08593	0.08712	0.06498
AGARCH	0.06122	0.04394	0.11272	0.03517	0.09507	0.07716	0.09111	0.08829	0.06604
EGARCH	**0.05944**	**0.04203**	0.11142	**0.03323**	**0.09164**	**0.07602**	0.09206	0.08826	0.06527
GARCH	0.06076	0.04232	**0.09036**	0.03334	0.09323	0.07820	**0.08636**	**0.08580**	**0.06493**
Panel B—Diebold-Marino Tests for Differences in Forecast Accuracy									
A-RLS vs. RLS	2.054	3.028	2.264	3.187	1.177	0.829	2.405	2.050	1.555
GARCH vs. EGARCH	−1.057	−0.010	1.327	−1.007	−0.657	−1.551	1.661	0.628	0.228
A-RLS vs. GARCH	1.520	2.432	2.924	6.833	2.829	1.506	2.718	0.946	1.110
OBS	8575	9337	6988	6988	8575	8575	8575	7566	8575

Notes. Root mean squared errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 40 trading days using the procedures listed in column 1 are reported. 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. The lowest RMSFE among *STD*(*n*)'s, the lower RMSFE of A-RLS and RLS, and the lowest RMSFE among AGARCH, EGARCH, and GARCH are shown in bold face. The lowest RMSFE among all the 10 models is shaded. Under the null that two forecasts are equally accurate, the Diebold-Mariano (1995) S1 statistics in Panel B are normally distributed (0,1) in large samples.

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:

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

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:

$$\text{AGARCH}_t = \sqrt{(1/s)\sum_{k=1}^{s} E(S_{t+k}^2)} \tag{15}$$

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:

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

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 $k > 1$,

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

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 Japanese yen market, but positive in the two interest rate markets.

As shown in Table V, 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. None of the differences in forecast accuracy between GARCH and EGARCH are significant at the .05 level. Pairwise, GARCH dominates AGARCH in every market except GM and EGARCH dominates AGARCH in all save International Paper. While we found that the regression model based on absolute return deviations forecasts better than equivalent model based on squared deviations, that clearly does not hold true for GARCH type models.

As with GARCH, the EGARCH and AGARCH estimations were not always well-behaved in the 1260 observation subsamples. For the EGARCH model, in eight of the nine markets, $\hat{\beta} < 0$ in some subsamples and, in five markets, $\hat{\beta} > 1$ in some subsamples. In a couple of the markets, there were also instances when $\hat{\gamma} < 0$, implying that higher than normal volatility tends to be followed by lower than normal volatility. Similar problems were observed with AGARCH.

And the Winner Is . . .

In each column in Table V, we shade the cell with the lowest RMSFE for that market among the ten forecasting models which we consider. The results are dramatic and telling. In all nine markets, the out-of-sample

RMSFE of the A-RLS model is the lowest of the 10 forecasting models. Besides having lower RMSFE's than GARCH in every market, in five of the nine the null that the two models forecasts are equally accurate is rejected at the .01 level in Diebold-Marino tests. In three of nine markets, the null that A-RLS and RLS are equally accurate is rejected at the .01 level and in six of nine at the .05 level. Across the nine markets, A-RLS's RMSFE averages 5.6% below GARCH's, 7.7% below EGARCH's, and 4.3% below RLS's.

ROBUSTNESS CHECKS

We next explore the robustness of these results. One issue is whether our results are sensitive to the chosen forecast accuracy measure: the root mean squared forecast error or RMSFE. Accordingly, in Table VI, we present results for an alternative measure of forecast accuracy, the mean absolute forecast error, which places less weight on extreme forecast errors. As shown there, the results largely mirror those for the RMSFE in Table V. Importantly, A-RLS out-forecasts all nine other models in all nine markets by this metric also.

Another issue is whether our results are specific to the forecast horizon of 40 trading days or eight weeks. We repeated all calculations for all models for forecast horizons of 10, 20, 80, and 120 days as well and present the results for 10 and 120 days in Tables VII and VIII, respectively.

TABLE VI

Out-of-Sample Mean Absolute Forecast Errors: All Models

	S&P 500	*10-year T-Bond*	*90-day Eurodollar*	*Yen/ Dollar*	*Boeing*	*GM*	*Int'l Paper*	*Mc-Donald's*	*Merck*
STD(10)	0.04051	0.03876	0.07874	0.03173	0.09654	0.07146	0.07774	0.07519	0.06796
STD(20)	0.03746	0.03482	0.07202	0.02751	0.08206	0.06263	0.06684	0.06444	0.05899
STD(40)	0.03727	0.03292	0.06755	0.02531	0.07319	0.05713	0.05805	0.05945	0.05326
STD(80)	0.03827	0.03340	0.06582	0.02428	0.07228	0.05453	0.05456	0.06009	0.05104
STD(120)	0.03800	0.03384	0.06535	0.02352	0.07232	0.05426	0.05527	0.06151	0.05194
EWMA	0.03581	0.03200	0.06614	0.02507	0.07359	0.05673	0.05883	0.05920	0.05274
GARCH	0.03571	0.03198	0.07163	0.02518	0.07350	0.05227	0.05714	0.05838	0.04827
RLS	0.03845	0.03167	0.06740	0.02435	0.06979	0.05125	0.05710	0.06014	0.04731
A-RLS	0.03296	0.02971	0.06260	0.02318	0.06692	0.04915	0.05090	0.05485	0.04607
AGARCH	0.03471	0.03282	0.07879	0.02633	0.07206	0.05141	0.06089	0.06004	0.04794
EGARCH	0.03790	0.03494	0.08771	0.02783	0.07486	0.05344	0.06152	0.06141	0.04935
OBS	8575	9337	6988	6988	8575	8575	8575	7566	8575

Notes. Mean absolute errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 40 trading days using the procedures listed in column 1 are reported. In each market, the lowest of the 11 mean absolute errors is shaded.

TABLE VII

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*	*Yen/ Dollar*	*Boeing*	*GM*	*Int'l Paper*	*Mc-Donald's*	*Merck*
STD(10)	0.07426	0.05753	0.12331	0.04884	0.15788	0.11701	0.12592	0.12391	0.10401
STD(20)	0.07186	0.05356	0.11710	0.04412	0.14325	0.11035	0.11871	0.11372	0.09462
STD(40)	0.07064	0.05246	0.11183	0.04301	0.13429	0.10435	0.11314	0.10796	0.08994
STD(80)	0.07175	0.05326	0.11079	0.04232	0.13127	0.10230	0.10996	0.11023	0.08964
STD(120)	0.07219	0.05445	0.11091	0.04218	0.13075	0.10229	0.10980	0.11322	0.09121
EWMA	0.06795	0.05064	0.10919	0.04195	0.13374	0.10275	0.11024	0.10663	0.08879
GARCH	0.06557	0.05067	0.11124	0.04258	0.12864	0.09857	0.10758	0.10472	0.08657
RLS	0.06850	0.05112	0.11009	0.04127	0.12805	0.09791	0.10921	0.10557	0.08644
A-RLS	0.06310	0.04923	0.10470	0.04004	0.12347	0.09701	0.10406	0.10269	0.08488
AGARCH	0.06449	0.05061	0.11197	0.04172	0.12834	0.09807	0.10874	0.10552	0.08655
EGARCH	0.06435	0.05075	0.16312	0.04200	0.12638	0.09814	0.10882	0.10515	0.08653
OBS	8495	9257	6908	6908	8495	8495	8495	7486	8495

Notes. Root mean squared errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 10 trading days using the procedures listed in column 1 are reported. In each market, the lowest of the 11 RMSFE is shaded.

TABLE VIII

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*	*Yen/ Dollar*	*Boeing*	*GM*	*Int'l Paper*	*Mc-Donald's*	*Merck*
STD(10)	0.07613	0.05571	0.11402	0.04311	0.13205	0.10780	0.11196	0.11794	0.09262
STD(20)	0.07127	0.04960	0.10055	0.03681	0.11099	0.09526	0.09757	0.10476	0.08005
STD(40)	0.06771	0.04565	0.08870	0.03253	0.09665	0.08551	0.08645	0.09677	0.07231
STD(80)	0.06518	0.04242	0.07973	0.02970	0.08877	0.07847	0.08117	0.09209	0.06796
STD(120)	0.06309	0.04031	0.07584	0.02953	0.08439	0.07408	0.08052	0.08961	0.06605
EWMA	0.06703	0.04504	0.08882	0.03258	0.09768	0.08588	0.08744	0.09654	0.07218
GARCH	0.05983	0.04157	0.08076	0.02800	0.08014	0.07005	0.08421	0.08616	0.05653
RLS	0.05782	0.03703	0.07326	0.02668	0.07160	0.06305	0.08328	0.08416	0.05648
A-RLS	0.05344	0.03653	0.07287	0.02579	0.07143	0.06327	0.07027	0.08127	0.05544
AGARCH	0.06985	0.05296	0.11925	0.03939	0.08948	0.07128	0.12360	0.09698	0.06000
EGARCH	0.05839	0.04417	0.09846	0.03002	0.08196	0.06719	0.12072	0.08760	0.05736
OBS	8495	9257	6908	6908	8495	8495	8495	7486	8495

Notes. Root mean squared errors for forecasts of the annualized standard deviation, *AS*, of daily returns over the next 120 trading days using the procedures listed in column 1 are reported. In each market, the lowest of the 11 RMSFE is shaded.

As shown there, the A-RLS model generally dominates all other models at these horizons also—the single exception being forecasts of GM's volatility over a 120-day horizon (the 80 day horizon also) where RLS's out-of-sample RMSFE is lower. Comparing RLS and GARCH, the former tends to forecast better than GARCH at the longer horizons, such as

80 and 120 days, while GARCH performs relatively better at short forecast horizons. Comparing GARCH and EGARCH, GARCH forecasts somewhat better at the longer horizons and EGARCH at the shorter horizons.

CONCLUSIONS

We have compared the forecasting ability of several volatility models—the historical standard deviation, an 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 in that the regression parameter estimates put more weight on older observations. They are also more stable across subsamples. 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.

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. In all nine markets at most horizons, the forecasting model with the lowest root mean squared forecast error and 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. In terms of forecast accuracy, it clearly dominates the widely accepted GARCH and EGARCH models.

BIBLIOGRAPHY

Akgiray, V. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. Journal of Business, 62, 55–80.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579–625.

Andersen, T. G., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. Journal of Empirical Finance, 6, 457–477.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74, 3–30.

Bollerslev, T., & Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. Journal of Econometrics, 73, 151–184.

Brailsford, T., & Faff, R. (1996). An evaluation of volatility forecasting techniques. Journal of Banking and Finance, 20, 419–438.

Brooks, C. (1998). Predicting stock index volatility: Can market volume help. Journal of Forecasting, 17, 59–80.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13, 253–263.

Dimson, E., & Marsh, P. (1990). Volatility forecasting without data snooping. Journal of Banking and Finance, 14, 399–421.

Ding, Z., & Granger, C. (1996). Modeling volatility persistence of speculative returns: A new approach. Journal of Econometrics, 73, 185–215.

Figlewski, S. (1997). Forecasting volatility. Financial Markets, Institutions, and Instruments 6, 1–88.

Hansen, P. R., & Lunde, A. (2001). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? Working paper 01-04, Brown University Department of Economics.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. Journal of Finance, 50, 507–528.

Lopez, J. (2001). Evaluating the predictive accuracy of volatility models. Journal of Forecasting, 20, 87–109.

Loudon, G. F., Watt, W. H., & Yadav, P. K. (2000). An empirical analysis of alternative parametric ARCH models. Journal of Applied Econometrics, 15, 117–136.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347–370.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models of conditional stock volatilities. Journal of Econometrics, 45, 267–290.

Poon, S. H., & Granger, C. (2003). Forecasting volatility in financial markets: A review. Journal of Economic Literature, 41, 478–539.
Riskmetrics. (1996). Riskmetrics Technical Document. Available at Riskmetrics.com.
Schmidt, P. (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, K., & Cho, D. (1995). The predictive ability of several models of exchange rate volatility. Journal of Econometrics, 69, 367–391.
Zumbach, G. (2003). Volatility processes and volatility forecast with long memory. Working paper, Consulting in Financial Research.