INVESTIGATION OF A CLASS OF VOLATILITY ESTIMATORS

Jacob Boudoukh
is a professor at the Stern School of Business at New York University.

Matthew Richardson
is a professor at the Stern School of Business at New York University.

Robert F. Whitelaw
is a professor at the Stern School of Business at New York University.

This article examines a class of volatility estimation models, all of them based on a weighted sum of squared deviations from the mean for historical returns. We show how some popular methods, such as RiskMetrics™, GARCH, and non-parametric density estimation, fall into this class. We also conduct a brief empirical comparison of these methods. We find density estimation and RiskMetrics™ forecasts to be the most accurate for forecasting short-term interest rate volatility.

Financial institutions and corporations are becoming increasingly aware of the importance of risk management, which has led to considerable interest in tools for risk measurement. This increased interest has prompted some discussion of the ability of various techniques to forecast volatility, tail behavior, and correlations, in both the academic literature and practitioners' publications (see Figlewski [1997] for a survey; see also Longerstaey and Zangari [1995]).

Our article provides an analysis of various methods for estimating volatility. The goal is to show how some of the most prominent methods for volatility estimation are related, and, as an aside, to discuss approaches for benchmarking and comparing volatility forecasts.

From a methodological perspective, this article provides a unified framework for estimating volatility that includes, as special cases, *historical volatility, exponentially smoothed volatility, GARCH,* and *MDE* (*multivariate non-parametric density estimation*).We demonstrate that these methods differ only in the *weights* given to historical asset return data (e.g., for a currency) or to changes in the level of an economic variable (e.g., short-term interest rates). We describe the advantages and disadvantages of each of these approaches, and discuss such issues as the trade-off between parametric and non-parametric models for forecasting volatility.

A key point is that we are interested in *forecasting* volatility, which imposes a cost on highly parameterized models. Imposing as little structure as possible, via non-parametric modeling, is a priori a potentially promising approach. The empirical example is from the world of fixed-income securities, namely, forecasts of the volatility of changes in U.S. short-term interest rates.

I. A UNIFYING APPROACH

Asset returns (denoted R_{t+1}) can be broken up into a forecastable part, $\mu_t \equiv E_t[R_{t+1}]$ (i.e., the conditional mean, where E_t reflects conditioning on the information available at time t), and an unpredictable portion, ε_{t+1}:

$$R_{t+1} = \mu_t + \varepsilon_{t+1} \quad (1)$$

The conditional variance of asset returns, σ_t^2, is then just

$$\sigma_t^2 = E_t[R_{t+1}^2] - \mu_t^2 \equiv E_t[\varepsilon_{t+1}^2] \quad (2)$$

Let us abstract from the issue of estimating the conditional mean and focus on methods for estimating the conditional second moment. For the purposes of the discussion, assume $\mu_t = 0$, so that the square root of the conditional second moment is in fact volatility. This is a common assumption in the asset pricing literature, corresponding closely to the notion of market efficiency and the random walk hypothesis. In our empirical example below, we forecast the volatility of short-term interest rates. There, we chose to mean-adjust the series to allow for the effect of mean reversion in interest rates.

Most empirical methods for predicting volatility on the basis of past data start with the premise that volatility clusters through time; that is, large returns tend to be followed by additional large returns of either sign. Motivated by this viewpoint, one popular class of volatility estimates can be written as the weighted sum of past squared returns. Specifically:

$$\hat{\sigma}_t^2 \equiv E_t[R_{t+1}^2] = \omega_0 + \sum_{i=1}^{k} \omega_i(t) R_{t+1-i}^2 \quad (3)$$

where $\omega_i(t)$ are the weights, which can time-vary depending on information available today, i.e., period t; and ω_0 is a constant term (that may be zero).

The class of estimators in Equation (3) captures many of the empirical volatility forecasting methods currently used in finance. Volatility forecasting models that do not fall into this class include 1) weighting past absolute errors, a method that some believe is superior in terms of statistical robustness (to deviations from normality, for example); 2) using implied volatility from explicit asset pricing models such as Black-Scholes, which, modeling error aside, may provide a better estimate because the approach is forward-looking in nature; and 3) more complex models that cannot be written as a simple weighted sum of past changes squared (e.g., stochastic volatility models).

While these models are popular in some specific cases, they may be inappropriate or inapplicable elsewhere (e.g., using Black-Scholes for options on fixed- income securities). They thus fail to provide a unified approach to volatility forecasting.

There are various widely applicable volatility forecasting models that fall under our unifying framework. The models we examine are, for that very reason, the most popular for risk management. We show how each of the models we examine relates to the representation in Equation (3).

II. NAIVE MODELS

Historical Volatility

Perhaps the simplest and most popular way to forecast the volatility of a series is to take its historical volatility over some prior window. This model is often termed "naive" because it places no structure on how volatility might evolve through time and puts constant weights on past observations. Its advantage, of course, is that it is less subject to problems associated with overfitting, and involves only one choice variable, namely, the observation period. In terms of Equation (3),

$$\hat{\sigma}_t^2 = \frac{1}{s} \sum_{i=1}^{s} R_{t+1-i}^2 \quad (4)$$

where $\omega_i(t) = 1/s$, i.e., equal weights on the past s observations, and the constant term is zero.

Exponentially Smoothed Volatility Forecasts

Even within the class of "naive" models, there are two obvious criticisms of historical volatility. First, if volatility clusters, it follows that more recent returns should be given more weight. Returns in the last period provide more information about current volatility than returns some time ago. This is a result of the fact that the relevant state variables, which may differ across return series, are often autoregressive, or, more generally, of a periodic nature. Such is the case for financial variables such as interest rates and spreads, and for related economic variables such as real growth rates, unemployment rates, or storage costs.

Second, the choice of the length of the observation period is somewhat arbitrary. Weight is given to observations that occur within the most recent s periods, while no weight is given to observations beyond this window. This procedure is ad hoc at best.

From a practical viewpoint, volatility estimates may often exhibit spurious clustering (due to the equal weights) and spurious drops (when large observations leave the window). Informally, a reasonable way to think of the relevance and effect of economic news after it becomes publicly known is that it has a diminishing effect through time. In this sense, we may expect the pattern of volatility to be such that large increases occur every once in a while, and then volatility smoothly diminishes.

The resulting model is one where exponentially declining weights are given to past volatilities, proxied by squared returns (as in RiskMetrics™, for example). This weighting procedure gives more weight to recent observations, yet smooths the series as all observations from period to period are given only slightly different emphasis. In terms of Equation (3):

$$\hat{\sigma}_t^2 = (1 - \theta)\sum_{i=1}^{\infty}\theta^{i-1}R_{t+1-i}^2 \tag{5}$$

where $\omega_i(t) = \theta^{i-1}(1 - \theta)$, and θ equals the exponential weighting parameter ($0 < \theta < 1$).

Of course, the researcher must truncate the return series since only a finite number of observations are available for estimation. This is, however, a minor consideration with financial data that are observed with high frequency.[1] More important, the value of the weighting parameter needs to be chosen by the researcher, and the model itself implies a restrictive set of weights. One parsimonious way of choosing these weights is to estimate them from historical data using a parametric model of returns, giving rise to the ARCH/GARCH class of models.

III. ARCH/GARCH MODELS

ARCH(p) Volatility

Engle [1982] introduced the autoregressive conditional heteroscedasticity (ARCH) model for estimating and forecasting volatility. The idea behind ARCH modeling is that volatility is persistent, but the exact form of this persistence is unknown. Thus, one way to estimate volatility using the class of models in Equation (3) is to estimate the weights as a function of past data. In particular, the ARCH (p) model can be written as:

$$\hat{\sigma}_T^2 = \hat{\alpha}_0 + \sum_{i=1}^{p}\hat{\alpha}_i R_{T+1-i}^2 \tag{6}$$

where p represents the order of the ARCH process, and $\omega_i(T) \equiv \hat{\alpha}_i$, the weights from the ARCH (p) estimation.

In order to estimate parameters governing the weights, the researcher may specify the conditional distribution of returns (possibly ignoring the conditional mean). While one popular model is the normal distribution, other distributional assumptions (such as Student-t) are also possible. Although exact distributional assumptions are not in general required (see Engle and Gonzalez-Rivera [1991]), it is convenient to make explicit distributional assumptions because the estimation can then be performed easily using maximum-likelihood techniques.

The econometrician must also choose both an order for the ARCH process and the length of the sample period in which to estimate the ARCH parameters. This can lead to several difficulties in predicting volatility. First, like the historical volatility procedure, ARCH volatility forecasts can produce spurious peaks. This phenomenon is especially pronounced in the case of ARCH, because only p observations obtain any weight. Since p is most commonly small (say, one to ten), large moves will produce even larger peaks. The problem arises from the fact that researchers ignore information in returns going back more than p periods.[2]

Second, while there are standard criteria for determining the order p in-sample, the effect of estimation error on forecasting is an unknown quantity. Third, the estimation itself is subject to misspecification, depending on the appropriateness of the assumption about the returns distribution. Fourth, even under the null, the small sample properties of the estimates are weak; hence, large amounts of data may be required.

The tremendous advantage of ARCH estimation is that the weights are no longer determined in an ad hoc manner. If, for example, the weights are not equal or proportional, ARCH modeling will capture this; "naive" models will not. That is, if the true process is given by Equation (6) with any set of α_is, then, given enough data, ARCH will recover the parameters (which are the weights). Furthermore, some of the disadvantages of volatility estimation can be reduced by generalizing the ARCH model.

GARCH (p, q) Volatility

Bollerslev [1986] extended the ARCH (p) model for volatility forecasting to the now commonly used generalized autoregressive conditional heteroscedasticity [GARCH (p, q)] model. The idea behind the GARCH extension is similar to that of the exponentially smoothed version of historical volatility. That is, ARCH (p) is too restrictive because it implies that returns more than p periods ago have no influence on current volatility. GARCH (p, q) places the same type of distributional assumption on returns but now assumes that the conditional variance follows:

$$\hat{\sigma}_T^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i R_{T+1-i}^2 + \sum_{i=1}^{p} \hat{\beta}_i \sigma_{T-i}^2 \quad (7)$$

where p and q represent the order of the GARCH process, and $\hat{\alpha}_j$ and $\hat{\beta}_j$ are the parameters from the GARCH (p, q) estimation.

Equation (7), like Equation (5), can be written as an infinite weighted average of past squared returns. That is:

$$\hat{\sigma}_T^2 = \hat{\delta}_0 + \sum_{i=1}^{\infty} \hat{\delta}_i R_{T+1-i}^2 \quad (8)$$

where $\omega_i(T) = \delta_i$, and the δ_i are non-linear functions of the parameters of the GARCH (p, q) process, α and β.

To gain some intuition for the weights implied by the parameters, consider the popular GARCH (1, 1) process. In this case, Equation (8) can be rewritten as

$$\hat{\sigma}_T^2 = \frac{\hat{\alpha}_0}{1 - \hat{\beta}_1} + \sum_{i=1}^{\infty} \hat{\alpha}_1 \hat{\beta}_1^{i-1} R_{T+1-i}^2 \quad (9)$$

where $\omega_i(T) = \alpha_1 \beta_1^{i-1}$. This looks very similar to Equation (5) for the exponentially smoothed estimator of σ_T^2. In fact, exponential smoothing is a constrained version of GARCH (1, 1).

Specifically, the parameter constraints on the GARCH (1, 1) process are $\alpha_0 = 0$, $\beta_1 = \theta$, and $\alpha_1 = 1 - \theta$. Unfortunately, this means that exponentially smoothed volatility, in general, and RiskMetrics™ volatility, in particular, are based on a non-stationary GARCH process, since $\alpha_1 + \beta_1 = 1$. Thus, the process implicit in RiskMetrics™ estimates with *any* smoothing parameter has an infinite fourth moment, meaning that the (unconditional) variance of the variance process is undefined. While this non-stationarity may be viewed as a theoretically unattractive feature in time series modeling, it does not invalidate exponentially smoothed volatility as a useful approach for forecasting volatility.

On the other hand, there are still potentially significant difficulties with GARCH (p, q) processes. While GARCH solves the problem of the spurious choppiness of the estimates that is characteristic of ARCH processes, and may require fewer numbers of parameters to fit the data, the remaining difficulties are similar to the ones described for ARCH.

All the techniques discussed so far imply, for a given sample, state-independent weights. That is, the only information used is information contained in past squared returns. This restricts the type of information that can impact volatility. Of course, in a rolling estimation setting, the weights will change as observations enter and leave the estimation window. Weights will not depend on whether these observations are more or less informative, given current information, however.[3]

IV. NON-PARAMETRIC MODELS: MULTIVARIATE DENSITY ESTIMATION

Multivariate density estimation (MDE) is a method for estimating the joint probability density function of a set of variables. For example, one could choose to estimate the joint density of returns and a set of predetermined factors such as the slope of the term structure, the inflation level, and the state of the economy. From this distribution, the conditional moments, such as the mean and volatility of returns, conditional on the state, can be calculated.[4]

The MDE volatility estimate provides an intuitive alternative to the standard set of approaches to weighting past (squared) changes in determining volatility forecasts. The key feature of MDE is that the weight function is no longer constant over time as in the other methods. Instead, it depends on how the current state of the world compares to past states of the world. If the current state of the world, as measured by the state vector x_t, is similar to a particular point in the past, then that past squared return is given a lot of weight in forming the volatility forecast, regardless of how far away in time it is. This contrasts with more standard approaches, which weight past observations according to how long ago they occurred, whatever the economic environment looked like at the time.

To the extent that the state variables are autocorrelated, the MDE weights will also capture some of the decay features present in the exponentially smoothed GARCH and RiskMetrics™ volatility forecasts. That is, recent observations will be weighted more heavily.

Of course, selecting the appropriate state variables is critical to the performance of the volatility estimate. These variables should be useful in describing the economic environment in general, and be related to volatility specifically. For example, suppose that the level of return volatility is related to the level of inflation; then inflation will be a good conditioning variable. Good candidates tend to be the same variables that one might select for use in the augmented GARCH model described in endnote 3. The advantages of the MDE estimate are that it can be interpreted in the context of weighted lagged returns, and that the functional form of the weights depends on the true (albeit estimated) distribution of the relevant variables.

Using the MDE method, the estimate of conditional volatility is

$$\hat{\sigma}_t^2 = \sum_{i=1}^{s} \omega(x_{t-i}) R_{t+1-i}^2 \qquad (10)$$

where

$$\omega(x_{t-i}) \equiv K\left(\frac{x_{t-i} - x_t}{h}\right) \Big/ \sum_{i=1}^{s} K\left(\frac{x_{t-i} - x_t}{h}\right)$$

Here, x_t, the vector of variables describing the economic state at time t (e.g., the term structure), determines the appropriate weight $\omega(x_{t-i})$ to be placed on observation $t - i$, as a function of the "distance" between the state x_{t-i} at that time and the current state x_t. The relative weight of "near" versus "distant" observations from the current state is measured via the kernel function $K(\cdot)$.

The kernel function has to obey some mild regularity conditions to ensure asymptotic convergence. For the application studied in this article, we choose $K(z) = (2\pi)^{-m/2} e^{-1/2z'z}$, where m is the number of conditioning variables. The reader will recognize this as a multivariate normal density. This is chosen for convenience, and relates only to the efficiency of the density estimate; it is not a distributional assumption per se.

The bandwidth, h, is related to sample size and the relative sparseness of the data. It determines how variable the weights are. As the bandwidth is increased, the weights become less variable, and, in the limit, the conditioning information is ignored and each observation gets equal weight. The bandwidth we use, as suggested by Scott [1992], is $h_j = \hat{\sigma}_j (t)^{-1/(m+4)}$, where $\hat{\sigma}_j$ is the standard deviation of the variable x_j.

MDE is extremely flexible in allowing the researcher to introduce dependence on state variables. For example, we may choose to include past squared returns as conditioning variables. In doing so, the volatility forecasts will depend non-linearly on these past changes. Or, the exponentially smoothed volatility estimate can be added to an array of relevant conditioning variables. This may be an important extension to the GARCH class of models. Of particular note, the estimated volatility is still based directly on past squared returns and thus falls into the class of models described by Equation (3).

The added flexibility becomes crucial when there are other relevant state variables that can be added to the current "state." For example, it is possible to capture 1) the dependence of interest rate volatility on the level of interest rates, 2) the dependence of equity volatility on current implied volatilities, and 3) the dependence of exchange rate volatility on interest rate spreads or proximity to intervention bands.

There are potential costs in using MDE. To estimate volatility, the researcher must choose a bandwidth, a kernel function, a set of conditioning variables, and the number of observations to be used. For our purposes, the bandwidth and kernel function are chosen objectively (using standard criteria). Although they may not be optimal choices, it is important to avoid problems associated with data snooping and overfitting.

While the choice of conditioning variables is a matter of discretion and subject to abuse, the choice does provide a considerable advantage. Theoretical models and existing empirical evidence may suggest relevant determinants for volatility estimation, which MDE can incorporate directly. These variables can be introduced in a straightforward way for the class of stochastic volatility models in Equation (3).

The most serious problem with MDE is that it is data-intensive. Considerable data are required in order to estimate the appropriate weights that capture the joint density function of the variables. The quantity of data needed increases quickly as the number of conditioning variables used in the estimation grows. On the other hand, for many of the relevant markets this concern is somewhat alleviated since the relevant state can be adequately described by a relatively small number of factors.

V. APPLICATION: A COMPARISON OF INTEREST RATE VOLATILITY FORECASTS

Our illustration of the implementation of these methods, and how one could go about benchmarking various methods against one another, is *not meant to provide an all-inclusive search* for the optimal set of parameters and horizon lengths for each method. Instead, the results of the "horse race" between methods should be considered at best suggestive of some of the issues that might arise.

We calculate volatility estimates of daily changes in the annualized interest rate on three-month Treasury bills. Each estimate is calculated using 150 daily observations on a rolling basis. In other words, the first 150 observations in the sample are used to compute an estimate for the following day (day 151), using each of the four methodologies described below. This estimate is then evaluated on an out-of-sample basis, i.e., in terms of the realized volatility. The whole exercise is then rolled forward one day. The models are reestimated using observations 2 through 151 in order to estimate volatility for day 152.

Interest rate changes are mean-adjusted using the sample mean from the 150-day estimation period, although the ranking of the methodologies and the relative performance are not sensitive to the fact that we mean-adjust. The analysis is performed over the period 1983–1992.

The models we compare are:

- Historical standard deviation.
- Exponentially smoothed volatility with $\theta = 0.94$ (the RiskMetrics™ model).
- GARCH (1, 1) volatility (estimated via maximum likelihood).
- MDE volatility with the state variables being x_i = (interest rate level, term structure slope).[5]

Note that in every case we use only 150 observations to estimate the models. An alternative strategy would be to increase the number of observations as the out-of-sample exercise moves forward through time. As the length of estimation sample increases, however, concerns about non-stationarity do so as well. That is, the parameters of the models (e.g., GARCH) may be changing over time. We make no effort to optimize this trade-off between estimation error and capturing time-varying parameters.

To illustrate the four methodologies, Exhibit 1 gives weights on past squared interest rate changes as of June 1, 1992. The weights for the standard deviation and exponentially smoothed weighting schemes are the same every period and will vary only with the window length and the smoothing parameter. The GARCH (1, 1) weighting scheme varies with the parameters, which are reestimated daily, given each day's previous 150-day history. The date is selected at random. For that particular day, $\beta_1 = 0.74$. Given that β_1 is relatively low, it is not surprising that the weights decay relatively quickly.

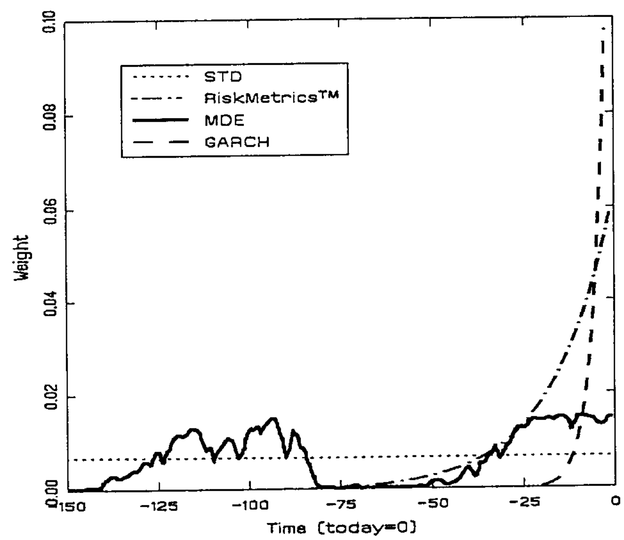
Exhibit 1 is particularly illuminating with respect to MDE. As with GARCH, the weights change over time. The weights are high for dates t (June 1, 1992) through t − 25 (twenty-five days prior), and then start to decay. The weights decrease because the economic environment, as described by the interest rate level and spread, is moving farther away from the conditions observed at date t. We do, however, observe an increase in the weights for dates t − 80 to t − 120. Economic conditions in this period (the level and spread) are similar to those at date t . MDE puts high weight on *relevant information*, regardless of how far in the past this information is.

Exhibit 2 compares, on a period-by-period basis, the extent to which the forecasts from the various models match realized future volatility. We define realized daily volatility as the average squared daily changes during the week following, from day t + 1 to day t + 5. These changes are mean-adjusted using the sample mean from the 150-day estimation period.

Note that we use this five-day average as an estimate of the realized daily volatility at time t + 1. An alternative estimate would be simply the squared interest rate change on day t + 1, but this estimate is extremely noisy. Using five observations to estimate the daily volatility reduces the magnitude of the estimation error at the cost of introducing a small bias associated with the change in daily volatility over the course of the week.

We compare realized and forecasted volatility in two ways. First, we compare the out-of-sample performance over the entire period using the mean squared error of the forecasts. That is, we take the difference between each model's volatility forecast and the realized volatility, square this difference, and average through time. Second, we regress realized volatility on the forecasts and document the regression coefficients and R^2s.

EXHIBIT 1
WEIGHTS ON PAST SQUARED INTEREST RATE CHANGES



The first part of Exhibit 2 is quite illuminating. First, while all the means of the volatility forecasts are of a similar order of magnitude (approximately 7 basis points per day), the standard deviations are quite different, with the most volatile forecast provided by GARCH (1, 1). This result is somewhat surprising, because GARCH (1, 1) is supposed to give a relatively smooth volatility estimate (due to the moving average term). For rolling, out-of-sample forecasting, however, the variability of the parameter estimates from sample to sample induces variability in the forecasts. These results are, however, upwardly biased, since GARCH would commonly require much more data to yield stable parameter estimates.

From a practical perspective, everything else the same, volatile forecasts for volatility are a disadvantage. In particular, to the extent that such numbers serve as inputs in setting time-varying rules in a risk management system (for example, by setting trading limits), smoothness of rules is a virtue.

To measure the forecasting performance of the various volatility models, Exhibit 2 reports the mean squared error (MSE). For this particular sample and window length, MDE comes in first, with the lowest MSE (0.887), and RiskMetrics™ is a close second,

EXHIBIT 2
OUT-OF-SAMPLE VOLATILITY FORECASTING

	STD	RM	MDE	GARCH
Mean	0.070	0.067	0.067	0.073
Std. Dev.	0.022	0.029	0.024	0.030
Autocorr.	0.999	0.989	0.964	0.818
MSE	0.999	0.930	0.887	1.115
LINEAR REGRESSION				
β	0.577	0.666	0.786	0.559
(s.e.)	(0.022)	(0.029)	(0.024)	(0.030)
R^2	0.100	0.223	0.214	0.172

Daily data on the three-month T-bill rate are provided by the Federal Reserve over 1983–1992 for a total of 2,250 observations. Volatility is estimated using mean-adjusted interest rate changes over a moving window of 150 days. The methods for volatility estimation are:
STD — simple standard deviation.
RM — RiskMetrics™ volatility with $\theta = 0.94$.
MDE — non-parametric multivariate density estimation using the level and spread as state variables.
GARCH — GARCH (1, 1) estimated via maximum likelihood.

Summary statistics are calculated for the standard deviation rather than the variance of interest rate changes. Interest rates are in percent; a mean forecast of 0.07 represents a standard deviation of 7 basis points per day. Comparison across methods is achieved by measuring: 1) the mean squared error (MSE) of the daily volatility forecast versus realized daily volatility $\sigma_{t+1}^{2\,*}$, which is estimated using the following five trading days, and 2) the regression coefficient and R^2 in a regression of realized volatility on the volatility forecast:

$$\sigma_{t+1}^{2\,*} = \alpha + \beta\hat{\sigma}_t^2 + \varepsilon_{t+1}$$

with an MSE of 0.930. Note that, while GARCH comes out the worst here, this comparison involves just one particular GARCH model [i.e., GARCH (1, 1)], over a short estimation window.

One would need to investigate other window lengths and specifications, as well as other data series, to reach more general conclusions. It is interesting to note, however, that, non-stationarity aside, exponentially smoothed volatility is a special case of GARCH (1, 1) in-sample, as shown above. The results here suggest the potential cost of estimation error on an out-of-sample basis.

An alternative approach to benchmarking the various volatility forecasting methods is via linear regression of realized volatility on the forecast. If the conditional volatility is measured without error, then the coefficient should equal one; if the forecast is unbiased but contains estimation error, then the coefficient will be biased downward. Deviations from one reflect a combination of this estimation error plus any systematic over- or underestimation.

The ordering in this "horse race" is quite similar to that in the previous one. In particular, MDE exhibits the β coefficient closest to one (0.786), and exponentially smoothed volatility comes in second, with a β of 0.666. The goodness of fit measures, the R^2s of the regressions, are similar for both methods.

VI. CONCLUDING REMARKS

We have provided a unifying framework for considering a number of popular volatility estimation methods, including exponentially smoothed historical volatility and GARCH, as well as a non-parametric method that uses conditioning information. These methods are illustrated and compared in an out-of-sample daily interest rate volatility forecasting exercise.

One natural and important question that we do not address is how to extend these volatility forecasts to longer horizons. For historical volatility estimates, the answer is simple — multiply the daily volatility estimate by the horizon length. That is, the volatility forecast over the next n days is simply $n\hat{\sigma}_t^2$, where the daily forecast $\hat{\sigma}_t^2$ comes from Equation (4) or Equation (5). ARCH and GARCH processes imply more complex, but still tractable, multiperiod volatility forecasts based on Equations (6) and (7).

The only caveat in these cases is that, while the one-day innovation is normally distributed for standard specifications, the multiperiod innovation has a much more complicated distribution, which makes interpretation of the volatility estimate more difficult. Finally, multiperiod volatility forecasting using MDE requires more work. Daily volatility depends on the level of the conditioning variables, so multiperiod volatility depends on the future levels of these conditioning variables.

One might think of estimating the one-day ahead distribution of these variables using the same non-parametric methods and conditioning information as those used to estimate volatility. Expectations of the convolution of this distribution function provide the appropriate values of the conditioning variables for the calculation of weights as in Equation (10).

ENDNOTES

[1]Using one year of past data (250 daily observations) and a smoothing parameter of 0.96, for example, the truncation is done at a point where the weight on all the omitted observations would have been 0.000037. Note also that with high-frequency data, the chosen smoothing parameter is generally close to 1 so as to give non-negligible weight to more than just the few most recent observations.

[2]Note that information beyond p periods is used in estimating the parameters α; to the extent that these parameters are constant, however, the information is not explicitly incorporated into the volatility forecasts.

[3]There are several extensions to GARCH models that do allow other information to enter the estimation, but these cannot be written in the form of Equation (3). For example, modified GARCH models specify conditional volatility in terms of other predetermined variables, X_t; that is:

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i R_{t+1-i}^2 + \sum_{i=1}^{p} \hat{\beta}_i \sigma_{t-i}^2 + \sum_{i=1}^{r} \hat{\gamma}_i X_{t+1-i}$$

Conditional volatility above will depend on past values of X_t directly and not through the weights on past returns. Thus, modified GARCH models lie outside the family of volatility models described in Equation (3). Depending on the particular application, however, these models may be a useful alternative to models that do allow state-dependent weights.

[4]For more details on the intricacies of density estimation as it applies to our problem, and for relevant econometric references, please see Boudoukh et al. [1996].

[5]The state variables are defined as the yield on three-month T-bills and the spread between the yields on ten-year Treasury notes and three-month T-bills. This choice is motivated by results in the literature (see, for example, Boudoukh et al. [1996]).

REFERENCES

Bollerslev, Tim. "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics*, Vol. 31, 3 (1986), pp. 307-328.

Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert Whitelaw. "The Stochastic Behavior of Interest Rates." Working paper, Stern School of Business, New York University, 1996.

Engle, Robert F. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, Vol. 50, 4 (1982), pp. 987-1008.

Engle, Robert F., and Gloria Gonzalez-Rivera. "Semiparametric ARCH Models." *Journal of Business Economics and Statistics*, Vol. 9, 4 (1991), pp. 345-360.

Figlewski, Stephen. "Forecasting Volatility." *Financial Markets, Institutions and Instruments*, Vol. 6 (1997).

Longerstaey, Jacque, and Peter Zangari. "A Transparent Tool." *Risk*, January 1995, pp. 30-32.

Scott, David W. *Multivariate Density Estimation: Theory, Practice and Visualization.* New York: John Wiley & Sons, Inc., 1992.