Volatility Forecasting With Range-Based EGARCH Models

Michael W. Brandt
Fuqua School of Business, Duke University, and NBER, Durham, NC 27708 (*mbrandt@duke.edu*)

Christopher S. Jones
Marshall School of Business, University of Southern California, Los Angeles, CA 90089 (*christopher.jones@marshall.usc.edu*)

We provide a simple, yet highly effective framework for forecasting return volatility by combining exponential generalized autoregressive conditional heteroscedasticity models with data on the range. Using Standard and Poor's 500 index data for 1983–2004, we demonstrate the importance of a long-memory specification, based on either a two-factor structure or fractional integration, that allows for some asymmetry between market returns and volatility innovations. Out-of-sample forecasts reinforce the value of both this specification and the use of range data in the estimation. We find substantial forecastability of volatility as far as 1 year from the end of the estimation period, contradicting the return-based conclusions of West and Cho and of Christoffersen and Diebold that predicting volatility is possible only for short horizons.

KEY WORDS: Exponential generalized autoregressive conditional heteroscedasticity; High-low volatility estimator; Long memory in volatility; Multifactor volatility.

1. INTRODUCTION

The volatility of asset returns is time-varying and predictable, but forecasting the future level of volatility is difficult for at least three reasons. First, volatility forecasts are sensitive to the specification of the volatility model. In particular, it is important to strike the right balance between capturing the salient features of the data and overfitting the data. Second, correctly estimating the parameters of a volatility model can be difficult, because volatility is not observable. The further the estimated parameters are from the true parameters, the worse the volatility forecasts. Third, volatility forecasts are anchored at noisy proxies or estimates of the current level of volatility. Even with a perfectly specified and estimated volatility model, forecasts of future volatility inherit and potentially even amplify the uncertainty about the current level of volatility. The aim of this article is to provide a volatility forecasting framework that addresses these three issues.

We combine exponential generalized autoregressive conditional heteroscedasticity (EGARCH) models with data on the daily range, defined as the difference between the highest and lowest log asset prices recorded throughout the day. Although certainly not the only class of models that do so, EGARCH models capture the most important stylized features of stock return volatility, namely time-series clustering, negative correlation with returns, lognormality, and under certain specifications, long memory (e.g., Andersen, Bollerslev, Diebold, and Ebens 2001). The other ingredient to our approach is the use of range data, which has several advantages relative to the use of absolute or squared return data. The range is a much more efficient volatility proxy, a fact known at least since the work of Parkinson (1980) and recently formalized by Andersen and Bollerslev (1998). Furthermore, Alizadeh, Brandt, and Diebold (2002) established that the distribution of the log range conditional on volatility is approximately Gaussian, making range-based estimation of volatility models highly effective, yet surprisingly simple. Together, these two advantages of the range lead to more precise range-based estimates of the model parameters and of the current level of volatility.

Using daily Standard and Poor's 500 (S&P500) index data for 1983–2004, we document substantial gains in estimation efficiency from using range data instead of, or in addition to, return data. Furthermore, the information added by the range allows us to draw sharp distinctions between competing models that are indistinguishable when only returns are used in the estimation. Our results clearly illustrate our earlier claim that the quality of volatility forecasts depends crucially on a well-specified model and on the use of informative data. In particular, we find that two-factor range-based EGARCH models dominate, both in-sample and out-of-sample, the extensive set of model and data combinations that we consider. Finally, models that incorporate volatility asymmetries, or negative correlations between returns and volatility innovations, generally outperform models that do not.

Of course, we are not the first to point out the benefits of the EGARCH framework, of multifactor volatility models or of range-based volatility estimation. Specifically, EGARCH models have been advocated by Nelson (1989, 1991), Pagan and Schwert (1990), and Hentschel (1995), among others. Related work on multifactor volatility models includes that of Engle and Lee (1999), Gallant, Hsu, and Tauchen (1999), Alizadeh et al. (2002), Chernov, Ghysels, Gallant, and Tauchen (2003), Barndorff-Nielsen and Shephard (2001), and Bollerslev and Zhou (2002). Finally, the literature on range-based volatility estimation includes work of Parkinson (1980), Garman and Klass (1980), Schwert (1990), Gallant et al. (1999), Yang and Zhang (2000), Alizadeh et al. (2002), Brandt and Diebold (2006), and Chou (2005), among others. In particular, Chou (2005) also considered autoregressive volatility models

Journal of Business & Economic Statistics
October 2006, Vol. 24, No. 4
DOI 10.1198/073500106000000206

that involve the range. However, that approach is very different from ours in several respects. First, Chou's model involves the lagged range instead of the lagged log range. Our logarithmic specification allows us to apply the approximate normality result of Alizadeh et al. (2002) to dramatically simplify the estimation. Second, Chou's model describes the dynamics of the conditional mean of the range, whereas our model explicitly describes the dynamics of the conditional return volatility. Finally, Chou focused on estimation and in-sample prediction, whereas our interest lies primarily in model specification and out-of-sample forecasting.

Our article builds on this work by demonstrating the *joint* effectiveness for volatility forecasting of multifactor EGARCH models and range-based volatility estimation. A major contribution of our article is to provide new empirical evidence that volatility is predictable even over horizons as long as 1 year. This result addresses a paradox in the literature on the long-run dependence in stock market volatility. Studies such as those of Baillie, Bollerslev, and Mikkelsen (1996), Andersen et al. (2001), and Calvet and Fisher (2002) have documented strong evidence for long memory in volatility. This is reinforced by the findings of Bollerslev and Mikkelsen (1996) and Christoffersen, Jacobs, and Wang (2005) that option prices display evidence of long memory. At the same time, the forecasting results of West and Cho (1995) and Christoffersen and Diebold (2000) suggest that volatility predictability is a short-horizon phenomenon. Our results demonstrate striking forecastability in equity index volatilities at long horizons using easily obtainable data on the daily range. We attribute our results to the combination of a less misspecified volatility model and a more informative volatility proxy.

The article proceeds as follows. Section 2 provides some theoretical background for the use of the range as a volatility proxy. It also describes the data. Section 3 outlines the range-based EGARCH modeling framework. Section 4 summarizes the in-sample fit of the models, and Section 5 presents the out-of-sample forecasting results. Section 6 concludes.

2. THE RANGE AS VOLATILITY PROXY

2.1 Theoretical Background

Assume that the log stock price L follows the martingale process $dL_s = \sqrt{V_s}\,dW_s$, where V_s is possibly stochastic but is assumed continuous, at least within each period. Let Q_t denote the quadratic variation of L_t and define the discrete-time process

$$h_t = \sqrt{Q_t - Q_{t-1}} \tag{1}$$

for $t = 1, 2, 3, \ldots, T$, where h_t is commonly referred to as *integrated* volatility. Note that although the martingale assumption on the log stock price process can be relaxed (see, e.g., footnote 12 of Alizadeh et al. 2002), it is quite beneficial from a statistical perspective. In our analysis that follows, with the exception of Section 2.3, we work exclusively with daily returns and ranges. In this setting, assuming that a zero drift is known to cause only a very small bias, to the extent that the true drift differs slightly from 0. The advantage is that we greatly reduce the mean squared error relative to relying on some noisy drift estimator.

Under the assumption that V_s is constant throughout the period, Alizadeh et al. (2002) showed that the log range, defined as

$$D_t = \ln\Big(\max_{s\in[t-1,t]} L_s - \min_{s\in[t-1,t]} L_s\Big), \tag{2}$$

is to a *very good* approximation distributed as

$$D_t \sim \mathrm{N}[.43 + \ln h_t, .29^2], \tag{3}$$

where $\mathrm{N}[m, v]$ denotes a Gaussian distribution with mean m and variance v. A more detailed description of this result is provided in the Appendix. Thus, the log range is a *noisy* linear proxy of log volatility $\ln h_t$.

To relate the range data to return data, note that under the foregoing assumptions, the daily log absolute return, defined as

$$\ln|R_t| = \ln|L_t - L_{t-1}|, \tag{4}$$

is also a noisy linear proxy of log volatility. According to the results of Alizadeh et al. (2002), the log absolute return has a mean of $-.64 + \ln h_t$ and a variance of 1.11^2. However, the distribution of the log absolute return is far from Gaussian, a well-known problem in the literature on estimating stochastic volatility models (see, e.g., Jacquier, Polson, and Rossi 1994; Andersen and Sørensen 1997).

The fact that both the log range and the log absolute return are linear log volatility proxies (with the same loading of 1), but that the standard deviation of the log range is about one-quarter of the standard deviation of the log absolute return, makes clear that the range is a much more informative volatility proxy. It also makes sense of the finding of Andersen and Bollerslev (1998) that the daily range has approximately the same informational content as sampling intradaily returns every 4 hours.

It is well known that the range suffers from a discretization bias because the highest (lowest) stock price observed at discrete points in time is likely to be lower (higher) than the true maximum (minimum) of the underlying diffusion process. It follows that the observed range is a downward-biased estimate of the true range (which in turn is a noisy proxy of volatility). This is a particular problem when prices are unobserved over a significant portion of the day, as is the case for stock market indexes. Although Rogers and Satchell (1991) devised a correction of the observed range that virtually eliminates this bias, their method requires observation of both an opening price and a closing price each day, and in our data the reported opening price is most often simply a repeat of the previous day's closing price (more then 90% of the time in the second half of our sample, which we find indicative of some kind of recording bias). Therefore, our observed range is computed over the set of prices observed throughout the day in addition to the close of the previous day. We believe that the problems with observed range data make our results conservative, in that the results for a less-biased range are likely to be even better.

Even during the trading day, infrequent trading of the component stocks may cause an additional downward bias of the range. In less-liquid markets with infrequent price observations, the extremes of the underlying value processes are unlikely to be observed in a transaction. This effect might be offset, at least in part, by widening bid–ask spreads during periods of low liquidity. Because high prices are more likely to reflect transactions at the ask and low-price transactions at the bid, widening spreads will cause an upward bias in the range. Our results might be improved on by accounting for the impact of these biases, but we would expect the effects of doing so to be small. This is because all of our results are for a value-weighted stock market index, in which the most liquid stocks comprise the bulk of the portfolio and where at least some of the bid–ask effect is diversified away. Accounting for such biases would likely be more relevant in applications with individual stock data.

A final, and potentially important, source of bias might arise from conditional nonnormality in returns, such as that arising from estimation error or from replacing the underlying Brownian motion with a Levy-type process. Exploring properties of the range in these cases is interesting but beyond the scope of this article. In part because of this issue and the other potential biases discussed earlier, our forecast evaluations will primarily be regression-based, so that the regression intercept and slope can "absorb" some of these biases. Nevertheless, unadjusted forecasts do remarkably well, and we believe that our results could only be strengthened by explicitly accounting for nonnormality.

Except for the model of Chou (2005), GARCH-type volatility models rely on squared or absolute returns (which have the same information content) to capture variation in the conditional volatility h_t. Because the range is a more informative volatility proxy, it makes sense to consider range-based GARCH models, in which the range is used in place of squared or absolute returns to capture variation in the conditional volatility. This is particularly true for the EGARCH framework of Nelson (1991), which describes the dynamics of log volatility (of which the log range is a linear proxy). However, before we proceed along these lines, we first briefly describe the empirical relationship between the range data and return data.

2.2 Data

We collect daily return and range observations for the S&P500 index from January 2, 1962, to December 31, 2004 (10,787 observations). Over this period, the return-based volatility is 93 basis points per day, or 15.2% per year. The annualized average return is 6.9%. In addition, the returns are substantially left-skewed and fat-tailed, with a skewness of -1.5 and an excess kurtosis of 41.3, providing strong evidence for the presence of conditional nonnormality and/or heteroscedasticity.

As is all too often the case, the data are not as clean as the theory suggests. Specifically, the results of Alizadeh et al. (2002) summarized earlier imply that the average difference between the log absolute returns and log ranges should be approximately -1.07. This implication appears to be violated in the first half of our sample.

Figure 1(a) plots the differences between the annual average log absolute returns and annual average log ranges, together with two–standard error bands. [Here and throughout the article, we compute heteroscedasticity and serial correlation–adjusted standard errors using the covariance matrix estimator of Newey and West (1987) with five lags. The results are not particularly sensitive to the choice of lags.] The horizontal line marks the theoretical difference of -1.07. It is clear from this plot that the relationship between the two log volatility proxies experienced an abrupt change somewhere between 1982 and 1983. Before 1982–1983, the average difference between the log absolute returns and log ranges was significantly less than its theoretical value (about -1.5). Since then, the average difference has been only slightly above its theoretical value. Furthermore, the correlation between the two log volatility proxies changed at the same time. Figure 1(b) plots the correlation between the log volatility proxies estimated each year. The correlation averaged close to .4 before 1982–1983 and closer to .6 since then.

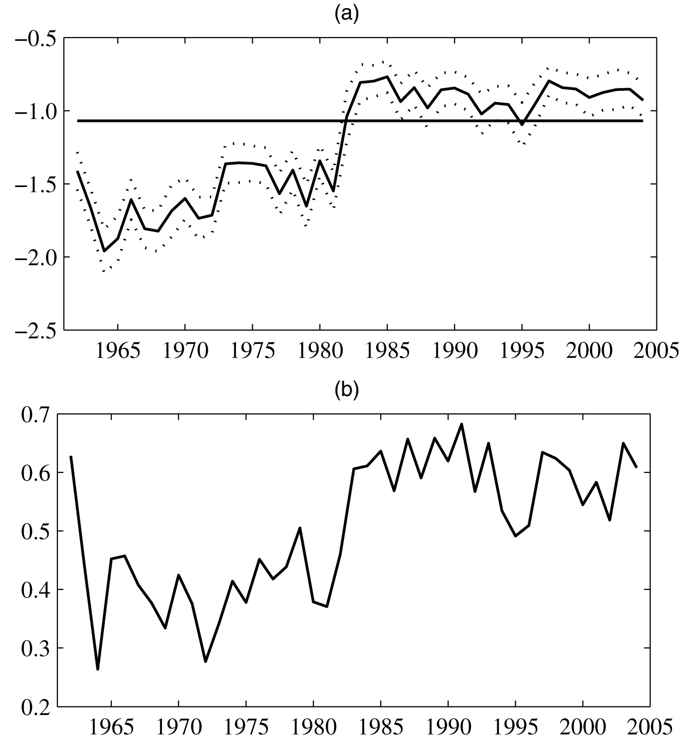

Figure 1. The Time-Varying Relationship Between Log Range and Absolute Log Returns. (a) Average log absolute return minus average log range. (b) Correlation between log absolute return and log range.

The reason for this structural break in the data is unclear. Nonetheless, an empirical analysis based on the whole sample may be quantitatively biased as well as qualitatively misleading. Therefore, all of the results in this article are based on the postbreak subsample from January 1, 1983 to December 31, 2004 (5,552 observations). For this subsample, the annualized mean and volatility of returns are 10.3% and 17.3%.

2.3 A Nonparametric Look at Range-Based Volatility Predictability

West and Cho (1995) and Christoffersen and Diebold (2000) both argued that volatility predictability is a short-horizon phenomenon. Christoffersen and Diebold proposed a nonparametric test of volatility forecastability at different horizons. They constructed a sequence of indicator variables based on nonoverlapping τ-day returns, $R_{t,t+\tau}$. This "hit sequence" is defined as

$$H_{t,t+\tau} = \begin{cases} 1 & \text{if } |R_{t,t+\tau}| < n\hat{\sigma} \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $n = 1.5$ or 2, and thus takes the value of 1 if the τ-day return lies within n unconditional standard deviations of 0. They then regressed, in a logit framework, this hit sequence on lagged volatilities measures (squared returns in their case). p-values of the joint significance of the regressors generally indicate the absence of volatility predictability at horizons greater than 5–10 days.

To give a first indication of the difference between return-based and range-based volatility forecasting, we repeat the tests of Christofferson and Diebold both with lagged squared returns and lagged ranges on the right side of the regression. In a first set of regressions, the right side variables include either five lags of squared daily returns or five lags of daily log ranges. In another set of regressions we aggregate the longer history of returns or ranges into a single predictor by averaging squared returns or ranges over the past 22 days.

The results of these two sets of regressions are presented graphically in Figure 2. The left and right pairs of columns correspond to the first and second set of regressions. Each plot shows the p-value of joint significance for horizons up to 40 days. The horizontal dashed line marks 5% significance. The results in the top row, in which lagged squared returns are used as volatility forecasters, roughly confirm the findings of Christoffersen and Diebold (2000). The p-values are reliably $<5\%$ only for horizons up to 10 days, sometimes less. Using a different set of volatility predictors changes the results dramatically, however. In the second row, in which lagged ranges are used as regressors, the long-horizon p-values are much smaller. Volatility appears to be predictable as far as 30 days into the future. The stark difference between the two sets of results can be reconciled by the much lower signal-to-noise ratio of the range compared with squared returns, as noted by Alizadeh et al. (2002).

As Andersen and Bollerslev (1998) demonstrated, more informative volatility proxies also aid in the detection and evaluation of volatility predictability. Based on this intuition, we repeat the analysis using the range to construct the hit sequence on the left side of the regressions. Specifically, instead of using a return-based hit sequence, we instead use the average log range,

$$H'_{t,t+\tau} = \begin{cases} 1 & \text{if } \dfrac{1}{\tau}\displaystyle\sum_{s=1}^{\tau} D_{t+s} < D^* \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The hit criterion is one-sided because of the positive relationship between volatility and average log ranges. The cutoff D^* is chosen as either the 90th or 95th percentile of the empirical distribution of the average log range for each forecast horizon.

The results are presented in Figure 3. Looking first at the first row with squared returns as predictors, we note that the long-horizon p values are generally lower than those in Figure 2, and the null of zero forecast power can be rejected for horizons as long as 15–20 days. This pattern is even more dramatic for the range-based tests in the second row, where with only few exceptions, the p values are close to 0 out to 30 days.

These results suggests that volatility forecastability exists for horizons up to at least 30 days and that the range considerably enhances our ability to forecast long-horizon volatility. The next logical step is to specify volatility models that use the range.

3. TWO–FACTOR EGARCH MODELS

We consider variants of the EGARCH framework introduced by Nelson (1991). In general, the EGARCH(1, 1) model

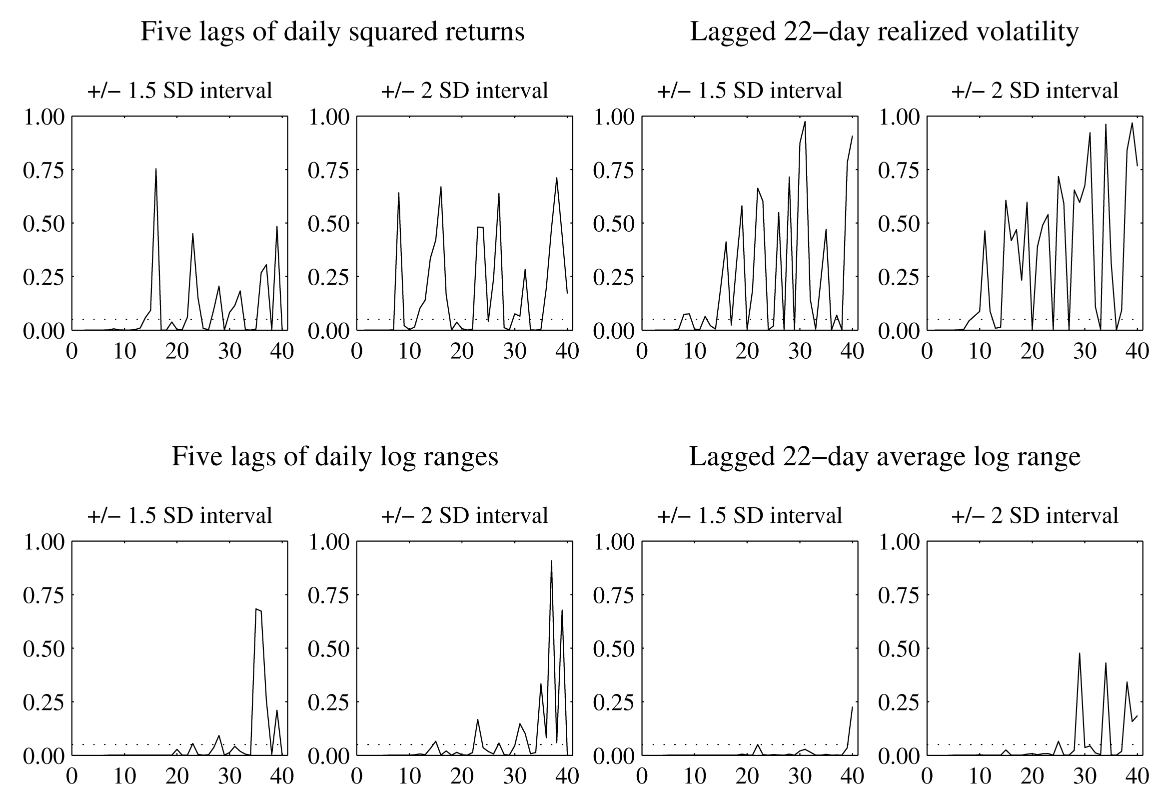

Figure 2. Nonparametric Tests of Volatility Forecastability Using Return-Based Hit Sequences.

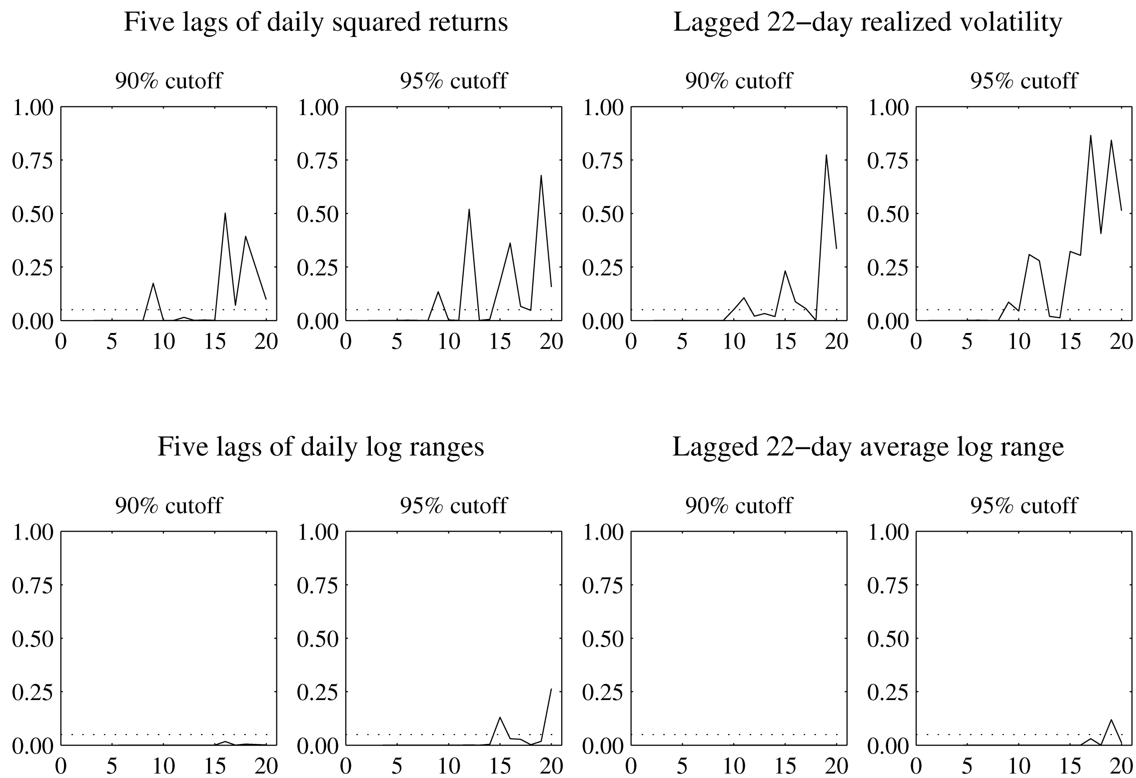

Figure 3. Nonparametric Tests of Volatility Forecastability Using Range-Based Hit Sequences.

performs comparably to the GARCH(1, 1) model of Bollerslev (1986). However, for stock indices, the in-sample evidence reported by Hentschel (1995) and the forecasting performance presented by Pagan and Schwert (1990) show a slight superiority of the EGARCH specification over plain GARCH. This is due in large part to EGARCH's accommodation of asymmetric volatility (often called the "leverage effect," which refers to one of the explanations of asymmetric volatility), where increases in volatility are associated more often with large negative returns than with equally large positive returns. Although the leverage effect can be generated by other members of the GARCH family (e.g., Glosten, Jagannathan, and Runkle 1993), we adopt the exponential specification because of its familiarity and the simplicity with which volatility asymmetry can be introduced. We make no claim about the superiority of exponential range-based GARCH models.

3.1 Return-Based Models

Before we describe the range-based EGARCH models, we present their traditional return-based counterparts, which we use as benchmark models. We consider a one-factor return-based EGARCH(1, 1) model (EGARCH1). Daily log returns are conditionally Gaussian,

$$R_t \sim \mathrm{N}[0, h_t^2], \tag{7}$$

with a conditional volatility h_t that changes from one day to the next according to

$$\ln h_t - \ln h_{t-1} = \kappa(\theta - \ln h_{t-1}) + \phi X_{t-1}^R + \delta R_{t-1}/h_{t-1}, \tag{8}$$

where

$$X_{t-1}^R = \frac{|R_{t-1}/h_{t-1}| - \mathrm{E}|R_{t-1}/h_{t-1}|}{\sqrt{\operatorname{var}|R_{t-1}/h_{t-1}|}} \tag{9}$$

is an innovation that depends on the standardized deviation of the absolute return from its expected value. Our notation, which is different from that of Nelson (1991), assigns a conceptually separate role to each parameter. θ is the long-run mean of the volatility process, κ is the speed of mean reversion, ϕ measures the sensitivity to lagged absolute returns, and δ is an asymmetry parameter that allows volatility to be affected differently by positively and negatively lagged returns. We consider the symmetric model with $\delta = 0$ as a special case.

Recent studies such those of Baillie et al. (1996) and Andersen et al. (2001) have found that volatility dynamics display pronounced long memory, so we consider two classes of models capable of generating this type of behavior. As Gallant et al. (1999) noted, multifactor models are attractive because they can replicate the autocorrelation structure of a long-memory process within a standard $I(0)$ environment. Therefore, following Engle and Lee (1999), we consider a two-factor return-based EGARCH model (EGARCH2), in which the conditional volatility dynamics in (8) are replaced with

$$\begin{aligned}
&\ln h_t - \ln h_{t-1} \\
&\quad = \kappa_h(\ln q_{t-1} - \ln h_{t-1}) + \phi_h X_{t-1}^R + \delta_h R_{t-1}/h_{t-1}, \\
&\ln q_t - \ln q_{t-1} \\
&\quad = \kappa_q(\theta - \ln q_{t-1}) + \phi_q X_{t-1}^R + \delta_q R_{t-1}/h_{t-1},
\end{aligned} \tag{10}$$

where $\ln q_t$ can be interpreted as a slowly moving stochastic mean around which log volatility $\ln h_t$ makes large but transient deviations (with a process determined by κ_h, ϕ_h, and δ_h).

Here θ, κ_q, ϕ_q, and δ_q determine the long-run mean, speed of mean reversion, sensitivity of the stochastic mean to lagged absolute returns, and asymmetry of absolute return sensitivity. This model can be seen to nest the previous one-factor specification by setting $\kappa_q = 1$, $\phi_q = 0$, and $\delta_q = 0$. As with the EGARCH1 model, we consider the special cases of a fully symmetric model with $\delta_q = \delta_h = 0$, as well as a partly symmetric model with only $\delta_q = 0$. The motivation for the partly symmetric model is the finding of Engle and Lee (1999) that asymmetric volatility is primarily a short-run phenomenon.

The final class of return-based model that we consider is the fractionally integrated EGARCH model (FIEGARCH) of Bollerslev and Mikkelsen (1996). As Barndorff-Nielsen and Shephard (2001) pointed out, both fractional integration and multifactor $I(0)$ models are alternative approaches for generating long memory, so it is of some interest to see which model performs better in practice. The FIEGARCH model that we consider is

$$(1-\omega L)(1-L)^d(\ln h_t - \theta) = \phi X^R_{t-1} + \delta R_{t-1}/h_{t-1}, \quad (11)$$

where X^R is as given earlier and the parameters θ, ϕ, and δ have similar interpretations as before. The other two parameters, ω and d, determine the persistence properties of log volatility, where d is the fractional differencing parameter. As with the other specifications, we consider the special case in which $\delta = 0$.

To understand the role of volatility proxies in GARCH-type models, we interpret (8) as a standard AR process for log volatility. Ignoring the asymmetry term, the log volatility innovations $\ln h_t - \mathrm{E}_{t-1}[\ln h_t]$, which are unobservable, are proxied for by the demeaned and standardized lagged absolute returns X^R_{t-1}. The intuition is that when the lagged absolute return is large (small) relative to the lagged level of volatility, volatility is likely to have experienced a positive (negative) innovation. Unfortunately, as we explained earlier, the absolute return is a rather noisy proxy of volatility, suggesting that a substantial part of the volatility variation in GARCH-type models is driven by *proxy noise* as opposed to true information about volatility. In other words, the noise in the volatility proxy introduces noise in the inferred volatility process. In a volatility forecasting context, this noise in the inferred volatility process deteriorates the quality of the forecasts through less precise parameter estimates and, more importantly, through less precise estimates of the current level of volatility to which the forecasts are anchored.

3.2 Range-Based Models

Because the range is more informative about the true volatility than the absolute return, it is sensible to consider EGARCH models in which the demeaned and standardized log range serves as a proxy for the log volatility innovation. (Standardizing the volatility innovation term has no effect on the model dynamics, but it allows us to directly compare the parameter estimates across the return- and range-based specifications.) In particular, we consider the one-factor range-based model (REGARCH1)

$$D_t \sim \mathrm{N}[.43 + \ln h_t, .29^2], \quad (12)$$

$$\ln h_t - \ln h_{t-1} = \kappa(\theta - \ln h_{t-1}) + \phi X^D_{t-1} + \delta R_{t-1}/h_{t-1}, \quad (13)$$

where the innovation is now defined as the standardized deviation of the log range from its expected value,

$$X^D_{t-1} = \frac{D_{t-1} - .43 - \ln h_{t-1}}{.29}. \quad (14)$$

We also consider the two-factor range-based model (REGARCH2), in which the conditional volatility dynamics in (13) is replaced with

$$\begin{aligned} &\ln h_t - \ln h_{t-1} \\ &\quad = \kappa_h(\ln q_{t-1} - \ln h_{t-1}) + \phi_h X^D_{t-1} + \delta_h R_{t-1}/h_{t-1}, \\ &\ln q_t - \ln q_{t-1} \\ &\quad = \kappa_q(\theta - \ln q_{t-1}) + \phi_q X^D_{t-1} + \delta_q R_{t-1}/h_{t-1}. \end{aligned} \quad (15)$$

Finally, we introduce a fractionally integrated range-based model (FIREGARCH),

$$(1-\omega L)(1-L)^d(\ln h_t - \theta) = \phi X^D_{t-1} + \delta R_{t-1}/h_{t-1}. \quad (16)$$

Because the range does not reflect the direction of the price movement, we still use lagged returns to generate volatility asymmetry. As in the return-based specifications, we also consider symmetric and partially symmetric versions of these models.

4. IN–SAMPLE FIT

We estimate a total of 14 model specifications, consisting of 7 return-based models and the corresponding 7 range-based models, by maximum likelihood. Specifically, we estimate the EGARCH2 and REGARCH2 specifications, along with their symmetric and partially symmetric versions. We also fit the EGARCH1 and REGARCH1 specifications, along with their symmetric versions, which can be considered as special cases of the two-factor models with the stochastic long-run mean process q_t set to a constant. Finally, we consider return-based and range-based fractionally integrated models (FIEGARCH and FIREGARCH), both asymmetric and symmetric.

For return-based models, likelihoods are calculated using

$$R_t \sim \mathrm{N}[0, h_t^2], \quad (17)$$

whereas range-based models rely on the approximate result,

$$D_t \sim \mathrm{N}[.43 + \ln h_t, .29^2]. \quad (18)$$

Volatility dynamics follow the processes given in Sections 3.1 and 3.2. Standard errors are computed using the outer product of gradients estimator.

Several regularities emerge from the estimates presented in Table 1. First, the standard errors are usually smaller for the range-based estimates than for the corresponding return-based estimates (e.g., REGARCH2 vs. EGARCH2), reflecting the greater precision of the range as a volatility proxy. This reflects the same informational advantage of using the range that has been a primary motivation of all previous work in range-based volatility estimation. Second, the parameters of the two-factor models are generally highly significant, indicating that the volatility dynamics contain very distinct long-run and short-run components. In particular, the two-factor models display high long-run persistence through the coefficient κ_q, but also exhibit quick short-run reversion through much larger values

Table 1. Full-Sample Maximum Likelihood Estimates

	κ or κ_q	θ	ϕ or ϕ_q	δ or δ_q	κ_h	ϕ_h	δ_h	d	ω
EGARCH1	.0222	−4.4826	.0400	−.0448					
	(.0017)	(.0271)	(.0022)	(.0023)					
EGARCH1,	.0156	−4.3949	.0530						
symmetric	(.0021)	(.0550)	(.0010)						
EGARCH2	.0097	−4.4154	.0275	−.0175	.3275	.0248	−.1058		
	(.0013)	(.0475)	(.0024)	(.0029)	(.0375)	(.0040)	(.0050)		
EGARCH2,	.0046	−4.2626	.0207		.1054	.0632			
symmetric	(.0013)	(.1063)	(.0033)		(.0254)	(.0020)			
EGARCH2,	.0049	−4.3731	.0208		.1467	.0347	−.0815		
part symmetric	(.0010)	(.0718)	(.0020)		(.0130)	(.0029)	(.0035)		
FIEGARCH		−3.7829	.0548	−.0573				.5247	.4985
		(.1499)	(.0035)	(.0037)				(.0228)	(.0553)
FIEGARCH		−3.3129	.0551					.5935	.5551
symmetric		(.3821)	(.0026)					(.0361)	(.0688)
REGARCH1	.0177	−4.9148	.0302	−.0277					
	(.0011)	(.0191)	(.0008)	(.0009)					
REGARCH1,	.0106	−5.0125	.0331						
symmetric	(.0010)	(.0302)	(.0008)						
REGARCH2	.0119	−4.9304	.0273	−.0137	.3772	.0238	−.0538		
	(.0010)	(.0254)	(.0009)	(.0012)	(.0287)	(.0014)	(.0017)		
REGARCH2,	.0021	−5.0387	.0122		.0495	.0354			
symmetric	(.0006)	(.0609)	(.0014)		(.0068)	(.0009)			
REGARCH2,	.0063	−4.9985	.0214		.1400	.0288	−.0423		
part symmetric	(.0009)	(.0351)	(.0010)		(.0080)	(.0011)	(.0013)		
FIREGARCH		−4.3782	.0344	−.0320				.6609	.3749
		(.1534)	(.0013)	(.0012)				(.0146)	(.0329)
FIREGARCH		−5.1997	.0293					.6870	.5538
symmetric		(.1965)	(.0011)					(.0213)	(.0346)

NOTE: Maximum likelihood estimates and standard errors (in parentheses) of return-based EGARCH models with likelihood function $R_t \sim \mathrm{N}[0, h_t^2]$ and range-based REGARCH models with likelihood function $D_t \sim \mathrm{N}[.43 + \ln h_t, .29^2]$. The conditional volatility follows the single-factor process (8) or (13), the two-factor process (10) or (15), or the fractionally integrated process (11) or (16). Daily data on the S&P500 index are for January 1983–December 2004 (5,552 observations).

of κ_h. Third, the fractional differencing parameter in the FIEGARCH and FIREGARCH models is between .5 and .7 in all cases, consistent with the value of .63 that Bollerslev and Mikkelsen (1996) reported for a similar specification. Finally, volatility asymmetry (the leverage effect) appears to be important, because the asymmetry parameters δ are generally significant. For the two-factor models, however, past returns appear to have a much greater effect in the short run than they do in the long run, because $\delta_h < \delta_q$. In fact, for the EGARCH2 specification, there is no significant leverage effect in the long-run component of volatility (δ_q is not significantly nonzero).

These observations are reinforced by Table 2, which presents a variety of model selection criteria and diagnostic tests for each specification. Given the fitted volatilities h_t from either the return-based or range-based estimates, we define both the return-based errors, $\eta_t^R = R_t^2 - h_t^2$, and the range-based errors, $\eta_t^D = D_t - .43 - \ln h_t$. We then use these two sets of errors to compute adjusted R^2's, Akaike information criteria (AICs), and Schwartz criteria (SCs). We compute unadjusted R^2's as $1 - (\eta^R)'(\eta^R)/[(R^2)'(R^2)]$ and $1 - (\eta^D)'(\eta^D)/[(D^2)'(D^2)]$ and then perform the usual adjustment. We calculate the AICs as $\ln(\eta'\eta/T) + 2K/T$, where K is the number of model parameters and T is the number of observations. Finally, we compute the SCs as $\ln(\eta'\eta/T) + K\ln(T)/T$. Regardless of whether we use range or return data, high adjusted R^2's and low AICs and SCs indicate support for a model.

The most consistent and perhaps least surprising result in Table 2 is that the range-based models explain ranges better, whereas the return-based models explain squared returns better. The less foreseeable result is that two-factor and fractionally integrated models with some asymmetry are generally favored over their one-factor or symmetric counterparts. (There are some exceptions for the return-based SCs.) Finally, the model selection criteria exhibit greater dispersion across different specifications when they are computed with range data as opposed to return data. This means that the information contained in the range allows us to draw distinctions between competing models that are virtually indistinguishable based on returns.

The table also shows for each specification the p value of Engle's (1982) ARCH–LM test, performed with 10 lags, to check whether ARCH effects remain in the residuals η_t^D and $\eta_t^{R*} = R_t/h_t$. Similar to the model selection criteria, the range-based tests are substantially more powerful than the return-based test. In particular, the range-based tests reinforce the importance of both multiple factors or fractional integration and at least partial asymmetries.

The final three columns of the table report t-statistics for three return-based specification tests proposed by Engle and Ng (1993). These tests are designed to reveal various violations of the assumption that η_t^{R*} is iid $\mathrm{N}[0, 1]$. The results must be interpreted with some caution because violations can be due

Table 2. In-Sample Diagnostics

	Adjusted R^2	AIC	SC	ARCH LM(10) p value	Sign bias t-statistic	−Size bias t-statistic	+Size bias t-statistic
Return-based diagnostics							
EGARCH1	.040	−14.408	−14.403	.980	2.186	−1.213	−4.846
EGARCH1, symmetric	.031	−14.398	−14.395	.491	3.187	−1.314	−6.941
EGARCH2	.040	−14.407	−14.398	.657	−.275	−.540	−.324
EGARCH2, symmetric	.035	−14.402	−14.396	.690	3.204	−1.509	−7.404
EGARCH2, part symmetric	.042	−14.410	−14.403	.757	.777	−.816	−1.855
FIEGARCH	.044	−14.411	−14.405	.680	1.542	−.836	−4.702
FIEGARCH, symmetric	.037	−14.404	−14.400	.596	3.175	−1.442	−7.137
REGARCH1	.025	−14.387	−14.383	.377	1.910	−1.285	−4.558
REGARCH1, symmetric	.016	−14.379	−14.375	≤.001	2.537	−1.603	−5.757
REGARCH2	.024	−14.386	−14.378	.288	.680	−.934	−2.093
REGARCH2, symmetric	.016	−14.378	−14.372	≤.001	2.523	−1.598	−5.836
REGARCH2, part symmetric	.024	−14.386	−14.379	.249	1.237	−1.087	−3.086
FIREGARCH	.026	−14.388	−14.382	.365	1.671	−1.137	−4.547
FIREGARCH, symmetric	.016	−14.378	−14.374	≤.001	2.594	−1.637	−5.727
Range-based diagnostics							
EGARCH1	.367	−1.262	−1.258	≤.001			
EGARCH1, symmetric	.355	−1.219	−1.215	≤.001			
EGARCH2	.379	−1.296	−1.288	≤.001			
EGARCH2, symmetric	.357	−1.225	−1.219	≤.001			
EGARCH2, part symmetric	.381	−1.295	−1.288	≤.001			
FIEGARCH	.366	−1.260	−1.254	≤.001			
FIEGARCH, symmetric	.354	−1.219	−1.215	≤.001			
REGARCH1	.397	−1.664	−1.660	.023			
REGARCH1, symmetric	.378	−1.634	−1.630	≤.001			
REGARCH2	.406	−1.679	−1.671	.077			
REGARCH2, symmetric	.380	−1.637	−1.631	≤.001			
REGARCH2, part symmetric	.405	−1.677	−1.670	.039			
FIREGARCH	.399	−1.668	−1.662	.096			
FIREGARCH, symmetric	.399	−1.668	−1.662	.096			

NOTE: In-sample diagnostics for the maximum likelihood estimates in Table 1. Return- and range-based volatility errors are defined as $\eta_t^R = R_t^2 - h_t^2$ and $\eta_t^D = D_t - .43 - \ln h_t$.

either to neglected GARCH effects or to conditional nonnormality in returns. Nevertheless, we include these tests for completeness.

For the first test, the so-called "sign bias" test, we regress $(\eta_t^{R*})^2$ on a constant and a dummy variable d_t that is equal to 1 if $R_t < 0$ and 0 otherwise. For the other two tests, the negative size bias test and the positive size bias test, we perform similar regressions, except with $d_t \times R_t$ and $(1 - d_t) \times R_t$ in place of d_t. In all cases, the t-statistic is for a test that the explanatory variable enters the regression significantly. For most specifications, at least one t-statistic is >2 in absolute value, indicating the violation of one of the specification tests. Only the return-based two-factor models with asymmetries pass all three specification tests.

5. OUT–OF–SAMPLE VOLATILITY FORECASTS

Ultimately, the usefulness of volatility models depends on their ability to accurately forecast future volatility. Therefore, we perform a variety of out-of-sample forecasting exercises to determine which specification performs best by this criterion. The forecasts are based on parameter estimates from rolling samples with fixed sample size of 1,000 days, as opposed to an expanding sample. We believe that the choice of a fixed sample size is conservative, in that it favors traditional single-factor models, for two reasons. First, single-factor models have fewer parameters and thus require less data to estimate precisely. Second, if single-factor models are misspecified, they have a better chance of explaining the volatility dynamics over relatively short rolling samples than over the whole sample period. Finally, results computed using 1,500 observations do not appear to be qualitatively different from those reported herein.

For every date $t \geq 1{,}000$ in the sample, we estimate the parameters of each specification over the 1,000 days of data up to and including date t. To insure reasonable results, the θ parameter is required to be within the interval $(-6, -3)$, corresponding to a long-run volatility between .25% and 5% per day. We then form forecasts of the integrated log volatility over the period $t + \tau_0$ to $t + \tau_1$. To measure the accuracy of these forecasts, we compare them with the following realized volatility measures:

- Average log absolute return: $\frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} \ln |R_{t+\tau}|$
- Average log range: $\frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} D_{t+\tau}$
- Realized standard deviation: $\sqrt{\frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} R_{t+\tau}^2}$
- Realized variance: $\frac{1}{\tau_1 - \tau_0 + 1} \sum_{\tau=\tau_0}^{\tau_1} R_{t+\tau}^2$.

Note that when the return is 0 (as a result of price discreteness) and the log is undefined, we set the return equal to .0001, or one basis point.

A significant disadvantage of EGARCH models is that the proper aggregation for long-horizon forecasting is rather inconvenient, in that is difficult to compute unbiased forecasts of volatility over multiperiod intervals. We forecast log volatility simply by setting the shock terms in the volatility equation, X_{t-1}^R and X_{t-1}^D, and the lagged return R_{t-1} equal to 0. This results in a forecast of future *log* volatility and under no reasonable conditions would imply an unbiased forecast of the *level* of volatility more than 1 day ahead. For this reason, we place greater emphasis on the first two measures, the average log absolute return and average log range, because their expectations are linear in integrated log volatility. In particular, we know from the results of Alizadeh et al. (2002) that

$$\frac{1}{\tau_1 - \tau_0 + 1}\sum_{\tau=\tau_0}^{\tau_1} \ln|R_{t+\tau}| = -.64 + \frac{1}{\tau_1 - \tau_0 + 1}\sum_{\tau=\tau_0}^{\tau_1} \ln h_{t+\tau} + u \quad (19)$$

and

$$\frac{1}{\tau_1 - \tau_0 + 1}\sum_{\tau=\tau_0}^{\tau_1} D_{t+\tau} = .43 + \frac{1}{\tau_1 - \tau_0 + 1}\sum_{\tau=\tau_0}^{\tau_1} \ln h_{t+\tau} + v, \quad (20)$$

where u and v are 0 mean error terms.

Therefore, we can construct forecasts of the average log absolute return and average log range by replacing the future log volatility $\sum_{\tau=\tau_0}^{\tau_1} \ln h_{t+\tau}$ with the corresponding forecasts $\sum_{\tau=\tau_0}^{\tau_1} \widehat{\ln h_{t+\tau}}$. Finally, recall from Section 2.1 that from the properties of the two log volatility proxies, $\text{var}[u] \simeq 15\,\text{var}[v]$, which means that the average log range provides a much more informative forecast evaluation than the average log absolute return.

Table 3 summarizes the 4,549 sets of rolling sample parameter estimates for each model specification. The first estimation period ends in December 1986, so the corresponding forecast period includes the crash of October 1987. The results confirm the finding from Table 1 that the range-based estimates are generally more precise. Also as in Table 1, asymmetries are important; however, the role of asymmetry in the long-term component of the two-factor volatility models is somewhat unstable, as can be seen from the distributions of the δ_q parameter estimates.

Given these rolling sample estimates, we examine both short-horizon and long-horizon volatility forecasts. First, we consider the problem of forecasting volatility over the first τ_1 days immediately after the estimation period. Specifically, we set $\tau_0 = 1$ and consider three volatility intervals τ_1 of 1, 5, and 22 days, approximately representing daily, weekly, and monthly integrated volatilities. To avoid serious problems with autocorrelation, we examine only *nonoverlapping* forecasts, which means that there are fewer monthly volatility forecasts than daily volatility forecasts. In total, we have 4,548 1-day forecasts, 909 1-week forecasts, and 206 1-month forecasts. Second, we consider the problem of forecasting 1-month (22-day) volatility 1–12 months from the end of the estimation period.

Table 3. Rolling Sample Maximum Likelihood Estimates

	κ or κ_q	θ	ϕ or ϕ_q	δ or δ_q	κ_h	ϕ_h	δ_h	d	ω
EGARCH1	.0518 (.0439)	−4.5087 (.3545)	.0349 (.0214)	−.0570 (.0272)					
EGARCH1, symmetric	.0315 (.0243)	−4.4689 (.3396)	.0504 (.0312)						
EGARCH2	.0119 (.0096)	−4.4318 (.4480)	.0208 (.0129)	−.0155 (.0222)	.3296 (.1663)	.0132 (.0163)	−.1080 (.0415)		
EGARCH2, symmetric	.0130 (.0156)	−4.2791 (.5986)	.0184 (.0138)		.2150 (.2182)	.0539 (.0334)			
EGARCH2, part symmetric	.0135 (.0158)	−4.3927 (.4646)	.0226 (.0149)		.1802 (.1146)	.0208 (.0165)	−.0908 (.0364)		
FIEGARCH		−4.1464 (.9385)	.0392 (.0216)	−.0762 (.0375)				.4922 (.1977)	.4584 (.2186)
FIEGARCH symmetric		−3.9440 (.6983)	.0493 (.0289)					.3974 (.2641)	.7079 (.2530)
REGARCH1	.0400 (.0214)	−4.9615 (.2205)	.0282 (.0063)	−.0369 (.0182)					
REGARCH1, symmetric	.0271 (.0118)	−4.9912 (.2572)	.0344 (.0095)						
REGARCH2	.0183 (.0110)	−4.9767 (.2866)	.0213 (.0087)	−.0129 (.0177)	.3680 (.2314)	.0164 (.0106)	−.0621 (.0274)		
REGARCH2, symmetric	.0130 (.0137)	−4.7858 (.6009)	.0131 (.0140)		.1752 (.2528)	.0344 (.0103)			
REGARCH2, part symmetric	.0215 (.0167)	−4.9611 (.2967)	.0223 (.0116)		.1960 (.1127)	.0216 (.0064)	−.0544 (.0218)		
FIREGARCH		−4.8198 (.7658)	.0315 (.0067)	−.0453 (.0224)				.5067 (.2014)	.4667 (.2388)
FIREGARCH symmetric		−4.9316 (.6167)	.0317 (.0091)					.5204 (.2239)	.6386 (.1827)

NOTE: Averages and standard deviations (in parentheses) of 4,549 sets of rolling samples (with sample size of 1,000 days) maximum likelihood estimates of the return-based EGARCH model with likelihood function $R_t \sim \text{N}[0, h_t^2]$ and range-based REGARCH models with likelihood function $D_t \sim \text{N}[.43 + \ln h_t, .29^2]$. The conditional volatility follows the single-factor process (8) or (13), the two-factor process (10) or (15), or the fractionally integrated process (11) or (16). Daily data on the S&P500 index are subsamples of the period January 1983–December 2004.

We set $\tau_0 = 22(m-1)+1$ and $\tau_1 = 22m$ for $m = 1, 2, \ldots, 12$, where $m = 1$ corresponds to the 1-month forecasts already considered. Because the length of the volatility interval does not increase, the number of nonoverlapping forecasts declines only slightly for longer horizons, from 206 1-month forecasts to 195 1-year-ahead 1-month forecasts.

5.1 Regression Tests

We use regression-based tests to assess the ability of each model to predict the four realized volatility measures described above. The tests are based on the regression

$$\begin{aligned}&\text{realized proxy}(t+\tau_0, t+\tau_1)\\&\quad = \alpha_i + \beta_i \times \text{model } i\text{'s forecast}_t(t+\tau_0, t+\tau_1)\\&\qquad + \varepsilon_{i,t}(t+\tau_0, t+\tau_1),\end{aligned} \tag{21}$$

where unbiasedness of the forecast requires that $\alpha_i = 0$ and $\beta_i = 1$. In addition to these coefficients, we also examine the regression R^2's and residual autocorrelations.

5.1.1 Short-Horizon Results. Table 4 presents estimates of the intercepts α_i along with heteroscedasticity and serial correlation–adjusted t-statistics. For forecasting log absolute returns, the return-based models produce forecasts with insignificant intercepts, thus providing no indication of bias. The range-based models perform relatively poorly, displaying a bias at the 1-day and 5-day horizons in five out of seven cases. However, for the other three volatility measures (the log range, realized standard deviation, and realized variance, all of which are more informative than the log absolute return), the return-based models fare much worse. In 44 out of 63 cases (7 models, 3 volatility measures, and 3 horizons), the return-based forecasts are associated with significant intercepts. In contrast, only six of the range-based model intercepts are significantly nonzero.

Table 5 presents estimates of the slope coefficient β_i, providing further evidence against both the return-based models and the use of the log absolute return regressions for forecast evaluation. In the log absolute return regressions, none of the slopes for the return-based models is significantly different from 1, but in the log range, realized volatility, and realized variance regressions, the slopes are significantly <1 in 48 out of 63 cases. The range-based forecasts display significant biases in 34 out of the same 63 cases, with most of the rejections occurring for forecasts of realized variance.

Table 6 reports the R^2's of the forecast regressions. The results clearly demonstrate the almost uniform superiority of the range-based models for forecasting all four volatility measures. Only the 1-day-ahead forecasts of the realized variance show some support for the return-based models. In every other

Table 4. Short-Horizon Forecast Regression Intercepts

	Average log absolute return $1/\tau_1 \sum_{\tau=1}^{\tau_1} \ln\lvert R_{t+\tau}\rvert$			*Average log range* $1/\tau_1 \sum_{\tau=1}^{\tau_1} D_{t+\tau}$			*Realized standard deviation* $\sqrt{1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2}$			*Realized variance* $1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2$		
	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$
EGARCH1	−.162	−.255	−.017	**−.580**	**−.690**	**−.799**	.709	.899	.432	**.695**	**.524**	.246
	(.594)	(.712)	(.028)	**(5.937)**	**(3.851)**	**(2.063)**	(1.486)	(1.708)	(.435)	**(3.940)**	**(2.499)**	(.785)
EGARCH1,	−.172	−.221	−.464	**−.542**	**−.652**	**−1.140**	**1.398**	**2.444**	**4.756**	**.706**	**.870**	**1.086**
symmetric	(.606)	(.605)	(.884)	**(5.109)**	**(3.682)**	**(3.059)**	**(2.463)**	**(2.146)**	**(2.491)**	**(5.599)**	**(4.321)**	**(4.149)**
EGARCH2	−.064	−.022	−.006	**−.620**	**−.585**	**−.773**	.494	1.215	1.914	**.432**	**.650**	.704
	(.252)	(.066)	(.012)	**(6.689)**	**(3.598)**	**(2.501)**	(.877)	(1.603)	(1.293)	**(2.876)**	**(2.947)**	(1.912)
EGARCH2,	−.265	−.301	−.415	**−.617**	**−.714**	**−1.070**	**1.039**	**1.890**	**2.560**	**.623**	**.775**	**.727**
symmetric	(.994)	(.865)	(.810)	**(6.360)**	**(4.307)**	**(3.285)**	**(2.530)**	**(2.210)**	**(1.982)**	**(4.206)**	**(3.598)**	**(2.148)**
EGARCH2,	−.049	−.002	.069	**−.562**	**−.568**	**−.733**	.360	.444	1.181	**.600**	**.236**	.449
part symmetric	(.189)	(.006)	(.126)	**(5.891)**	**(3.221)**	**(2.159)**	(1.054)	(.944)	(1.106)	**(3.089)**	**(2.005)**	(1.365)
FIEGARCH	−.354	−.341	−.296	**−.782**	**−.815**	**−.989**	**.793**	.634	1.053	**.663**	**.189**	.312
	(1.477)	(1.082)	(.616)	**(8.920)**	**(4.947)**	**(3.315)**	**(2.463)**	(1.407)	(1.346)	**(3.684)**	**(2.008)**	(1.102)
FIEGARCH	−.174	−.171	−.122	**−.568**	**−.639**	**−.877**	.839	.873	1.091	**.653**	**.491**	.365
symmetric	(.662)	(.527)	(.263)	**(6.020)**	**(4.111)**	**(3.033)**	(1.895)	(1.875)	(1.314)	**(4.045)**	**(2.263)**	(1.188)
REGARCH1	**.545**	**.600**	**.896**	.005	−.025	−.016	−.599	.038	.608	−.135	.118	.362
	(2.204)	**(2.087)**	**(2.086)**	(.058)	(.235)	(.067)	(1.397)	(.112)	(.642)	(.888)	(1.288)	(1.063)
REGARCH1,	**.554**	**.701**	**.816**	.056	.060	−.047	−.387	.033	1.134	−.008	.148	.464
symmetric	**(2.150)**	**(2.414)**	**(2.091)**	(.628)	(.578)	(.219)	(.963)	(.089)	(1.171)	(.065)	(1.274)	(1.402)
REGARCH2	**.583**	**.657**	**.814**	−.018	−.013	−.068	−.683	−.051	.678	−.194	.085	.320
	(2.387)	**(2.342)**	**(2.183)**	(.222)	(.128)	(.361)	(1.530)	(.142)	(.970)	(1.207)	(.956)	(1.233)
REGARCH2,	.478	**.624**	.666	−.021	−.026	−.189	−.427	−.060	.788	−.061	.094	.325
symmetric	(1.883)	**(2.154)**	(1.776)	(.236)	(.231)	(.941)	(1.062)	(.161)	(1.106)	(.471)	(.963)	(1.161)
REGARCH2,	**.627**	**.705**	**.843**	.064	.067	−.015	−.776	−.199	.752	−.227	.038	.344
part symmetric	**(2.489)**	**(2.352)**	**(2.099)**	(.756)	(.581)	(.072)	(1.679)	(.489)	(.951)	(1.298)	(.504)	(1.260)
FIREGARCH	.295	.326	.293	**−.230**	**−.278**	**−.495**	−.373	.260	**1.183**	−.177	.060	.315
	(1.268)	(1.193)	(.863)	**(2.925)**	**(2.393)**	**(2.757)**	(.893)	(.660)	**(2.145)**	(1.120)	(.946)	(1.313)
FIREGARCH	.377	.467	.422	−.114	−.148	**−.349**	−.250	.248	**1.255**	−.042	.122	.374
symmetric	(1.525)	(1.679)	(1.307)	(1.320)	(1.336)	**(2.271)**	(.638)	(.633)	**(2.204)**	(.331)	(1.537)	(1.589)

NOTE: Estimates of the intercept coefficient α in the regression: realized proxy$(t+1, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+1, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on December 30, 2004, resulting in sample sizes of 4,548, 909, and 206 for $\tau_1 = 1$, 5, and 22. T-statistics computed using Newey–West standard errors are in parentheses. Intercepts for the standard deviation and variance have been multiplied by 10^3 and 10^4.

Table 5. Short-Horizon Forecast Regression Slopes

	Average log absolute return $1/\tau_1 \sum_{\tau=1}^{\tau_1} ln\lvert R_{t+\tau}\rvert$			Average log range $1/\tau_1 \sum_{\tau=1}^{\tau_1} D_{t+\tau}$			Realized standard deviation $\sqrt{1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2}$			Realized variance $1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2$		
	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$
EGARCH1	.99	.97	1.02	**.92**	**.90**	.88	**.69**	**.84**	.94	**.46**	**.64**	.94
	(.25)	(.43)	(.17)	**(3.29)**	**(2.40)**	(1.36)	**(6.12)**	**(2.91)**	(.58)	**(8.23)**	**(4.46)**	(.33)
EGARCH1,	.99	.98	.94	**.94**	**.92**	**.81**	**.61**	**.66**	**.47**	**.43**	**.30**	**.11**
symmetric	(.16)	(.25)	(.57)	**(2.41)**	**(2.04)**	**(2.22)**	**(6.57)**	**(3.09)**	**(3.19)**	**(8.50)**	**(12.82)**	**(16.59)**
EGARCH2	1.01	1.02	1.02	**.92**	**.93**	.89	**.71**	**.80**	.78	**.67**	**.51**	**.47**
	(.14)	(.25)	(.25)	**(3.88)**	**(1.96)**	(1.60)	**(4.68)**	**(2.66)**	(1.62)	**(2.09)**	**(5.70)**	**(2.69)**
EGARCH2,	.97	.97	.95	**.92**	**.90**	**.82**	**.64**	**.72**	**.68**	**.50**	**.38**	**.41**
symmetric	(.52)	(.48)	(.49)	**(3.40)**	**(2.55)**	**(2.33)**	**(8.25)**	**(3.58)**	**(2.80)**	**(8.31)**	**(9.70)**	**(3.90)**
EGARCH2,	1.01	1.02	1.03	**.93**	.93	.89	**.73**	.89	.87	**.55**	.91	.74
part symmetric	(.16)	(.27)	(.33)	**(3.26)**	(1.78)	(1.39)	**(7.07)**	(1.85)	(1.24)	**(5.22)**	(1.01)	(1.38)
FIEGARCH	.95	.96	.97	**.88**	**.87**	**.84**	**.67**	**.85**	.86	**.46**	.90	.81
	(1.04)	(.73)	(.34)	**(5.87)**	**(3.30)**	**(2.36)**	**(9.82)**	**(2.68)**	(1.91)	**(8.67)**	(.96)	(1.36)
FIEGARCH	.99	.99	1.01	**.93**	**.92**	**.87**	**.66**	**.82**	**.83**	**.48**	**.64**	.74
symmetric	(.20)	(.13)	(.08)	**(3.02)**	**(2.27)**	**(1.97)**	**(7.33)**	**(3.89)**	**(2.15)**	**(7.27)**	**(4.67)**	(1.79)
REGARCH1	1.07	1.07	1.13	1.00	.99	.99	1.09	**1.23**	**1.25**	**2.10**	**1.77**	1.48
	(1.48)	(1.44)	(1.63)	(.13)	(.39)	(.15)	(1.31)	**(4.16)**	**(2.29)**	**(2.73)**	**(4.29)**	(1.46)
REGARCH1,	1.07	1.09	1.11	1.01	1.01	.99	1.06	**1.23**	1.17	**1.94**	**1.74**	1.28
symmetric	(1.47)	(1.79)	(1.60)	(.52)	(.45)	(.29)	(.96)	**(3.91)**	(1.52)	**(2.72)**	**(4.20)**	(.94)
REGARCH2	1.07	1.09	1.12	1.00	1.00	.99	1.08	**1.22**	**1.20**	**2.13**	**1.77**	**1.45**
	(1.70)	(1.74)	(1.73)	(.26)	(.14)	(.33)	(1.22)	**(3.73)**	**(2.48)**	**(2.74)**	**(4.06)**	**(2.32)**
REGARCH2,	1.06	1.08	1.09	.99	.99	.96	1.06	**1.23**	**1.19**	**1.99**	**1.78**	**1.45**
symmetric	(1.23)	(1.57)	(1.33)	(.26)	(.27)	(.94)	(.89)	**(3.69)**	**(2.27)**	**(2.79)**	**(4.14)**	**(2.08)**
REGARCH2,	1.08	1.10	1.12	1.01	1.01	1.00	1.11	**1.26**	**1.21**	**2.23**	**1.88**	**1.46**
part symmetric	(1.80)	(1.78)	(1.65)	(.66)	(.53)	(.10)	(1.47)	**(3.69)**	**(2.16)**	**(2.73)**	**(3.84)**	**(2.03)**
FIREGARCH	1.02	1.03	1.03	**.95**	**.94**	**.89**	1.03	**1.17**	1.11	**2.04**	**1.74**	**1.38**
	(.57)	(.61)	(.42)	**(2.92)**	**(2.39)**	**(2.69)**	(.50)	**(2.47)**	(1.69)	**(2.65)**	**(3.00)**	**(2.60)**
FIREGARCH	1.04	1.05	1.05	.97	.97	**.92**	1.03	**1.19**	1.12	**1.94**	**1.72**	**1.34**
symmetric	(.85)	(1.07)	(.78)	(1.38)	(1.42)	**(2.29)**	(.51)	**(2.78)**	(1.83)	**(2.71)**	**(3.22)**	**(2.50)**

NOTE: Estimates of the slope coefficient β in the regression: realized proxy$(t+1, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+1, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on December 30, 2004, resulting in sample sizes of 4,548, 909, and 206 for $\tau_1 = 1$, 5, and 22. *T*-statistics computed using Newey-West standard errors are in parentheses.

case, the R^2 of the range-based forecasts is higher than that of the corresponding return-based forecasts (e.g., REGARCH2 vs. EGARCH2). Another striking feature of the results is the consistent, although modest, superiority of the two-factor and fractionally integrated models. In addition, models that incorporate some form of asymmetry offer significant advantages in short-horizon forecasts. Because the partially symmetric models perform about the same as their unrestricted counterparts at long horizons, the asymmetry again seems to be important mostly for the short-run component of the volatility dynamics.

Finally, Table 7 gives the first-order autocorrelations of the forecast residuals. For a forecast to be optimal, the residual must be uncorrelated with all information available at the time of the forecast, including the residual from the previous forecast. However, the table shows that the residual autocorrelations from many of the return-based and range-based models are reliably nonzero. Although there is no clear pattern in the results, it appears that the residuals of the range-based forecasts of the return-based volatility measures are frequently autocorrelated. Likewise, the residuals of return-based forecasts of the range-based volatility measures tend to be autocorrelated. Thus all specifications display some failings.

5.1.2 Long-Horizon Results. We now assess the ability of each model to forecast the 1-month integrated volatility up to 1 year from the end of the estimation period. To reduce the quantity of results, we consider only 3-, 6-, and 12-month forecasts of the average log absolute return and average log range (the volatility measures that are linear in log volatility). Furthermore, we report only the estimates and t-statistics of the slope coefficient β_i, as measures of forecast bias and statistical significance, and the corresponding R^2's, as a description of forecast accuracy. The omitted results and forecast horizons are consistent with the reported results and are available on request.

The first three columns of Table 8 present the estimates and t-statistics of the slope coefficients for the log absolute return regressions. The pattern in the results is clear. In 15 out of 21 cases, the return-based forecasts are significantly biased. In contrast, the range-based forecasts are significantly biased in only three cases, with all violations occurring at the longest forecast horizon. The final three columns show analogous results for the average log range regressions. In this table, all of the slope coefficients for the return-based models are significantly biased. Unbiasedness is rejected about half the time for range-based models, with most of the rejections coming from the symmetric or fractionally integrated specifications.

To determine which forecasts are more informative, Table 9 reports the regression R^2's. Again, the results are easy to summarize. In every case, the range-based model outperforms its return-based counterpart. Furthermore, the performance of some of the two-factor range-based models at long horizons is

Table 6. Short-Horizon Forecast Regression R^2's

	Average log absolute return $1/\tau_1 \sum_{\tau=1}^{\tau_1} \ln\|R_{t+\tau}\|$			Average log range $1/\tau_1 \sum_{\tau=1}^{\tau_1} D_{t+\tau}$			Realized standard deviation $\sqrt{1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2}$			Realized variance $1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2$		
	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$
EGARCH1	.091	.282	.445	.367	.543	.502	.158	.281	.269	.027	.044	.055
EGARCH1, symmetric	.086	.287	.476	.353	.558	.529	.130	.239	.171	.024	.030	.013
EGARCH2	.105	.327	.533	.401	.610	.608	.170	.294	.276	.032	.041	.039
EGARCH2, symmetric	.086	.291	.495	.355	.565	.562	.136	.258	.244	.025	.035	.035
EGARCH2, part symmetric	.100	.316	.506	.389	.589	.573	.166	.291	.270	.028	.050	.047
FIEGARCH	.099	.315	.518	.386	.588	.586	.166	.302	.308	.029	.051	.062
FIEGARCH, symmetric	.088	.299	.519	.360	.574	.586	.139	.278	.275	.026	.044	.050
REGARCH1	.107	.353	.566	.429	.672	.666	.175	.332	.326	.029	.052	.066
REGARCH1, symmetric	.099	.347	.577	.405	.664	.687	.149	.312	.307	.020	.045	.057
REGARCH2	.112	.366	.602	.442	.689	.712	.179	.330	.339	.030	.051	.072
REGARCH2, symmetric	.100	.352	.602	.406	.667	.706	.149	.314	.330	.021	.045	.067
REGARCH2, part symmetric	.109	.356	.583	.440	.684	.700	.175	.324	.318	.030	.052	.064
FIREGARCH	.111	.369	.618	.436	.688	.712	.176	.333	.346	.029	.052	.073
FIREGARCH, symmetric	.101	.355	.613	.408	.671	.724	.150	.315	.335	.021	.046	.068

NOTE: Adjusted R^2's of the regression: realized proxy$(t+1, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+1, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on December 30, 2004, resulting in sample sizes of 4,548, 909, and 206 for $\tau_1 = 1$, 5, and 22. Corresponding asymptotic standard errors are .0003, .0016, and .0071.

Table 7. Short-Horizon Forecast Regression Residual Autocorrelations

	Average log absolute return $1/\tau_1 \sum_{\tau=1}^{\tau_1} \ln\|R_{t+\tau}\|$			Average log range $1/\tau_1 \sum_{\tau=1}^{\tau_1} D_{t+\tau}$			Realized standard deviation $\sqrt{1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2}$			Realized variance $1/\tau_1 \sum_{\tau=1}^{\tau_1} R_{t+\tau}^2$		
	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$	$\tau_1 = 1$	$\tau_1 = 5$	$\tau_1 = 22$
EGARCH1	−.034	.051	.391	.100	.357	.534	−.036	.017	.086	−.020	−.010	−.028
EGARCH1, symmetric	−.042	−.001	.249	.090	.300	.422	.008	.057	.034	.027	.034	.027
EGARCH2	−.033	.015	.197	.078	.304	.353	.001	.031	−.032	.033	.003	−.049
EGARCH2, symmetric	−.045	−.001	.284	.082	.301	.465	−.009	.027	.037	.011	.008	−.028
EGARCH2, part symmetric	−.034	.031	.269	.084	.329	.428	−.038	.070	.057	−.033	.015	−.023
FIEGARCH	−.044	.024	.340	.068	.336	.485	−.047	.056	.075	−.018	.025	−.015
FIEGARCH, symmetric	−.044	−.023	.260	.083	.269	.451	−.004	.007	.061	.027	−.007	−.018
REGARCH1	−.047	−.018	.174	−.030	.055	.163	.008	.060	−.030	.072	.064	−.062
REGARCH1, symmetric	−.046	−.025	.103	−.015	.052	.087	.033	.094	−.041	.086	.091	−.058
REGARCH2	−.045	−.031	.066	−.022	.048	.033	.009	.085	−.044	.066	.079	−.056
REGARCH2, symmetric	−.048	−.031	.119	−.018	.058	.139	.031	.101	−.006	.086	.098	−.041
REGARCH2, part symmetric	−.046	−.018	.084	−.032	.060	.072	.008	.090	−.015	.068	.072	−.036
FIREGARCH	−.054	−.031	.158	−.050	.084	.222	.004	.099	.050	.072	.088	−.006
FIREGARCH, symmetric	−.045	−.039	.104	−.007	.042	.126	.035	.104	.019	.086	.101	−.015

NOTE: First-order autocorrelation of the residuals of the regression: realized proxy$(t+1, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+1, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on December 30, 2004, resulting in sample sizes of 4,548, 909, and 206 for $\tau_1 = 1$, 5, and 22. Corresponding asymptotic standard errors are .014, .034, and .071.

Table 8. Long-Horizon Forecast Regression Slopes

	Average log absolute return $1/(\tau_1-\tau_0+1)\sum_{\tau=\tau_0}^{\tau_1} \ln\lvert R_{t+\tau}\rvert$, $\tau_0=22(m-1)+1$, $\tau_1=22m$			Average log range $1/(\tau_1-\tau_0+1)\sum_{\tau=\tau_0}^{\tau_1} D_{t+\tau}$, $\tau_0=22(m-1)+1$, $\tau_1=22m$		
	$m=3$	$m=6$	$m=12$	$m=3$	$m=6$	$m=12$
EGARCH1	.790	.640	**.380**	**.640**	**.540**	**.320**
	(1.160)	(1.730)	**(2.650)**	**(2.430)**	**(2.640)**	**(3.740)**
EGARCH1, symmetric	**.640**	**.460**	**.240**	**.530**	**.390**	**.210**
	(2.380)	**(3.070)**	**(4.100)**	**(4.170)**	**(4.350)**	**(5.530)**
EGARCH2	.840	**.630**	**.310**	**.690**	**.500**	**.260**
	(1.140)	**(2.370)**	**(4.230)**	**(3.000)**	**(3.820)**	**(5.650)**
EGARCH2, symmetric	**.640**	**.450**	**.200**	**.530**	**.350**	**.160**
	(2.460)	**(3.610)**	**(5.610)**	**(4.400)**	**(5.220)**	**(7.220)**
EGARCH2, part symmetric	.780	**.570**	**.260**	**.650**	**.460**	**.190**
	(1.470)	**(2.680)**	**(4.510)**	**(3.330)**	**(4.120)**	**(6.330)**
FIEGARCH	.800	**.670**	**.500**	**.680**	**.570**	**.420**
	(1.500)	**(2.120)**	**(2.930)**	**(3.190)**	**(3.420)**	**(4.150)**
FIEGARCH symmetric	.800	**.660**	**.430**	**.660**	**.550**	**.360**
	(1.590)	**(2.240)**	**(3.340)**	**(3.440)**	**(3.680)**	**(4.730)**
REGARCH1	1.140	.990	.720	.980	.870	.660
	(.800)	(.030)	(.970)	(.150)	(.690)	(1.490)
REGARCH1, symmetric	1.020	.800	.530	.870	.690	**.470**
	(.090)	(.890)	(1.800)	(.990)	(1.650)	**(2.420)**
REGARCH2	1.170	1.050	.700	.970	.900	.620
	(1.340)	(.230)	(1.190)	(.310)	(.680)	(1.900)
REGARCH2, symmetric	.940	.800	**.640**	**.800**	**.680**	**.550**
	(.520)	(1.290)	**(1.990)**	**(2.000)**	**(2.580)**	**(3.060)**
REGARCH2, part symmetric	1.150	1.040	.780	.960	.880	.660
	(1.060)	(.200)	(.860)	(.340)	(.790)	(1.620)
FIREGARCH	.930	.860	**.690**	**.790**	**.740**	**.600**
	(.770)	(1.340)	**(2.320)**	**(3.280)**	**(3.380)**	**(3.970)**
FIREGARCH symmetric	.930	.840	**.620**	**.790**	**.720**	**.530**
	(.730)	(1.330)	**(2.520)**	**(3.050)**	**(2.940)**	**(3.750)**

NOTE: Estimates of the slope coefficient β in the regression: realized proxy$(t+\tau_0, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+\tau_0, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days of data before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on November 30, 2004, resulting in sample sizes ranging from 204 ($m=3$, or about 3 months ahead) to 195 ($m=12$, or about 12 months ahead). T-statistics computed using Newey–West standard errors are in parentheses.

Table 9. Long-Horizon Forecast Regression R^2's

	Average log absolute return $1/(\tau_1-\tau_0+1)\sum_{\tau=\tau_0}^{\tau_1} \ln\lvert R_{t+\tau}\rvert$, $\tau_0=22(m-1)+1$, $\tau_1=22m$			Average log range $1/(\tau_1-\tau_0+1)\sum_{\tau=\tau_0}^{\tau_1} D_{t+\tau}$, $\tau_0=22(m-1)+1$, $\tau_1=22m$		
	$m=3$	$m=6$	$m=12$	$m=3$	$m=6$	$m=12$
EGARCH1	.214	.133	.047	.212	.146	.051
EGARCH1, symmetric	.210	.106	.027	.216	.113	.031
EGARCH2	.308	.184	.053	.314	.178	.058
EGARCH2, symmetric	.230	.131	.035	.241	.121	.033
EGARCH2, part symmetric	.269	.161	.042	.278	.157	.037
FIEGARCH	.326	.238	.149	.357	.260	.165
FIEGARCH, symmetric	.312	.224	.111	.325	.232	.117
REGARCH1	.329	.218	.116	.366	.254	.146
REGARCH1, symmetric	.327	.185	.079	.359	.209	.097
REGARCH2	.431	.300	.136	.447	.335	.167
REGARCH2, symmetric	.382	.280	.188	.425	.303	.213
REGARCH2, part symmetric	.408	.293	.161	.435	.316	.179
FIREGARCH	.455	.386	.279	.491	.431	.319
FIREGARCH, symmetric	.456	.374	.222	.496	.413	.252

NOTE: Adjusted R^2's of the regression: realized proxy$(t+\tau_0, t+\tau_1) = \alpha + \beta \times$ model forecast$_t(t+\tau_0, t+\tau_1) + \varepsilon$. Forecasts are constructed from 1,000 days of data before each nonoverlapping forecast interval. The first forecast is on December 16, 1986, and the last is on November 30, 2004, resulting in sample sizes ranging from 204 ($m=3$, or about 3 months ahead) to 195 ($m=12$, or about 12 months ahead). Asymptotic standard errors all range from .0071 to .0075.

quite impressive. The FIREGARCH model, which displayed many biases and performed relatively poorly at shorter horizons, is the best model for long-term forecasting. At a 1-year horizon, for instance, the FIREGARCH model results in an R^2 of about .28 for forecasting the average log absolute return and .32 for forecasting the average log range. This result clearly contradicts the usual perception and related empirical findings of West and Cho (1995) and Christoffersen and Diebold (2000) that volatility predictability is solely a short-horizon phenomenon.

5.2 Qualitative Forecast Comparisons

To gain a better understanding of the differences between the volatility forecasts that drive the statistical results in the previous section, we generate pairwise scatterplots of the forecasts generated by six leading model specifications (EGARCH1, partially symmetric EGARCH2, FIEGARCH, REGARCH1, partially symmetric REGARCH2, and FIREGARCH). Figure 4 shows these plots both for 1-day and 22-day volatility forecasts. Plots above the diagonal correspond to 1-day forecasts, whereas forecasts of 1-day volatility 22 days from the sample end are below the diagonal. Longer horizon forecasts (not plotted) are visually similar to the 22-day forecasts.

Several patterns emerge from the figure. Least surprising of these is that the 22-day forecasts are not as dispersed as those for the 1-day horizon, reflecting mean reversion in volatility that is captured to some degree by all of the models. More interestingly, there is an obvious qualitative difference between return-based and range-based forecasts, with return-based forecasts occasionally taking on far more extreme values than those resulting from the range-based models. Finally, although all three range-based models imply roughly identical 1-day forecasts, differing substantially only at the longer horizon, the three return-based forecasts are easily distinguished at both horizons.

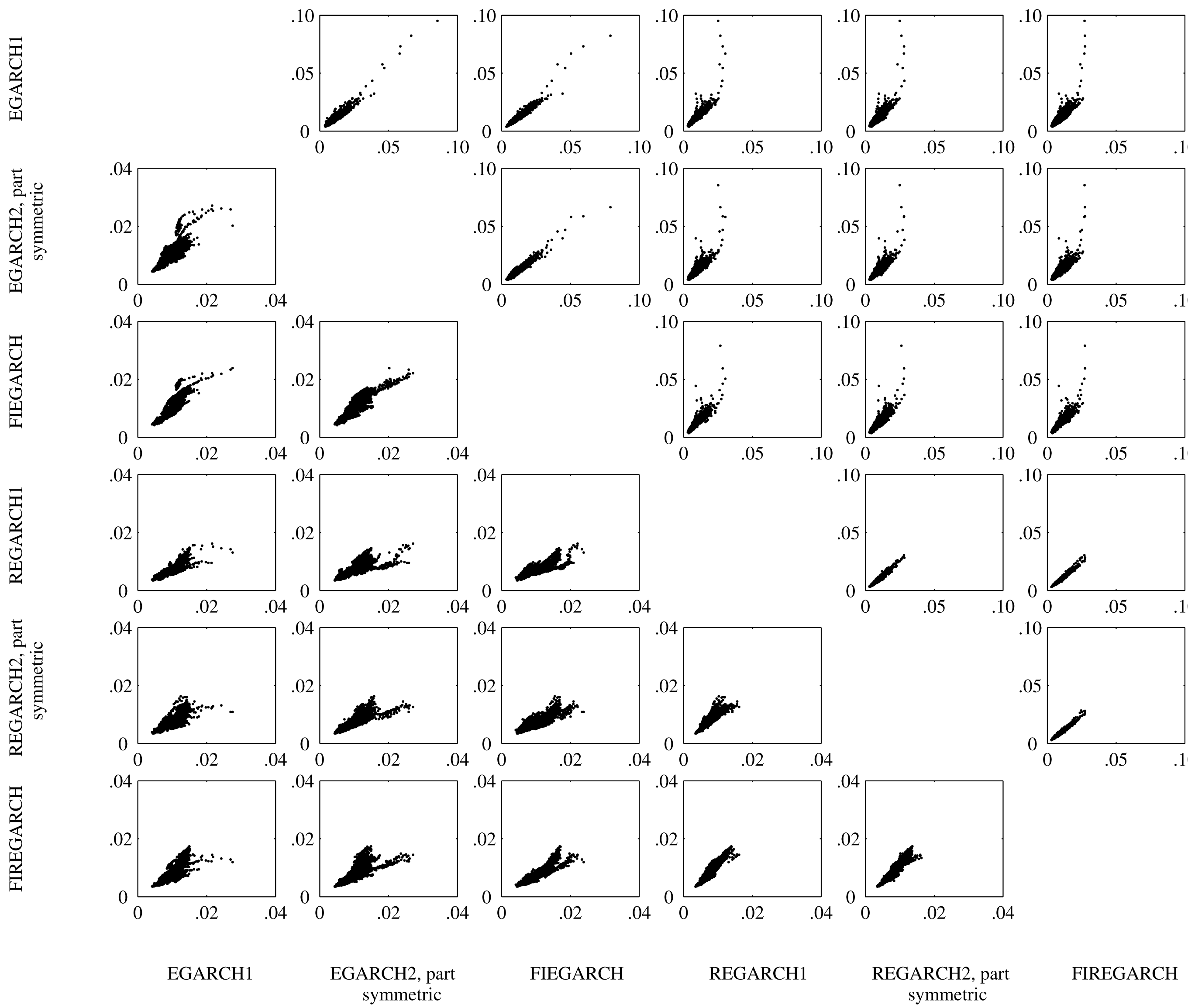

Figure 4. Pairwise Comparisons of Volatility Forecasts From Different Models. 1-day forecasts are above the diagonal, and 22-day forecasts are below it.

5.3 Formal Forecast Comparisons

West and Cho (1995) and Diebold and Mariano (1995) have suggested alternative statistics for comparing the forecasting performance of different model specifications. The differentiating feature of these statistics relative to the regression tests is that these statistics do not correct for any bias. If range-based volatility measures are systematically different from return-based volatility measures due to market microstructure effects, then this lack of bias correction should cause return-based models to better forecast return-based volatility measures and, likewise, range-based models to better forecast range-based volatility measures.

Although both West and Cho (1995) and Diebold and Mariano (1995) considered a variety of forecast evaluation measures, we focus here on the root mean squared error (RMSE), where the error of model i at date t is defined as

$$\varepsilon_{i,t}(t+\tau_0, t+\tau_1) \equiv \text{realized proxy}(t+\tau_0, t+\tau_1) - \text{model } i\text{'s forecast}_t(t+\tau_0, t+\tau_1). \quad (22)$$

Specifically, we are interested in measuring the *relative* forecasting performance of the different model specifications. Following Diebold and Mariano (1995), we can directly test the superiority of model i over model j with a t-test of the $\mu_{i,j}$ coefficient in

$$\varepsilon_{j,t}^2 - \varepsilon_{i,t}^2 = \mu_{i,j} + \eta_t, \quad (23)$$

where a positive estimate of $\mu_{i,j}$ indicates support for model i. Because the number of pairwise tests is rather large, we focus on a comparison of only six of the candidate models. In particular, we consider the partially symmetric REGARCH2 model, because it appears to forecast well at most horizons, and the REGARCH1 model as its single-factor competitor. We also consider FIREGARCH as an alternative range-based model capable of generating long memory. Among the return-based models, we consider the EGARCH1, partially symmetric EGARCH2, and FIEGARCH models.

Table 10 presents the t-statistics, computed with heteroscedasticity and serial correlation–adjusted standard errors, for the estimates of $\mu_{i,j}$. The top half of the table focuses on forecasts of the log absolute return. Although it is difficult to conclude that the range-based forecasts of absolute returns are significantly better than the return-based forecasts, it is clear that in most cases the two-factor and fractionally integrated models offer significant improvements over their one-factor counterparts (e.g., partially symmetric EGARCH2 vs. EGARCH1). Furthermore, in the bottom half of the table, which focuses on forecasts of the log range, we see highly significant improvements from both range-based forecasts and two-factor models.

Finally, Table 11 repeats the analysis for selected long-horizon forecasts. Again, the results in the top half of the table are a bit inconclusive as to the advantages of using range-based forecasts, although a few of the range-based models are significantly better than their return-based counterparts in forecasting log absolute returns. As in the last table, the results in the bottom half clearly show the benefits of the range. In particular, when we use the range in the estimation, the two-factor specifications outperform their simpler one-factor counterparts. When we use only returns in the estimation, in contrast, the added complexity of the two-factor specifications appears to backfire in some cases. We conclude that returns are not sufficiently informative to accurately pin down the parameters of the more complicated (but also more accurate) two-factor volatility dynamics. The resulting estimation errors translate into greater forecast errors that often make the simpler EGARCH1 model preferable.

Because the two-factor and fractionally integrated specifications provide different ways to capture the apparent long memory in volatility, it is worth comparing the relative performance of these two classes of models. For return-based models, two-factor models have some advantages over fractional integration at shorter horizons, whereas the latter's strength is in forecasting at the longest horizons. For range-based models, fractional integration seems to offer moderate advantages at both short and long horizons, particularly for forecasting log absolute returns. The one exception is the superior performance of

Table 10. Pairwise Comparisons of Short-Run Forecast RMSEs

Average log absolute return: $1/\tau_1 \sum_{\tau=1}^{\tau_1} \ln|R_{t+\tau}|$

$\tau_1 = 1$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	3.63	3.30	−.01	.58	1.19
E2		−1.37	−1.42	−.84	−.24
EF			−.92	−.32	.26
R1				2.59	5.66
R2					2.66

$\tau_1 = 5$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	3.19	3.24	.56	.97	1.91
E2		−1.01	−.73	−.34	.62
EF			−.37	.02	.93
R1				1.52	5.08
R2					3.47

$\tau_1 = 22$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	1.81	2.30	.31	.78	1.71
E2		−.22	−.49	−.01	1.05
EF			−.36	.08	.98
R1				1.34	3.60
R2					2.86

Average log range: $1/\tau_1 \sum_{\tau=1}^{\tau_1} D_{t+\tau}$

$\tau_1 = 1$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	3.66	−.12	17.19	17.84	17.87
E2		−4.58	16.02	17.39	17.12
EF			17.90	18.86	19.15
R1				4.05	2.70
R2					−1.96

$\tau_1 = 5$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	2.20	.19	10.15	10.36	10.60
E2		−2.51	10.55	11.01	11.30
EF			10.63	10.99	11.43
R1				1.87	2.31
R2					.23

$\tau_1 = 22$

Over model	Superiority of model E2	EF	R1	R2	RF
E1	1.31	.42	4.70	4.96	5.18
E2		−1.10	5.56	6.00	6.27
EF			5.15	5.51	5.84
R1				1.75	1.70
R2					.06

NOTE: Each table entry represents a t-statistic of the estimate of $\mu_{i,j}$ in $\varepsilon_{j,t}^2 - \varepsilon_{i,t}^2 = \mu_{i,j} + \eta_t$, where $\varepsilon_{i,t}$ is the forecast error of model i in period t. A t-statistic >2 indicates that model i is preferred to model j, whereas a t-statistic <-2 indicates the opposite. The models are EGARCH1 (E1), partially symmetric EGARCH2 (E2), FIEGARCH (EF), REGARCH1 (R1), partially symmetric REGARCH2 (R2), and FIREGARCH (RF).

Table 11. Pairwise Comparisons of Long-Run Forecast RMSEs

Average log absolute return: $1/(\tau_1 - \tau_0 + 1)\sum_{\tau=\tau_0}^{\tau_1} \ln|R_{t+\tau}|$, $\tau_0 = 22(m-1)$, $\tau_1 = 22m$

$m = 3$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	.53	1.38	.58	1.19	2.14
model	**E2**		.83	.37	1.04	2.08
	EF			.10	.70	1.67
	R1				2.01	3.24
	R2					3.06

$m = 6$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	−.96	.36	.69	1.08	2.18
model	**E2**		1.58	.95	1.31	2.33
	EF			.54	.90	1.97
	R1				1.85	3.12
	R2					3.10

$m = 12$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	−2.06	−.91	1.11	1.29	2.54
model	**E2**		2.26	1.77	1.93	2.75
	EF			1.27	1.46	2.52
	R1				.66	2.13
	R2					2.46

Average log range: $1/(\tau_1 - \tau_0 + 1)\sum_{\tau=\tau_0}^{\tau_1} D_{t+\tau}$, $\tau_0 = 22(m-1)$, $\tau_1 = 22m$

$m = 3$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	.19	.51	3.46	3.71	4.04
model	**E2**		.38	3.89	4.29	4.70
	EF			3.59	3.93	4.45
	R1				1.96	1.41
	R2					.26

$m = 6$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	−1.66	−.72	3.17	3.39	3.87
model	**E2**		1.19	3.25	3.49	3.95
	EF			3.23	3.49	4.14
	R1				1.56	1.23
	R2					.73

$m = 12$

		Superiority of model				
		E2	**EF**	**R1**	**R2**	**RF**
Over	**E1**	−2.51	−1.87	3.51	3.74	4.06
model	**E2**		1.67	3.30	3.47	3.69
	EF			3.48	3.76	4.19
	R1				.39	.07
	R2					−.19

NOTE: Each table entry represents a t-statistic of the estimate of $\mu_{i,j}$ in $\varepsilon^2_{j,t} - \varepsilon^2_{i,t} = \mu_{i,j} + \eta_t$, where $\varepsilon_{i,t}$ is the forecast error of model i in period t. A t-statistic >2 indicates that model i is preferred to model j, whereas a t-statistic <-2 indicates the opposite. The models are EGARCH1 (E1), partially symmetric EGARCH2 (E2), FIEGARCH (EF), REGARCH1 (R1), partially symmetric REGARCH2 (R2), and FIREGARCH (RF).

REGARCH2 over FIREGARCH for forecasting the log range at a 1-day horizon.

6. CONCLUSION

The goal of this article is to provide a simple, yet highly effective framework for forecasting the volatility of asset returns by combining two-factor EGARCH models with data on the range. Our empirical analysis of S&P500 index data makes the following points:

- Incorporating the information about volatility contained in the range into EGARCH models significantly improves the in-sample fit and, more importantly, the accuracy of out-of-sample forecasts of the models.
- Two-factor and fractionally integrated range-based EGARCH models that incorporate some volatility asymmetry dominate, both in-sample and out-of-sample, the extensive set of model and data combinations that we consider. The importance of volatility asymmetries is apparent in both return-based and range-based models, whereas the value of the second factor relies on the superior information about volatility contained in the range (which mirrors the conclusion of Alizadeh et al. 2002).
- Fractionally integrated ranged-based models offer comparable and sometimes superior performance to two-factor models. Among the return-based models, fractional integration is dominated by the two-factor specifications.
- In evaluating the forecasting performance of competing model specifications, the choice of volatility measure to which the forecasts are compared matters. In particular, it is more difficult to obtain unbiased forecasts of log ranges than log absolute returns, because the latter contains much more idiosyncratic noise (or the former is more informative about the true volatility). Andersen and Bollerslev (1998) made a similar point in advocating the use of realized volatility for forecast evaluation.
- Finally, volatility is forecastable over relatively long horizons, contrary to the claims of West and Cho (1995) and Christoffersen and Diebold (2000). The fractionally integrated range-based EGARCH model forecasts the average log absolute return or log range as far as 1 year from the end of the estimation period with R^2's of .28 and .32.

ACKNOWLEDGMENTS

The authors thank Torben Andersen, Frank Diebold, Eric Ghysels, Fernando Zapatero, an anonymous associate editor, an anonymous referee, and seminar participants at the University of Pennsylvania for their comments and suggestions. Financial support from the Rodney L. White Center for Financial Research at the Wharton School of the University of Pennsylvania is gratefully acknowledged.

APPENDIX: DISTRIBUTION OF THE LOG RANGE

To keep this article relatively self-contained, we briefly review the main theoretical result of Alizadeh et al. (2002). Consider a driftless Brownian motion x, with origin x_0 and constant diffusion coefficient σ, over an interval with finite length τ. Feller (1951) derived the distribution of the range of this process. A transformation of this distribution yields the following distribution of the log range $Y_t = \ln(\sup_{0<t\leq\tau} x_t - \inf_{0<t\leq\tau} x_t)$:

$$\begin{aligned} q(\sigma, \tau, y)\, dy &= \Pr[Y_t \in dy | \sigma, \tau] \\ &= 8 \sum_{k=1}^{\infty} (-1)^{k-1} \frac{k^2 e^y}{\sigma\sqrt{\tau}} \phi\left(\frac{k e^y}{\sigma\sqrt{\tau}}\right) dy, \end{aligned} \qquad \text{(A.1)}$$

where ϕ denotes the density function of a standard normal random variable. Given the homotheticity of a driftless Brownian motion in volatility, the distribution of the log range can be rearranged as

$$\begin{aligned} q(\sigma, \tau, y)\, dy &= q(1, 1, y - .5\ln\tau - \ln\sigma)\, dy \\ &= q(y - .5\ln\tau - \ln\sigma)\, dy, \end{aligned} \qquad \text{(A.2)}$$

where

$$q(y)\,dy = 8\sum_{k=1}^{\infty}(-1)^{k-1}\,k^2 e^y \phi(ke^y)\,dy. \qquad \text{(A.3)}$$

The moments of $q(y)$, which are independent of σ and τ, can be computed using brute-force numerical integration with a truncated sum. (Specifically, we use Gaussian quadrature with integration limits of $[-2, 3]$, which, based on an initial set of simulations of y, spans a range of about 16 standard deviations centered roughly on the mean. The sum is truncated at $k = 1{,}000$.) As Alizadeh et al. (2002) reported, the mean is .426, the variance is .082, the skewness is .169, and the kurtosis is 2.801. Therefore, the log range has a mean of $.426 + .5 \ln \tau + \ln \sigma$ with the same higher-order moments. Furthermore, because the skewness and kurtosis of the log range are close to the Gaussian benchmark levels of 0 and 3, Alizadeh et al. (2002) argued that the distribution of the log range is approximately Gaussian.

[Received February 2002. Revised September 2005.]

REFERENCES

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002), "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047–1091.

Andersen, T. G., and Bollerslev, T. (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

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

Andersen, T. G., and Sørensen, B. E. (1997), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business & Economic Statistics*, 14, 328–352.

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

Barndorff-Nielsen, O. E., and Shephard, N. (2001), "Non-Gaussian Ornstein-Uhlenbeck–Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society*, Ser. B, 63, 167–241.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

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

Bollerslev, T., and Zhou, H. (2002), "Estimating Stochastic Volatility Diffusions Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33–65.

Brandt, M. W., and Diebold, F. X. (2006), "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," *Journal of Business*, 79, 61–73.

Calvet, L., and Fisher, A. J. (2002), "Multifractality in Asset Returns: Theory and Evidence," *Review of Economics and Statistics*, 84, 381–406.

Chernov, M., Ghysels, E., Gallant, A. R., and Tauchen, G. E. (2003), "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 106, 225–257.

Chou, R. Y. (2005), "Forecasting Financial Volatilities With Extreme Values: The Conditional Autoregressive Range (CARR) Model," *Journal of Money, Credit and Banking*, 37, 561–682.

Christoffersen, P., and Diebold, F. X. (2000), "How Relevant Is Volatility Forecasting for Financial Risk Management?" *Review of Economics and Statistics*, 82, 1–11.

Christoffersen, P., Jacobs, K., and Wang, Y. (2005), "Option Valuation With Long-Run and Short-Run Volatility Components," working paper, McGill University.

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R. F., and Lee, G. G. J. (1999), "A Permanent and Transitory Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, eds. R. F. Engle and H. White, pp. 475–497.

Engle, R. F., and Ng, V. K. (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

Feller, W. (1951), "The Asymptotic Distribution of the Range of Sums of Independent Random Variables," *The Annals of Mathematical Statistics*, 22, 427–432.

Gallant, A. R., Hsu, C. T., and Tauchen, G. E. (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, 81, 617–631.

Garman, M. B., and Klass, M. J. (1980), "On the Estimation of Price Volatility From Historical Data," *Journal of Business*, 53, 67–78.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Hentschel, L. (1995), "All in the Family: Nesting Symmetric and Asymmetric GARCH Models," *Journal of Financial Economics*, 39, 71–104.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12, 371–389.

Nelson, D. B. (1989), "Modeling Stock Market Volatility," in *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp. 93–98.

——— (1991), "Conditional Heteroskedasticity in Asset Pricing: A New Approach," *Econometrica*, 59, 347–370.

Newey, W. K., and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pagan, A. R., and Schwert, G. W. (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business*, 53, 61–65.

Revuz, D., and Yor, M. (1999), *Continuous Martingales and Brownian Motion*, Berlin: Springer-Verlag.

Rogers, L. C. G., and Satchell, S. E. (1991), "Estimating Variance From High, Low, and Closing Prices," *The Annals of Applied Probability*, 1, 504–512.

Schwert, G. W. (1990), "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77–102.

West, K. D., and Cho, D. (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics*, 69, 367–391.

Yang, D., and Zhang, Q. (2000), "Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices," *Journal of Business*, 73, 477–491.