CEMA WORKING PAPER SERIES

Forecasting Nonlinear Functions of Returns Using LINEX Loss Functions

Soosung Hwang
Department of Banking and Finance
City University Business School
London, UK

John Knight
Department of Economics
University of Western Ontario
London, Ontario, Canada

Stephen E. Satchell
Trinity College and
Faculty of Economics and Politics
University of Cambridge
Cambridge, UK

Working Paper 49

May 2000

Working Paper 49, 1–27

Forecasting Nonlinear Functions of Returns Using LINEX Loss Functions

Soosung Hwang

Department of Banking and Finance
City University Business School
London, UK

John Knight*

Department of Economics
University of Western Ontario
London, Ontario, Canada

and

Stephen E. Satchell

Trinity College and
Faculty of Economics and Politics
University of Cambridge
Cambridge, UK

This paper applies LINEX loss functions to forecasting nonlinear functions of variance. We derive the optimal one-step-ahead LINEX forecast for various volatility models using data transformations such as $ln(y_t^2)$ where y_t is the return of the asset. Our results suggest that the LINEX loss function is particularly well-suited to many of these forecasting problems and can give better forecasts than conventional loss functions such as mean square error (MSE).

Key Words: LINEX Loss Function; Forecasting, Volatility.
JEL Classification Numbers: C22, C53.

* Corresponding Author: John Knight, Department of Economics, University of Western Ontario, London, Ontario Canada N6A 5C2. E-mail: jknight@julian.uwo.ca

1. INTRODUCTION

Forecasting volatility is a major issue in finance. For example, volatility forecasts are used to price options and to forecast option prices; they can be used to produce confidence intervals for the prices of the underlying assets and the forecasts can be used as a component of multi-period investment strategies. Volatility forecasts are also an integral part of forecasting value at risk. The recent growing concern about risk management and the rapid growth in financial derivative markets has resulted in volatility forecasting attracting a great deal of interest.

The major development in modelling and forecasting volatility has been the introduction of ARCH models by Engle (1982). Since then, numerous conditional volatility models have been suggested and tested. Empirical evidence suggests that volatility, however measured, has strong autocorrelations over time, see Ding, Granger, and Engle (1994). Many studies on volatility forecasting use symmetric loss functions to evaluate the efficacy of volatility forecasts; e.g., mean squared error or mean absolute error. We refer readers to Day and Lewis (1992), Engle, Hong, Kane, and Noh (1993), Harvey and Whaley (1992), Lamoureux and Lastrapes (1993), Noh, Engle, and Kane (1994), Hwang and Satchell (1998), and Knight and Satchell (1998b) for more details on volatility forecasting.

There have been a number of papers concerning the appropriateness of using symmetric loss functions to evaluate the efficacy of forecasts. Most studies on asymmetric loss functions have concentrated on the return process. See Varian (1975), Zellner (1986), Christoffersen and Diebold (1996, 1997), and Batchelor and Peel (1998) for example. These studies suggest that a symmetric loss function is not, in general, appropriate, and that other approaches need to be considered. This is because symmetric loss functions weigh returns above the mean as heavily as those below, which could be somewhat counter-intuitive to common notions of risk. Asymmetric loss functions such as semi-variance or lower partial moments are more appropriate for investors who want to consider downside risk.

In this study we advocate the use of an asymmetric loss function, in particular, the LINEX loss function for optimal forecasts of volatility processes and when the variable of interest is some function of returns. The rationale for the use of asymmetric loss function for the forecast of volatility may not be the same as those for the return process in the above. Empirically, we find that forecast errors from certain volatility models such as GARCH models are heavily skewed to the right (positive skewed) and asymmetric. The empirical results of the GARCH forecasts suggest that GARCH forecasts may not be optimal for an investor with a specific utility function.

Recent results by Bollerslev, Diebold, and Labys (1999) indicate that volatility can be observed by measuring daily volatility with summations of intra-day squared returns. They show that the volatility is model free and has very small measurement error. Whilst these are exciting results, high frequency data for returns where transaction costs are high and liquidity is low could not allow us to "converge" to continuous time processes in the required way. The conditions we describe above are likely to be met in all but a few financial markets, the exceptions being foreign exchange and some derivative (futures) markets. Therefore, we have to resort to comparing our forecasts against squared returns or some other non-linear transformation even though this is essentially using a volatility proxy with a lot of noise.

This study uses the LINEX loss function to proxy a utility function which enjoys certain optimal properties. Our results show that under the assumption of a LINEX loss function, the optimal predictor for a volatility process is the sum of conditional volatility and an adjustment factor. Under the assumption of normality the adjustment factor becomes a constant which is a function of an asymmetry parameter. If we do not assume conditional normality, then we need higher conditional moments for the volatility forecasts as an adjustment factor. These results are similar to the results of Christoffersen and Diebold (1997).

Our study is an extension of previous studies especially Christoffersen and Diebold (1996, 1997) so that volatility processes are discussed rather than return processes. Christoffersen and Diebold (1997) showed that the optimal LINEX predictor of a return process is the sum of the conditional expected return and a loss function that includes conditional higher moments including the second moment. They also showed that when returns are conditionally normal, the optimal LINEX predictor is sum of the conditional expected return and a loss function that includes the conditional variance.

This study focuses on the optimal volatility forecasts under the assumption of an asymmetric loss function. In the following sections, we first show why we need asymmetric loss functions to obtain the optimal forecasts in volatility processes. Then, the optimal forecasts with a LINEX loss function will be derived. It turns out that the LINEX optimal forecasts can be explicitly computed for a range of currently used volatility models. We extend the results of Christoffersen and Diebold (1997) by presenting results for conditional and unconditional one-step-ahead forecasts for GARCH, Exponential GARCH, stochastic volatility, and a moving average conditional heteroskedasticity model. Finally, an empirical example using LINEX forecasts will be shown and conclusions follow.

2. PROPERTIES OF FORECASTING ERROR OF GARCH (1,1) MODEL

In this section, we first calculate the properties of forecast errors to investigate if a symmetric loss function is an appropriate tool for volatility forecasting. We use a GARCH(1,1) model for the FTSE100 index. Although we could use other volatility models, we focus our attention on GARCH(1,1) because of its great popularity.

The return volatility is calculated from the log-return less the mean log-return. In what follows, we shall use y_t^2 for the return volatility at time t. More formally, y_t^2 is obtained from log-return series, r_t, as follows:

$$y_t^2 = 250[r_t - \overline{r_t}]^2$$

where the number 250 is used to annualise the squared daily return series and $\overline{r_t}$ is the in-sample mean of r_t at time t. Note that $\overline{r_t}$ is calculated using only past observations to avoid any look-ahead bias. We use a total number of 2044 daily log-returns from 21 January 1992 to 20 January 2000, which is the full set of data available to us.

We use a rolling sample of the past volatilities. On day t, the conditional volatilities of the next 60 periods ahead, $t+1, t+2, ..., t+60$, are constructed by using the estimates which are obtained from only the past observations. Therefore, allowing 60 forecasting horizons and 250 iterations from the total 2044 observations, we have 1734 observations to estimate the model. By recursive substitution of the conditional volatility, a set of one to 60 steps ahead forecasts is constructed. On the next day $(t+1)$, using recent 1734 observations (i.e., we drop the first observtion and add the observation of $t+1$), we estimate the parameters again and get another set of one to sixty steps ahead forecasts. The estimation and forecasting procedures are performed 250 times using rolling windows of 1794 observations. Estimations are carried out using the Berndt, Hall, Hall, and Hausman (BHHH) algorithm for the maximisation of the log-likelihood of the GARCH (1,1) model.

To investigate whether or not the out-of-sample forecasts above have consistent properties over different time periods and different returns, we use the following additional return series. We divide our entire sample into two equal subperiods and apply the same procedure. Here, each subsample consists of 1022 observations, of which we have 712 observations for the estimation, 60 observations for forecasting horizon and 250 observations for the iteration. We also use the S&P500 index daily log-returns; a total of 2088 observations from 21 January 1992 to 20 January 2000. These results are similar and only the FTSE100 case is reported here.[1]

[1]Other results can be obtained upon request.

The GARCH(1,1) model we use in this example is

$$\begin{aligned} y_t &= z_t h_t^{1/2} \\ h_t &= \alpha + \beta h_{t-1} + \gamma y_{t-1}^2 \end{aligned} \tag{1}$$

where $z_t \tilde{} N(0,1)$. Let f be the forecast horizon. The one step ahead forecast, $h_{t+1|t}$, and the f step ahead forecasts, $h_{t+f|t}$, of the GARCH(1,1) model are

$$\begin{aligned} h_{t+1|t} &= \alpha + \beta h_t + \gamma y_t^2, \\ h_{t+f|t} &= \alpha \sum_{j=0}^{f-1} (\beta+\gamma)^j + (\beta+\gamma)^{f-1}(\beta h_t + \gamma y_t^2), \text{ when } f > 1, \end{aligned}$$

where $h_{t+f|t}$ represents f step ahead volatility conditional on the information available at time t. For large f, $h_{t+f|t}$ approaches $\frac{\alpha}{1-\beta-\gamma}$, the unconditional varaince for the case $0 < \beta + \gamma < 1$.

Table 1 reports forecast errors of various forecast horizons. We choose f=1,5,20,60, which roughly represent one day, one week, one month, and one quarter for the forecast horizon. We use three most widely used non-linear functions of return that are used for risk evaluation. These are y_t^2, $|y_t|$, lny_t^2, see Ding, Granger, and Engle (1994) for example.

The forecast errors for the three non-linear functions of return are defined as follows. The forecast errors for the conditional variance (panel A), $v_{v,t,f}$, and the conditional standard deviation (panel B), $v_{s,t,f}$, are defined as

$$v_{v,t,f} = y_{t+f}^2 - h_{t+f|t}, \tag{2}$$

$$v_{s,t,f} = |y_{t+f}| - h_{t+f|t}^{1/2}, \tag{3}$$

and those for the conditional log-variance (panel C), $v_{l,t,f}$, are

$$v_{l,t,f} = \ln(y_{t+f}^2) - \ln(h_{t+f|t}). \tag{4}$$

A few interesting points can be made. First, panel A of table 1 shows that when GARCH(1,1) forecasts are measured by (2), they perform well; the average value of the forecast errors is very close to zero and the standard deviation of the forecast errors is small. This is because

$$h_{t+f|t} = E(y_{t+f}^2)$$

for all f in the GARCH(1,1) model. However, other measures such as (3) and (4) show that GARCH(1,1) forecasts are always larger than the

TABLE 1.

Properties of Out-of-Sample Forecasting Error of Conditional Volatility from GARCH(1,1) Model

A. Conditional Variance[1]

Forecasting Horizon	1	5	20	60
Mean	−0.0011	−0.0004	−0.0008	−0.0005
Standard Deviation	0.0529	0.0526	0.0509	0.0467
Skewness	2.2853*	2.2419*	2.4569*	3.1360*
Excess Kurtosis	6.9073*	6.7349*	7.5059*	16.1381*

B. Conditional Standard Deviation[2]

Forecasting Horizon	1	5	20	60
Mean	−0.0398	−0.0383	−0.0388	−0.0371
Standard Deviation	0.1196	0.1195	0.1157	0.1115
Skewness	0.8340*	0.7833*	0.9785*	0.9364*
Excess Kurtosis	0.5301	0.4444	0.6739*	1.1508*

C. Conditional Log-Variance[3]

Forecasting Horizon	1	5	20	60
Mean	−1.2936	−1.2788	−1.2820	−1.3297
Standard Deviation	2.2989	2.2958	2.2491	2.3373
Skewness	−1.4513*	−1.4374*	−1.4019*	−1.4101*
Excess Kurtosis	2.8996*	2.8060*	2.8344*	2.6423*

Notes: FTSE100 index daily log-returns were used for the our-of-sample forecast test of the GARCH(1,1) model. Total number of observations is 2044 from 21 January 1992 to 20 January 2000.
* represents significance at 95% level. The forecast errors used in the above panels are defined as follows.
1. Forecasting errors of conditional variance for forecast horizon f, v_{t+f}, are defined as

$$v_{t+f} = y_{t+f}^2 - h_{t+f}$$

where y_{t+f}^2 is realised variance at time $t+f$ and h_{t+f} is GARCH(1,1) forecasted variance for forecast horizon f at time t.
2. Forecasting errors of conditional standard deviation for forecast horizon f, v_{t+f}, are defined as

$$v_{t+f} = |y_{t+f}| - h_{t+f}^{1/2}$$

where $|y_{t+f}|$ is realised standard deviation at time $t+f$ and $h_{t+f}^{1/2}$ is GARCH(1,1) forecasted standard deviation over forecast horizon f at time t.
3. Forecasting errors of conditional log-variance for forecast horizon f, v_{t+f}, are defined as

$$v_{t+f} = \ln(y_{t+f}^2) - \ln(h_{t+f})$$

where $\ln(y_{t+f}^2)$ is realised log-variance at time $t+f$.

FIG. 1a. Realised Volatility and One-step-ahead Forecasted Volatility from GARCH(1,1) Model

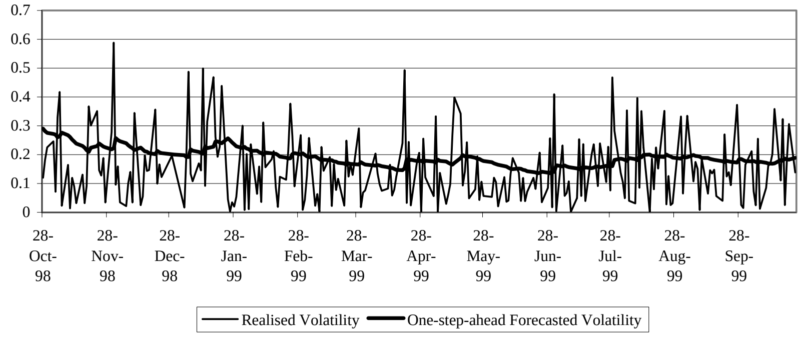
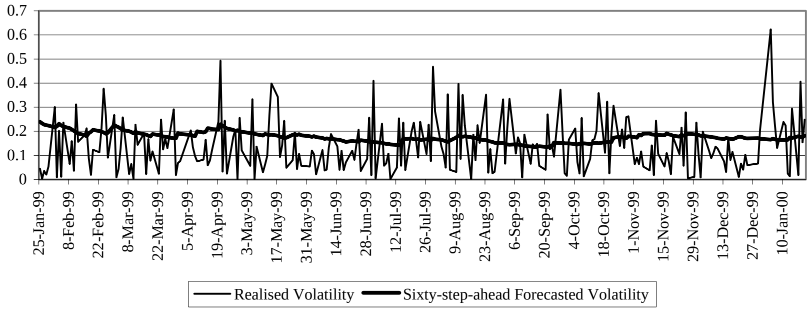


FIG. 1b. Realised Volatility and Sixty-step-ahead Forecasted Volatility from GARCH(1,1) Model



realised volatility over all horizons. See panels B and C. This is because by Jensen's inequality,

$$\begin{aligned}
\ln h_{t+1} &= \ln[E(y_{t+1}^2|\Omega_t)] > E[\ln y_{t+1}^2|\Omega_t], \\
h_{t+1}^{1/2} &= [E(y_{t+1}^2|\Omega_t)]^{1/2} > E[|y_{t+1}||\Omega_t],
\end{aligned}$$

and thus, the expected forecast errors for the conditional standard deviation and log-variance are expected to be negative;

$$\begin{aligned}
E(v_{s,t,f}) &= E[|y_{t+f}||\Omega_t] - h_{t+f}^{1/2} < 0, \qquad (5) \\
E(v_{l,t,1}) &= E[\ln(y_{t+1}^2|\Omega_t)] - \ln(h_{t+1}) < 0.
\end{aligned}$$

An interesting and important finding is that the forecast errors, as defined by (2), (3) and (4), are not symmetric. See panels A, B and C of table 1. They are significantly positively skewed (or negatively skewed in the conditional log-variance). Figure 1 shows GARCH(1,1) forecasts and realised volatility for one-step ahead forecasting and sixty-step ahead forecasting. The figure shows that the frequency of large shocks is less than that of small shocks, and volatility models are inadequate in explaining and predicting the large unanticipated shocks.[2] Figures 2 and 3 show the empirical distributions of the forecast errors of (2) and (4). All of them display forecast errors which are not symmetric. We expect that stochastic volatility (SV) forecasts have similar properties since their asymptotic properties of the two models are similar under certain conditions, see Nelson and Foster (1994), Nelson (1996).

The models such as GARCH models or SV models do not reflect investors' attitude to different levels of risk. It seems plausible that many investors pay more attention to a few high volatilities rather than a large number of lower-than-average volatilities. We need an appropriate loss function to reflect investors different attitude to high and low volatilities.

One method to obtain the optimal forecasts for investors who have different utilities for different levels of volatilities is to use an asymmetric loss function. The optimal predictor for volatility processes can be derived under the asymmetric loss function.

3. LINEX LOSS FUNCTION, OPTIMAL FORECASTS AND UTILITY MOTIVATION

In this section we consider some alternative procedures for forecasting that take into account the asymmetry of loss. We shall initially consider LINEX loss functions, see Varian (1975), Zellner (1986), and Christoffersen and Diebold (1996, 1997) for the detailed explanation of this method. One of the most significant differences between the most frequently used loss function, i.e., the mean square loss function, and LINEX loss functions is that the mean square loss function is symmetrical, while LINEX loss functions are asymmetic.

The asymmetric LINEX loss function $L(x)$ is given by:

$$L(x) = \exp(-ax) + ax - 1 \tag{6}$$

where x is the loss associated with the predictive error and a is a given parameter. With an appropriate LINEX parameter a, we can reflect small (large) losses for underestimation or overestimation. In particular, a negative a will reflect small losses for overprediction and large losses for underprediction.

[2]See Hwang (1997) for an application of outlier detection models to volatility models.

FIG. 2a. Empirical Distribution of Forecasting Error of One-step-ahead Conditional Standard Deviation from GARCH(1,1) Model

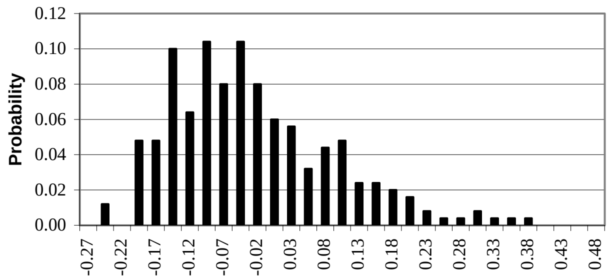


FIG. 2b. Empirical Distribution of Forecasting Error of Five-step-ahead Conditional Standard Deviation from GARCH(1,1) Model

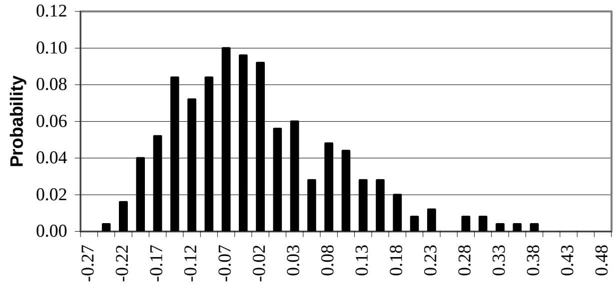


FIG. 2c. Empirical Distribution of Forecasting Error of Twenty-step-ahead Conditional Standard Deviation from GARCH(1,1) Model

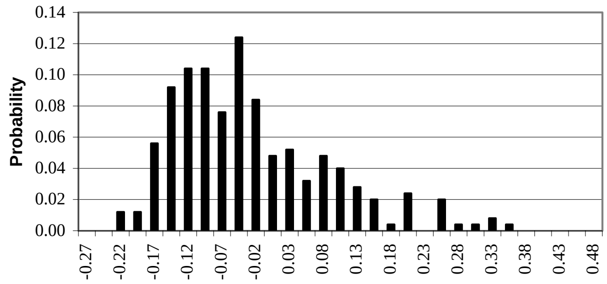


FIG. 2d. Empirical Distribution of Forecasting Error of Sixty-step-ahead Conditional Standard Deviation from GARCH(1,1) Model

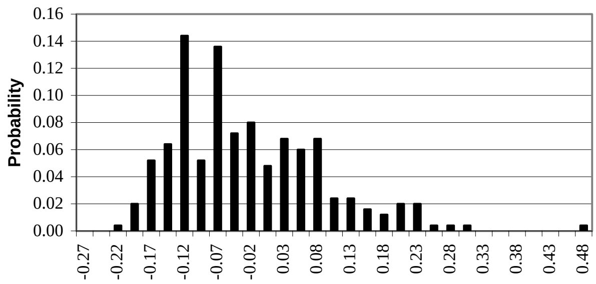


FIG. 3a. Empirical Distribution of Forecasting Error of One-step-ahead Conditional Log-Variance from GARCH(1,1) Model

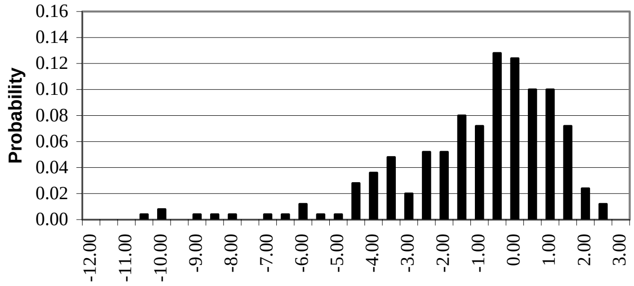


FIG. 3b. Empirical Distribution of Forecasting Error of Five-step-ahead Conditional Log-Variance from GARCH(1,1) Model

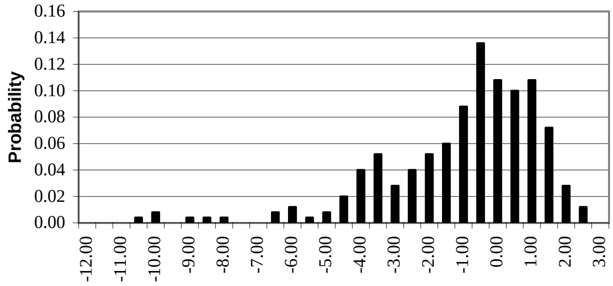


FIG. 3c. Empirical Distribution of Forecasting Error of Twenty-step-ahead Conditional Log-Variance from GARCH(1,1) Model

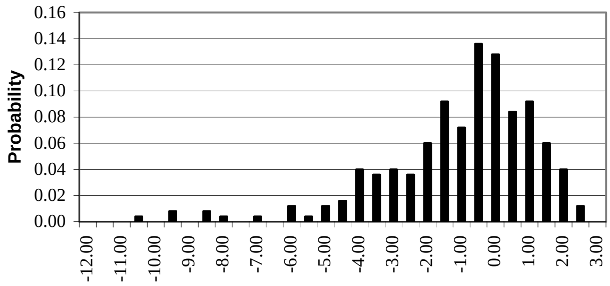
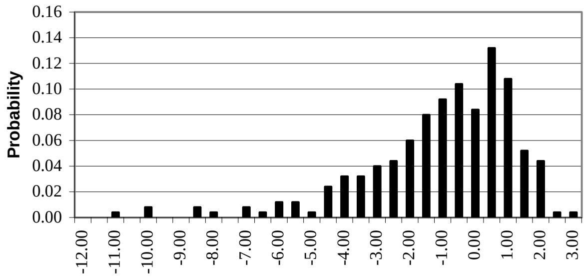


FIG. 3d. Empirical Distribution of Forecasting Error of Sixty-step-ahead Conditional Log-Variance from GARCH(1,1) Model



A forecast h is computed by carrying out the following optimization

$$\min_h \int L(y-h)pdf(y)dy \tag{7}$$

where y is the variable we wish to forecast. $pdf(y)$ is the unconditional or conditional probability function of y, depending on the context. If we substitute (6) into (7) we see that

$$\int L(y-h)pdf(y)dy = \exp(ha)m_y(-a) + a\mu_y - ah - 1$$

where $m_y(t)$ is the moment generating function of y evaluated at t, $\mu_y = E(y)$. Differentiating the above with respect to h, we find that the optimal h is given by

$$\widehat{h} = -\ell n(m_y(-a))/a \tag{8}$$

This is essentially the result given in equation (3.2) in Zellner (1986).

Consider some fairly general returns process, y_t

$$y_t = \mu_t + \sigma_t e_t \tag{9}$$

where μ_t is a deterministic mean and σ_t^2 is the conditional variance, e_t is $N(0,1)$, the unconditional mgf of $y_t, m_y(-a)$, is given by

$$m_y(-a) = \exp(-a\mu_t)m_{\sigma_t^2}(\frac{a^2}{2})$$

where $m_{\sigma_t^2}(\cdot)$ is the unconditional mgf of the stochastic volatility process.

It follows immediately that the optimal unconditional LINEX forecast h_t is given by

$$\widehat{h}_t = \mu_t - \ell n(m_{\sigma_t^2}(\frac{a^2}{2}))/a. \tag{10}$$

For $a > 0$, the extra term can be positive or negative depending on the distribution of σ_t^2. Furthermore, the expectation may only be defined for some values of a.

To illustrate the above, consider σ_t^2 following a $\chi^2(m)$ distribution, then

$$\widehat{h}_t = \mu_t + \frac{m}{2a}\ell n(1-a^2),\ 0 < a < 1$$

where μ_t is assumed deterministic.

In general, from (9)

$$\begin{aligned}\widehat{h} &= -\frac{\ell n(m_y(-a))}{a} \\ &= -\frac{1}{a}\ell n(\exp(-\mu a)E_{\sigma_{t+1}}(m_e(-a\sigma_{t+1})) \\ &= \mu - \frac{1}{a}\ell n E_{\sigma_{t+1}}(m_e(-a\sigma_{t+1}))\end{aligned} \tag{11}$$

where $m_e(-a) = E[\exp(-ae_t)]$.

Another justification for LINEX can be derived from expected utility. Bell (1995) presents the optimal properties of the utility function

$$u(w) = w - \lambda_1 \exp(-\lambda_2 w)$$

where λ_1 and λ_2 are positive constants. But this is just the LINEX loss function appropriately re-scaled; it is known that expected utility is invariant to multiplication by a positive constant, so choosing the forecast that minimises LINEX has an analogue in maximising expected utility. Bell shows in Theorem 3 (p29, Bell, 1995) that in a certain sense, this is the only utility function possessing certain desirable properties (see Bell (1995) for further details). It is likely that LINEX will enjoy similar desirable properties, although we do not explore this further.

4. LINEX VOLATILITY FORECASTS

Christoffersen and Diebold (1997) (CD) have examined the properties of LINEX forecasts for return process under the assumption that the statistical process is conditionally normal. We shall assume normality where distributional assumptions are required. However, there is accummulated evidence that innovations are non-normal even after GARCH type modelling has been done. Many of our formulae in this section could be analysed for non-normal distributions with known moment generating functions. We do not pursue this further.

We would write this as $y_{t+h}|\Omega_t \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$ where Ω_t is the information set up to time t, typically $\Omega_t = \{y_1, ..., y_t\}$, and where $\mu_{t+h|t}$ and $\sigma^2_{t+h|t}$ are the mean and variance of y_{t+h}, conditional on Ω_t, we can write $y_{t+h}|\Omega_t$ as $y_{t+h|t}$.

As shown in the previous section, the conditional volatility process $\sigma^2_{t+h|t}$ may not be optimal. The motivation for this paper is to extend CD's results to volatility forecasts. In this section we derive, in closed form where possible, conditional and unconditional LINEX forecasts for SV models and for the E-GARCH model of Nelson (1991) and a volatility process due to Knight and Satchell (1998b).

4.1. Conditioning on past information and volatility models.

We shall denote Ω_t as the information set appropriate to the conditioning. Whilst it is obvious that we would include $y_1, ..., y_t$ in Ω_t, it is by no means clear that conditional volatility, $h_1, ..., h_t$, should also be included since these variables are not observed by the econometrician for any of the models that shall be discussed in this section. However, the convenient assumption that the investors know the true parameter values but not the econometrician can be used to give a definition of available information. For this reason we shall adopt the following definition

Definition 4.1. We say that conditional volatility of time t, h_t, belongs to the conditioning set Ω_t if h_t can be computed exactly given knowledge of the true parameters, appropriate initial values for the stochastic process governing h_t, and the observed data, $y_1, ..., y_t$.

We shall apply Definition 1 when considering the different models under consideration. Summarising these future results we note that for a GARCH (1,1), where $h_t = \alpha + \beta h_{t-1} + \gamma y_{t-1}^2$, we could compute $h_1, ..., h_{t+1}$ given h_0, α, β, γ and $\{y_1, ...y_t\}$ so that $h_1, ..., h_{t+1}$ are clearly in Ω_t. Turning now to a stochastic volatility model (SVM), $y_t = z_t e^{(\xi + h_t)/2}$ and $h_t = \lambda + \alpha h_{t-1} + \nu_t$, it is apparent that knowledge of h_0, λ, ξ, α and $\{y_1, ...y_t\}$ is not enough to compute $h_1, ..., h_t$ so that these variables are not in Ω_t. It is interesting to see that Nelson's Exponential GARCH model (Nelson, 1991) has the same properties as GARCH as does the Knight and Satchell (1,1) model (Knight and Satchell, 1998b). See the following subsections for the definitions of models and further discussions.

4.2. Log-Volatility

We first need to calculate the loss associated with the prediction error. The prediction errors such as (2) and (3) are not appropriate to reflect the investors attitude to the different levels of volatilities in ARCH or SV models.

The optimal volatility predictor with the LINEX loss function needs some modification on the definition of volatility. In this study, we use the logarithmic transformation of volatility; i.e., $\ell n y_t^2$ for the realised volatility and derive the optimal log-volatility forecast for $\ell n y_t^2$. With this transformation, log-volatility in ARCH and SV models now becomes the sum of a log-chi-square variable and a log-conditional volatility (an unobserved volatility process in SV models), and thus we can calculate the loss associated with the predictive error; that is, the difference between realised log-volatility and a forecasted log-volatility.

However, logarithmic transformation of the conditional volatility of ARCH models, $\ell n h_t$, is not the optimal forecast for the log-volatility in the conventional mean square forecast error; the logarithmic value of ARCH conditional volatility is always biased upward. For example, the log-volatilty of ARCH models is

$$\ell n y_t^2 = \ell n z_t^2 + \ell n h_t.$$

Therefore,

$$\begin{aligned} E[\ell n y_t^2 | \Omega_{t-1}] &= E[\ell n z_t^2 | \Omega_{t-1}] + E[\ell n h_t | \Omega_{t-1}] \\ &= E[\ell n z_t^2] + \ell n h_t \\ &= -1.27 + \ell n h_t, \end{aligned}$$

and

$$E[\ell n y_t^2 | \Omega_{t-1}] - \ell n h_t = -1.27 < 0 \tag{12}$$

since h_t is conditional variance and for standard normal variable z_t, $E[\ell n z_t^2] = -1.27$. Equation (12) is the detailed explanation of (5).

Therefore, we need to adjust the bias in (12) which can be removed with the LINEX parameter.

4.3. ARCH Family Models

The ARCH family process is defined by

$$y_t = z_t h_t^{1/2}$$

where $z_t \sim iid\ N(0,1)$ and the conditional volatility, h_t, is a linear function of lagged values of h_t and/or y_t^2. For example, for the GARCH(p,q) process,

$$h_t = \alpha + \beta(L)h_t + \gamma(L)y_t^2,$$

where $\beta(L) = \beta_1 L + \beta_2 L^2 +, ..., +\beta_p L^p$ and $\gamma(L) = \gamma_1 L + \gamma_2 L^2 +, ..., +\gamma_q L^q$. The conditional volatility for the fractionally integrated GARCH(p,d,q) (FIGARCH) process introduced by Baillie, Bollerslev, and Mikkelsen (1996) is

$$h_t = \alpha + \beta(L)h_t + [1 - \beta(L) - \phi(L)(1-L)^d]y_t^2$$

where $\phi(L)$ is a polynomial of order $max\{p,q\} - 1$.

We shall compute conditional forecasts for $\ell n y_t^2$. The information set, according to Definition 1, includes $h_1, ..., h_{t+1}$. Firstly,

$$\begin{aligned} \ell n y_t^2 &= \ell n z_t^2 + \ell n h_t \\ &= \ell n \chi^2_{(1)} + \ell n h_t \end{aligned} \tag{13}$$

Thus the moment generating function of $\ell n y_t^2$ is

$$E[e^{-a\ell n(y_t^2)}] = E[e^{-a\ell n \chi^2_{(1)}}]E[e^{-a\ell n h_t}] \tag{14}$$

THEOREM 4.1. *Optimal One-step-ahead Conditional Forecast of $\ell n y_t^2$ and y_t in ARCH Family Models. The LINEX optimal conditional forecast of $\ell n(y_t^2)$ is*

$$E[\ell n(y_t^2)|\Omega_{t-1}] = \ell n(h_t) + \ell n(2) - \frac{1}{a}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}], \tag{15}$$

where h_t is defined by the conditional volatility model and the $a < \frac{1}{2}$.

Proof. The conditional moment generating function is

$$\begin{aligned} m_{\ell n(y_t^2)}(-a)|_{\Omega_{t-1}} &= E[e^{-a\ell n(y_t^2)}|\Omega_{t-1}] \\ &= E[e^{-a\ell n(y_t^2)}|\ell n h_t] \\ &= E[e^{-a\ell n \chi^2_{(1)}}]e^{-a\ell n h_t} \\ &= 2^{-a}\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}h_t^{-a} \end{aligned} \tag{16}$$

using Lemma A.1 in the Appendix. Thus

$$\begin{aligned} E[\ell n(y_t^2)|\Omega_{t-1}] &= -\frac{\ell n(m_{\ell n(y_t^2)}(-a))}{a}|_{\Omega_{t-1}} \\ &= \ell n(h_t) + \ell n(2) - \frac{1}{a}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}]. \end{aligned} \tag{17}$$

∎

The above theorem suggests that under the assumption of the normality of z_t and LINEX loss function, the optimal conditional forecasts for $\ell n(y_t^2)$ have an log correction factor (LCF), $\ell n(2) - \frac{1}{a}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}]$, which is constant and a function of LINEX parameter a.

We now investigate the effects of the LINEX paremater on the LCF. As in the above theorem, we require $a < 0.5$. However, this is not a restriction for the LINEX optimal volatility forecasts, since when $a \to 0.5$, $LCF \to -\infty$. This means that $-\infty < a < 0.5$ is enough for the LCF to lie between $\infty > LCF > -\infty$. In other words, the LCF can take any value with $a < 0.5$.

For log-volatiltiy, when $a < -1$, the LCF has a positive value and reflects large losses for underprediction whilst when $a > -1$, the LCF has a negative value and reflects large losses for overprediction. Note that when $a = -1$, the LCF is zero and the optimal log-volatility forecasts are the same as GARCH (1,1) log-conditional volatility. However, we saw that a negative

a reflects a large loss for underprediction. Apparently, when $a = -1$, the loss function is asymmetric and has large weights for underprediction. In this case equation (12) shows that $E[\ell n(y_t^2)|\Omega_{t-1}] < \ell n(h_t)$.

Remark 4.1. For the long memory structure of volatility processes (see Granger, Ding, and Spear, 1997, and Andersen, Bollerslev, Diebold, and Labys, 1999, for example), the above analysis allows us to use LINEX forecasts for long memory conditional volatility models such as FIGARCH models.

If we are concerned with the return process which does not need logarithmic transformation, then $a < 0$ reflects a large loss for underpredictions and $a > 0$ reflects a large loss for overpredictions as explained in section 3. This is shown in the following remark.

Remark 4.2. For the return process, y_t, the conditional mgf is

$$\begin{aligned} m_{y_t}(-a)_{\Omega_{t-1}} &= E[e^{-ay_t}|\Omega_{t-1}] = E[e^{-az_t h_t^{1/2}}|h_t^{1/2}] \\ &= e^{\frac{a^2 h_t}{2}}. \end{aligned} \tag{18}$$

Therefore, the one step ahead conditional forecast is

$$\begin{aligned} E[y_t|\Omega_{t-1}] &= -\frac{\ell n(m_{y_t}(-a))}{a}|_{\Omega_{t-1}} \\ &= -\frac{1}{2}\frac{a^2 h_t}{a} \\ &= -\frac{1}{2} a h_t. \end{aligned} \tag{19}$$

Note that the results in equation (19) agrees with the CD result; see section 3, Christoffersen and Diebold (1997).

On the other hand, we do not have a closed form solution for unconditional one-step-ahead forecasts, since the unconditional mgf of $\ell n h_t$ is typically unknown. In addition, we also do not suggest the unconditional LINEX forecast of y_t for the same reason.

4.4. Exponential GARCH

The Exponential GARCH model introduced by Nelson (1991) is given by (20) below. It is interesting to note that in the following definition (20), $h_1, ..., h_{t+1}$ belongs to the information set. We define y_t by,

$$\begin{aligned} y_t &= \sigma_t z_t \\ \sigma_t &= e^{h_t/2} \\ h_t &= \alpha_t + \sum_{j=1}^{\infty} \beta_j (\theta z_{t-j} + \gamma(\mid z_{t-j} \mid - E \mid z_{t-j} \mid)) \end{aligned} \tag{20}$$

Note that

$$\begin{aligned} \ell n y_t^2 &= \ell n \sigma_t^2 + \ell n z_t^2 \\ &= h_t + \ell n z_t^2 \end{aligned} \tag{21}$$

setting $\alpha_t = 0$ without loss of generality, we have

$$\ell n y_t^2 = h_t + \ell n \chi^2_{(1)}$$

with

$$h_t = \sum_{j=1}^{\infty} \beta_j (\theta z_{t-j} + \gamma(\mid z_{t-j} \mid - E \mid z_{t-j} \mid))$$

Thus

$$E[\exp(-a\ell n y_t^2)] = E[e^{-ah_t}]E[e^{-a\ell n \chi^2_{(1)}}] \tag{22}$$

since h_t depends only on lagged z_t's.

Theorem 4.2. *Optimal One-step-ahead Conditional Forecast of $\ell n y_t^2$ in E-GARCH Models. The LINEX optimal conditional forecast of $\ell n(y_t^2)$ is*

$$E[\ell n(y_t^2)|\Omega_{t-1}] = h_t + \ell n(2) - \frac{1}{a}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}] \tag{23}$$

where $h_t = \sum_{j=1}^{\infty} \beta_j (\theta z_{t-j} + \gamma(\mid z_{t-j} \mid - E \mid z_{t-j} \mid))$ *and* $a < \frac{1}{2}$.

Proof. Using the same method as in the GARCH(p,q) model, the moment generating function of $\ell n y_t^2$ conditioned on h_t is

$$\begin{aligned} m_{\ell n(y_t^2)}(-a)|_{\Omega_{t-1}} &= E[\exp(-a\ell n y_t^2)] \\ &= E[e^{-a\ell n \chi^2_{(1)}}]e^{-ah_t} \\ &= 2^{-a}\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}e^{-ah_t} \end{aligned} \tag{24}$$

using Lemma A in the Appendix. Thus the theorem can be obtained with

$$
\begin{aligned}
E[\ell n(y_t^2)|\Omega_{t-1}] &= -\frac{\ell n(m_{\ell n y_t^2}(-a))}{a}|_{\Omega_{t-1}} \qquad (25)\\
&= h_t + \ell n(2) - \frac{1}{a}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}].
\end{aligned}
$$

▮

Remark 4.3. For the fractionally integrated exponential GARCH (FIEGARCH) process introduced by Bollerslev and Mikkelsen (1996), we can also easily show that the LINEX optimal conditional forecast of $\ell n(y_t^2)$ is the same as (23) with the definition of the conditional volatility of the fractionally integrated exponential GARCH (FIEGARCH) process.

THEOREM 4.3. *Optimal One-step-ahead Unconditional Forecast of $\ell n y_t^2$ in E-GARCH Models. The optimal LINEX unconditional forecast for $\ell n y_{t+1}^2$ is given by*

$$
\begin{aligned}
E[\ell n y_t^2] &= \ell n2 - \frac{1}{a}\ell n[\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}] - \gamma\sqrt{\frac{2}{\pi}}\sum_{j=1}^{\infty}\beta_j\\
&\quad -\frac{1}{a}\sum_{j=1}^{\infty}\ell n[e^{a^2\beta_j^2(\theta+\gamma)^2/2}\Phi(-a\beta_j(\theta+\gamma))\\
&\quad +e^{a^2\beta_j^2(\theta-\gamma)^2/2}\Phi(-a\beta_j(\gamma-\theta))],
\end{aligned}
$$

where $\Phi(.)$ is the cumulative density function of the standard normal distribution,

Proof. See Appendix. ▮

4.5. Stochastic Volatility Model

In this section, we investigate LINEX optimal forecasts of the stochastic volatility model (SVM). This model is discussed in Taylor (1986) and Harvey and Shephard (1993, 1996). The SVM is given by

$$
\begin{aligned}
y_t &= z_t e^{h_t/2} \qquad (26)\\
h_t &= \lambda + \alpha h_{t-1} + \nu_t, \qquad \nu_t \sim iid\ N(0,\sigma^2)
\end{aligned}
$$

where $z_t \sim iid\ N(0,1)$ and it is assumed that z_t and ν_t are independent. Note that log-volatility can be represented as $\ell n y_t^2 = h_t + \ell n z_t^2$. Although not immediately obvious, according to Definition 1, h_1, ..., h_t, h_{t+1} are not in the information set, intuitively because there are two sources of noise.

THEOREM 4.4. *Optimal One-step-ahead Conditional Forecast of $\ell n y_t^2$ in SVM. The optimal LINEX forecast of $\ell n y_t^2$ conditional on h_t is*

$$E[\ell n y_t^2|\Omega_{t-1}] = E(h_t|\Omega_{t-1}) + \ell n 2 - \frac{1}{2}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}]$$

where h_t is defined in (26) and $a < \frac{1}{2}$.

Proof. The moment generating function of $\ell n y_t^2$ is

$$\begin{aligned} E[\exp(-a\ell n y_t^2)] &= E[\exp(-ah_t)\exp(-a\ell n z_t^2)] \qquad (27) \\ &= E[e^{-ah_t}]2^{-a}\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}. \end{aligned}$$

The optimal LINEX forecast of $\ell n y_t^2$ conditional on h_t is

$$\begin{aligned} E[\ell n y_t^2|\Omega_{t-1}] &= -\frac{\ell n(m_{\ell n y_t^2}(-a))}{a}|_{\Omega_{t-1}} \qquad (28) \\ &= E(h_t|\Omega_{t-1}) + \ell n 2 - \frac{1}{2}\ell n[\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}]. \end{aligned}$$

∎

In general $E(h_t|\Omega_{t-1})$ will depend upon lagged y_t values, but a simple expression for this term does not appear to be available in the SVM. We next look at the unconditional LINEX forecast of lny_t^2.

THEOREM 4.5. *Optimal One-step-ahead Unconditional Forecast of $\ell n y_t^2$ in SVM. The optimal LINEX prediction of $\ell n y_t^2$ is given by*

$$E[\ell n y_t^2] = \frac{\lambda}{1-\alpha} - \frac{\sigma^2 a}{2(1-\alpha^2)} + \ell n 2 - \frac{1}{a}\ell n\left(\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}\right).$$

Proof. The unconditional moment generating function of h_t is

$$E[e^{-ah_t}] = \exp(\frac{-a\lambda}{1-\alpha})\exp(\frac{a^2\sigma^2}{2(1-\alpha^2)}). \qquad (29)$$

∎

Therefore, the optimal LINEX prediction of $\ell n y_t^2$ is given by

$$\begin{aligned} E[\ell n y_t^2] &= -\frac{\ell n(m_{\ell n y^2}(-a))}{a} \qquad (30) \\ &= -\frac{1}{a}\ell n\left\{ e^{\frac{-a\lambda}{(1-\alpha)}} e^{\frac{\sigma^2 a^2}{2(1-\alpha^2)}} 2^{-a} \frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})} \right\} \\ &= -\frac{1}{a}\left\{ \frac{-a\lambda}{1-\alpha} + \frac{\sigma^2 a^2}{2(1-\alpha^2)} - a\ell n2 + \ell n\left(\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}\right) \right\} \\ &= \frac{\lambda}{1-\alpha} - \frac{\sigma^2 a}{2(1-\alpha^2)} + \ell n2 - \frac{1}{a}\ell n\left(\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}\right). \end{aligned}$$

4.6. Knight-Satchell Modified GARCH(p,q)

This model is presented in Knight and Satchell (1998b). Essentially, it writes h_t as linear in lagged h_t and lagged z_t^2, thereby eliminating the non-linearities implicit in a standard GARCH model. The Knight-Satchell (KS) Modified GARCH(p,q) can be represented as

$$\begin{aligned} y_t &= z_t h_t^{1/2} \qquad (31) \\ h_t &= \alpha + \sum_{i=1}^{p} \beta_i^2 h_{t-i} + \sum_{j=1}^{q} \gamma_j z_{t-j}^2 \end{aligned}$$

where $z_t \sim iid\ N(0,1)$. See Knight and Satchell (1998) for further discussion on this model. In this model the information set, Ω_{t-1}, contains $h_1, h_2, , , , h_t$.

THEOREM 4.6. *Optimal One-step-ahead Conditional Forecast of* $\ln y_t^2$ *in the KS Modified GARCH(p,q). The LINEX optimal one-step-ahead forecast is*

$$E[\ln(y_t^2)|\Omega_{t-1}] = \ln h_t + \ln(2) - \frac{1}{a}\ln\left(\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}\right),$$

where h_t *is defined in (31) and* $a < \frac{1}{2}$.

Proof. Using Lemma A in the Appendix, the mgf of lny_t^2 conditioning on the information set Ω_{t-1} is

$$\begin{aligned} m_{\ln(y_t^2)}(-a)|\Omega_{t-1} &= E[e^{-a(\ln y_t^2)}|\Omega_{t-1}] \\ &= E[e^{-a\ln\chi^2_{(1)}}]e^{-a\ln h_t} \\ &= 2^{-a}\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}e^{-a\ln h_t} \end{aligned} \tag{32}$$

where h_t is defined in equation (31). Therefore, the LINEX optimal one-step-ahead forecast is

$$\begin{aligned} E[\ln(y_t^2)|\Omega_{t-1}] &= -\frac{\ln(m_{\ln(y_t^2)}(-a))}{a}|_{\Omega_{t-1}} \\ &= \ln h_t + \ln(2) - \frac{1}{a}\ln\left(\frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})}\right), \end{aligned} \tag{33}$$

which is exactly the same as that for the ARCH family models in Theorem 1 except for the different conditional volatility process h_t. ∎

For the KS modified GARCH(p,q) model, the optimal one-step-ahead conditional forecast of y_t is also exactly the same as that of GARCH model in (15) except the definition of h_t. This is because, the process in (31) is equivalent to GARCH(p,q) process in (15). However, for the KS model, we can calculate the optimal LINEX one-step-ahead unconditional predictor of y_t. To see this, let us consider a simple case of $p = 1$ and $q = 1$. The mgf of the conditional volatility of the modified GARCH(1,1) model can be shown to be

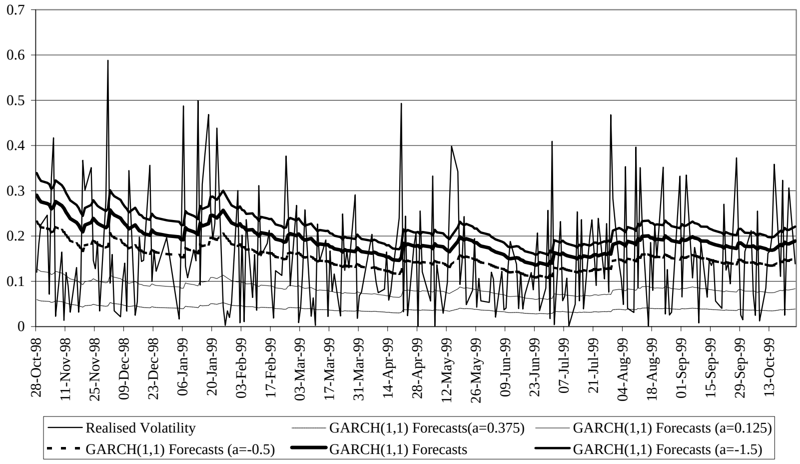
$$m_y(-a) = \exp\left(\frac{a^2\alpha}{2(1-\beta)}\right)\cdot\prod_{j=0}^{\infty}(1-a^2\gamma\beta^j)^{-1/2} \tag{34}$$

The optimal LINEX one-step-ahead unconditional predictor of y_t is given by

$$E[y_t] = -\frac{a\alpha}{2(1-\beta)} + \frac{1}{2a}\sum_{j=0}^{\infty}\ell n(1-a^2\gamma\beta^j).$$

See Knight and Satchell (1998) for proof. The optimal LINEX forecast for the more complicated KS GARCH(p,q) models where $p > 1$ and $q > 1$ will be obtained by an application of the above method.

FIG. 4. Realised Volatility and One-step-ahead Forecasted Volatility from GARCH(1,1) Model



5. AN EMPIRICAL EXAMPLE

The motivation for our empirical work is to see how our LINEX forecasts behave with different LINEX parameters. We use GARCH(1,1) model to investigate how the GARCH(1,1) optimal volatility forecasts under the assumption of the LINEX loss function and normality are different from the GARCH(1,1) forecasts.

We used the same data as used in section 2; the FTSE100 and the S&P500 index daily log-returns fron 21 January 1992 to 20 January 20 2000. Again GARCH(1,1) was used because of its great popularity. Other procedures such as the calculation of the return volatility and the procedure of out-of-sample forecast tests is the same as that in section 2.

Figure 4 plots the one-step-ahead LINEX optimal forecasts for GARCH(1,1) model with various LINEX parameters; see equation (15). The figure also plots realised volatility and one-step-ahead GARCH(1,1) forecasts, which are the same as figure 1a. Figure 4 shows that a value of $a < -1$ reflects small losses for underprediction and large losses for overprediction and thus results in higher forecasts. On the other hand, when $a > -1$, the optimal forecasts are less than those of the GARCH(1,1) model.

As explained in the previous section, when $a = -1$, GARCH(1,1) forecasts are the same as the LINEX optimal GARCH(1,1) forecasts. As expected, when $a = 0.375$, the LCF become -3.1657 and the LINEX optimal GARCH(1,1) forecasts are very low.

6. CONCLUSIONS

This study derives the one-step-ahead optimal LINEX forecasts for various nonlinear functions of returns associated with volatility. In addition, the empirical example in section 5 compares the GARCH(1,1) volatility forecasts with the LINEX forecasts of the GARCH(1,1). Our findings suggest that under the assumption of normality, we can easily obtain the LINEX forecasts of a range of volatility models with an additional adjustment component.

Further research needs to look at multiperiod LINEX conditional and unconditional forecasts. Other work of interest would be to extend our empirical results to all models. As yet we have no general results as to which models would be especially favoured by LINEX relative to mean squared estimates for an appropriate family of loss functions.

APPENDIX A

We first prove the moment generating function of log of chi-square which is key to the optimal volatility forecasts.

LEMMA A.1. *The moment generating function of* $\ell n\chi^2_{(1)}$ *is*

$$m_{\ell n\chi^2_{(1)}}(\theta) = 2^{\theta}\frac{\Gamma(\theta+\frac{1}{2})}{\Gamma(\frac{1}{2})}. \tag{A1}$$

Proof. The moment generating function of $\ell n\chi^2_{(1)}$ is

$$\begin{aligned} m_{\ell n\chi^2_{(1)}}(\theta) &= E[e^{\theta\ell n\chi^2_{(1)}}] \\ &= E[(\chi^2_{(1)})^{\theta}] \\ &= \int_0^{\infty} x^{\theta}\frac{1}{\Gamma(\frac{1}{2})2^{1/2}}x^{1/2-1}e^{-x/2}dx \\ &= \frac{1}{\Gamma(\frac{1}{2})2^{1/2}}\int_0^{\infty} x^{\theta+1/2-1}e^{-x/2}dx. \end{aligned}$$

Transforming from x to $w = x/2$, which implies $dx = 2dw$, we see that

$$\begin{aligned} m_{\ell n\chi^2_{(1)}}(\theta) &= \frac{2^\theta}{\Gamma(\frac{1}{2})}\int_0^\infty w^{\theta-1/2}e^{-w}dw \\ &= \frac{2^\theta}{\Gamma(\frac{1}{2})}\Gamma(\theta+\frac{1}{2}) \\ &= 2^\theta\frac{\Gamma(\theta+\frac{1}{2})}{\Gamma(\frac{1}{2})} \end{aligned}$$

where $\Gamma(.)$ is the gamma function and the LINEX parameter θ is restricted to be larger than than $-\frac{1}{2}$ since $\theta+\frac{1}{2}>0$. ▮

Proof (Proof of Theorem 3). Since $z_t \sim iid\ N(0,1)$ we have

$$E[e^{-ah_t}] = \prod_{j=1}^{\infty} E[\exp(-a\theta\beta_j z_{t-j} - a\beta_j\gamma\ |z_{t-j}|)]\exp(a\beta_j\gamma E\ |z_{t-j}|).$$

Examining $E[\exp(a_1 z_t + b_1|z_t|)]$, with $z_t \sim iid\ N(0,1)$, we have

$$\begin{aligned} E[\exp(a_1 z_t + b_1|z_t|)] &= \int_{-\infty}^{\infty} e^{a_1 z + b_1|z|}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz \\ &= \int_{-\infty}^{0} e^{a_1 z - b_1 z}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz + \int_0^\infty e^{a_1 z + b_1 z}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz \end{aligned}$$

Consider

$$\begin{aligned} &\int_0^\infty e^{a_1 z + b_1 z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\ &= \exp((a_1+b_1)^2/2)\int_0^\infty \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(z-(a_1+b_1))^2)dz \end{aligned}$$

If we put $q = z - (a_1+b_1)$, then $dz = dq$, so that we have

$$\begin{aligned} \int_0^\infty e^{a_1 z + b_1 z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz &= \exp((a_1+b_1)^2/2)\int_{-(a_1+b_1)}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{q^2}{2}}dq \\ &= \exp((a_1+b_1)^2/2)\Phi(a_1+b_1). \end{aligned}$$

where $\Phi(.)$ is the cumulative density function of the standard normal distribution. Next

$$\begin{aligned} \int_{-\infty}^{0} e^{(a_1-b_1)z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz &= \int_0^\infty e^{(b_1-a_1)w}\frac{1}{\sqrt{2\pi}}e^{-w^2/2}dw \\ &= \exp((b_1-a_1)^2/2)\Phi(b_1-a_1) \end{aligned}$$

by putting $w = -z$, then $dz = -dw$. Finally, for $E\ |z|$ when $z \sim N(0,1)$, we require

$$
\begin{aligned}
E(|z|) &= \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= -\int_{-\infty}^{0} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz + \int_{0}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= 2\int_{0}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-w} dw = \sqrt{\frac{2}{\pi}}
\end{aligned}
$$

Thus

$$
\begin{aligned}
E[e^{-ah_t}] &= m_h(-a) \\
&= \prod_{j=1}^{\infty} [\exp(a^2\beta_j^2(\theta+\gamma)^2/2)\Phi(-a\beta_j(\theta+\gamma)) \\
&\quad + \exp(a^2\beta_j^2(\theta-\gamma)^2/2)\Phi(-a\beta_j(\gamma-\theta))] \exp\left(a\beta_j\gamma\sqrt{\frac{2}{\pi}}\right).
\end{aligned}
$$

Therefore, using equations (A1) and (22), we have

$$
E[\exp(-a\ell n y_t^2)] = 2^{-a} \frac{\Gamma(-a+\frac{1}{2})}{\Gamma(\frac{1}{2})} m_h(-a)
$$

Therefore the optimal LINEX unconditional forecast for $\ell n y_{t+1}^2$ is given by

$$\begin{aligned}
E[\ell n y_t^2] &= -\frac{\ell n(m_{\ell n y^2}(-a))}{a} \\
&= -\ell n\{2^{-a}\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}\cdot m_h(-a)\}\big/ a \\
&= -\frac{1}{a}\{-a\ell n2+\ell n\Gamma(\frac{1}{2}-a)-\ell n\Gamma(\frac{1}{2}) \\
&\quad +\sum_{j=1}^{\infty}[a\beta_j\gamma\sqrt{\frac{2}{\pi}}+\ell n\{\exp(a^2\beta_j^2(\theta+\gamma)^2\big/2)\Phi(-a\beta_j(\theta+\gamma)) \\
&\quad +\exp(a^2\beta_j^2(\theta-\gamma)^2\big/2)\Phi(-a\beta_j(\gamma-\theta))\}] \\
&= \ell n2-\frac{1}{a}\ell n[\frac{\Gamma(\frac{1}{2}-a)}{\Gamma(\frac{1}{2})}]-\gamma\sqrt{\frac{2}{\pi}}\sum_{j=1}^{\infty}\beta_j \\
&\quad -\frac{1}{a}\sum_{j=1}^{\infty}\ell n[e^{a^2\beta_j^2(\theta+\gamma)^2/2}\Phi(-a\beta_j(\theta+\gamma)) \\
&\quad +e^{a^2\beta_j^2(\theta-\gamma)^2/2}\Phi(-a\beta_j(\gamma-\theta))].
\end{aligned}$$

∎

REFERENCES

Andersen, T. G. and T. Bollerslev, 1997, Long-run volatility dynamics and high frequency returns. *Journal of Finance* **52**, 975-1005.

Andersen, T. G., T. Bollerslev, F. Diebold, and P. Labys, 1999, The distribution of exchange rate volatility. Working Paper 99-08, Financial Instutions Center, The Wharton School, University of Pennsylvania.

Baillie, R. T. and T. Bollerslev, 1992, Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics* **52**, 91-113.

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

Bell, O. E., 1995, Risk, return, and utility. *Management Science* **41**, 23-30.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

Bollerslev, T. and H. O. Mikkelsen, 1996, Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* **73**, 151-184.

Christodoulakis, G. A. and S. E. Satchell, 1998, Hashing GARCH: A reassessment of volatility forecasting performance. In *Forecasting Volatility in the Financial Markets.* Edited by J. Knight, and S. E. Satchell. Butterworth-Heinemann.

Christoffersen, P. F. and F. X. Diebold, 1996, Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics* **11**, 561-571.

Christoffersen, P. F. and F. X. Diebold, 1997, Optimal prediction under asymmetric loss. *Econometric Theory* **13**, 808-817.

Christoffersen, P. F. and F. X. Diebold, 1998, How relevant is volatility forecasting for financial risk management. Working paper 6844, National Bureau of Economic Research.

Day, E. T. and C. M. Lewis, 1992, Stock market volatility and the information content of stock index options. *Journal of Econometrics* **52**, 267-287.

Ding, Z., C. W. Granger, and R. F. Engle, 1993, A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**, 83-106.

Engle, R. F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987-1007.

Engle, R. F., C. Hong, A. Kane, and J. Noh, 1993, Arbitrage valuation of variance forecasts with simulated options. *Advances in Futures and Options Research* **6**, 393-415.

Granger, C., Z. Ding, and S. Spear, 1999, Stylized facts on the temporal and distributional properties of daily data from speculative markets. Working Paper, Department of Economics, Univeristy of California San Diego.

Harvey, A. C. and N. Shephard, 1993, Estimation and testing of stochastic variance models. Econometrics discussion paper EM/93/268, London School of Economics.

Harvey, A. C. and N. Shephard, 1996, Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* **14**, 429-34.

Harvey, C. R. and R. E. Whaley, 1992, Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics* **31**, 43-73.

Hwang, S. and S. E. Satchell, 1998, Implied volatility forecasting: A comparison of different procedures including fractionally integrated models with applications to UK equity options. In *Forecasting Volatility in the Financial Markets.* Edited by J. Knight and S. E. Satchell. Butterworth-Heinemann.

Knight, J. and S. E. Satchell, 1998a, *Forecasting Volatility in the Financial Markets.* Butterworth-Heinemann.

Knight, J. and S. E. Satchell, 1998b, GARCH processes - some exact results, some difficulties and a suggested remedy. In *Forecasting Volatility in the Financial Markets.* Edited by J. Knight and S. E. Satchell. Butterworth-Heinemann.

Lamoureux, C. G. and W. D. Lastrapes, 1993, Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Review of Financial Studies* **6**, 293-326.

Nelson, D. B., 1991, Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347-370.

Nelson, D. B., 1996, Asymptotically optimal smoothing with ARCH models. *Econometrica* **64**, 561-573.

Nelson, D. B. and D. P. Foster, 1994, Asymptotic filtering theory for univariate ARCH models. *Econometrica* **62**, 1-41.

Noh, J., R. F. Engle, and A. Kane, 1994, Forecasting volatility and option prices of the S&P 500 index. *Journal of Derivatives* 17-30.

Taylor, S. J., 1986, *Modeling Financial Time Series.* Chichester: John Wiley.

Varian, Hal R., 1975, A Bayesian approach to real estate assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage.* Edited by S. E. Fienberg, and A. Zellner. Amsterdam: North-Holland.

Zellner, A., 1986, Bayesian estimation and prediction using asymmetric loss functions. *Journal of American Statistical Association* **81**, 446-451.