FORECASTING INTEREST RATES: A COMPARATIVE ASSESSMENT OF SOME SECOND GENERATION NON-LINEAR MODELS

Dilip M. Nachane
Indira Gandhi Institute of Development Research (IGIDR)
General Vaidya Marg, Goregaon (East)
Mumbai 400 065
India
Telephone :91-22-2840 2754
Fax : 91-22-2840 2752
E-Mail : nachane@igidr.ac.in
&
Jose G. Clavel
Facultad de Economía
30.100 Campus de Espinardo
Murcia
Spain
Telephone: +34-968-363757
Fax: +34-968-367905
E-mail: jjgarvel@um.es

Brief Descriptive Title

FORECASTING INTEREST RATES

ABSRACT: Modelling and forecasting of interest rates has traditionally proceeded in the framework of linear stationary models such as ARMA and VAR, but only with moderate success. We examine here four models which account for several specific features of real world asset prices such as non-stationarity and non-linearity. Our four candidate models are based respectively on wavelet analysis, mixed spectrum analysis, non-linear ARMA models with Fourier coefficients, and the Kalman filter. These models are applied to weekly data on interest rates in India, and their forecasting performance is evaluated vis-à-vis three GARCH models (GARCH (1,1), GARCH-M (1,1) and EGARCH (1,1)) as well as the random walk model. The Kalman filter model emerges at the top, with wavelet and mixed spectrum models also showing considerable promise.

Keywords: Interest rates; wavelets; mixed spectra; non-linear ARMA; Kalman filter; GARCH; Forecast encompassing.

Biographical Sketches

Dilip Nachane: Dilip Nachane is currently Senior Professor at the IGIDR (Indira Gandhi Institute of Development Research), Mumbai, India. His previous appointments include Director, Department of Economics, University of Mumbai and Hon. Director, Indian Council of Social Science Research (Western Region). He is also currently a member of the newly formed Monetary Policy Advisory Committee of the Reserve Bank of India. He has been Visiting Professor at the University of British Columbia, Canada, Jean Monnet Fellow at the EUI (European University Institute), Florence, Italy and Hallsworth Fellow at the University of Manchester, UK. His research interests include time series analysis, monetary economics and finance. He has several publications in international academic journals.

Jose Clavel : Jose Clavel is currently Professor at the University of Murcia, Spain. He did his Ph.D in 1997 on "Correspondence Analysis, Classification and Regression Trees" . He has been Visiting Professor at the Ontario Institute for Studies in Education, Toronto, Canada, the University Pompeu Fabra, Barcelona, Spain and the Indira Gandhi Institute of Development Research, Mumbai, India. His research is mainly in the field of categorical data analysis, time-series analysis and financial econometrics. He has published widely in reputed academic journals.

1. <u>INTRODUCTION</u>

Throughout its history, the world of finance has typified the well-known adage "Nothing is permanent except change". There have, however, been periods when change has been gradual and evolutionary, and others, when fundamental and cataclysmic "shape-shifting" (a term due to Kane (1984)) has occurred within a relatively short span of time. The last three decades have been witness to just such an episodic "shape-shifting" in financial markets and institutions, spurred on by the mutually reinforcing forces of financial deregulation and financial innovation, riding on the back of rapid strides in information technology.

One important consequence of the above developments has been the heightened uncertainty surrounding the behaviour of key macro-economic variables. The turbulence in interest rates, in particular, is of special significance, as this impinges on a wide range of economic activities (see James & Webber (2000) p.5).

Central banks committed to inflation control, output stabilization and exchange rate management, also need to monitor interest rate movements, in view of the intimate interconnections between these target variables and the interest rate. It should thus be hardly surprising that a great deal of attention has been focused in recent years on developing models and refining tools for interest rate forecasting. The demand for this activity stems from at least three principal sources :

 (i)   Banks, mutual funds and corporations, for managing the interest rate risk attached to their investment portfolios.
 (ii)   Banks and other authorized dealers in hedge funds, for appropriate pricing of their hedging instruments.
 (iii)   Central banks, from the viewpoint of monetary policy design.

India has had a long history of strictly regulated interest rates, and it was only in the early 1990s, with the onset of the financial liberalization program, that interest rates were progressively deregulated. The freeing of interest rates witnessed a surge in forecasting models (univariate as well as multivariate), but the success of these models was strictly limited (see e.g. Dua & Pandit (2002) and Dua *et al* (2003)). Most of these models were based on conventional linear econometric methodologies such as ARIMA, VAR, Bayesian VAR (BVAR) etc. Such models make heavy compromises with real world phenomena such as non-linearity and non-stationarity. It would therefore be interesting to examine whether resorting to more sophisticated models, paying special attention to these and related features, could deliver improved forecasting performance. This is what we attempt to examine in this paper, using recent (post-liberalized) data on interest rates in India. A major conclusion emerging from our paper is that models exploiting some of the newer methodologies (but based on simple univariate frameworks), can exhibit significant improvements over their more conventional counter-parts.

We develop forecasting models involving four *genres* of methodology viz.
(i)   Prediction based on a combination of *artificial neural networks (ANN)* and *wavelet decompositions* (Model I)
(ii) Prediction based on *mixed spectrum* analysis (Model II)
(iii) Prediction based on *Fourier coefficients of non-linear ARMA* models (Model III)
(iv) Prediction based on the *Kalman Filter* (Model IV)

Each of these methods is adapted towards addressing specific, non-standard features of the data. Thus Model I draws on a powerful and flexible new method, which adapts its parameters rapidly to changing patterns in the data. Model III is designed to accommodate a wide range of non-linearities. Model II is unique in some respects. It is not a new method, having been known to engineers for at least three decades or more. Its use in economics has however been rare, if at all. It pays special attention to deterministic cycles, which could be present in the data. In the case of interest rates, such deterministic features could arise from at least three major sources viz. strong seasonalities in money supply, the process of averaging characterizing published data on interest rates, and the strong presence of "technical analysts" in the markets for government securities. So far as Model IV is concerned, it is a highly flexible method, especially geared to account for rapid changes in parameters and structures. Even though extensively used in several forecasting situations in economics, its use in the interest rates context has been infrequent. Hence at the risk of some taxonomic over-simplification, all these models may be referred to as *second generation* non-linear models.

We assess the performance of these models vis-à-vis more traditional (*first generation*) non-linear models, usually used to forecast asset prices and interest rates, such as GARCH (1,1) (Model V), exponential or E-GARCH (1,1) (Model VI), and GARCH-in-mean or GARCH-M (1,1) (Model VII). We also use additionally, the standard benchmark –the Random Walk model (Model VIII).

The plan of our paper is as follows. Since Models I-III are not generally familiar to economists, their essential features are sketched in a non-rigorous fashion in the next three Sections (Sections 2, 3 and 4). Model IV is better known among economists but for the sake of completeness, Section 5 is devoted to a brief discussion of the same, while Section 6 presents an even more terse description of the conditional volatility models (Models V to VII). Section 7 presents the forecasting results and undertakes a comparative assessment of Models I –IV vis-à-vis each other as well as the benchmark Models V-VIII. Finally, Section 8 gathers the main conclusions.

2. WAVELET-BASED NEURAL NETWORKS

As mentioned in the previous Section, conventional time series analysis has always found it difficult to grapple with issues of non-stationarity. Since non-stationarity is pervasive in real-world economic and financial series, economists of late, have evinced a great deal of interest in mathematical, statistical and engineering techniques such as evolutionary spectral analysis (e.g. Nachane & Ray (1993), Nachane (2004)) artificial neural networks

(e.g. Refenes & Azema-Barac (1994), Swanson & White (1995), Gençay & Qi (2001) etc.) and wavelets (Ramsey *et al* (1995), Ramsey & Zhang (1997) etc.). Wavelet analysis shares several features in common with spectral analysis but has the advantage of capturing features in the underlying series that vary across both time and frequency.[i]

Wavelet Definitions : The following (largely heuristic) discussion on wavelets is intended to convey to the reader the main underlying ideas, and is modestly aimed at being "minimally sufficient" for the purpose of describing our forecasting model. The essence of wavelet analysis consists in projecting the time series of interest $\{x_t\}$, t = 0, 1,2….(N-1) onto a *discrete wavelet filter* (or often called the *mother wavelet*) $\{h_l\} = ($ h_0 , h_1 ….. h_{L-1},0…0). The *discrete wavelet filter* is supposed to satisfy three properties:

$$\sum_{l=0}^{L-1} h_l = 0 \tag{1}$$

$$\sum_{l=0}^{L-1} h_l^2 = 1 \tag{2}$$

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0, \quad \text{for all non-zero integers n} \tag{3}$$

and L is a suitably chosen positive integer. Note that we have assumed L < N, and padded the filter with (N-L) zeros at the end, so that $\{h_l\}$ has the same dimension N as $\{x_t\}$.

By virtue of (1), $\{h_l\}$ is a *high-pass filter.* [ii] Associated with $\{h_l\}$ is a so-called *scaling filter* (or *father wavelet)*, which is a *low-pass filter,* recoverable from $\{h_l\}$ via the relationship

$$g_l = (-1)^{l+1} h_{L-1-l}, \quad l = 0,1,.....L-1 \tag{4}$$

Various types of *mother wavelets* have been suggested in the literature, the most frequently used being the Haar (1910) and Daubechies (1992) wavelets[iii].

Discrete Wavelet Transform (DWT) : The next fundamental step in wavelet analysis is introducing the discrete wavelet transform (DWT), which can be done in several alternative ways. The intuitively most appealing procedure is the *pyramid algorithm,* suggested in Mallat (1989) (and fully explained in Percival & Walden (2000)).

For a discrete series $\{x_t\}$ of *dyadic* length N (i.e. N=2^J , where J is a positive integer), the algorithm yields the N-dimensional vector of *wavelet coefficients*

$$w_t = \left(w_t^{(1)}, w_t^{(2)}, ...w_t^{(J)}, v_t^{(J)} \right)^T. \tag{5}$$

where the $(N/2^J)$ vector $\left\{ w_t^{(j)} \right\}$ can be interpreted as the vector of *wavelet coefficients*

associated with the dynamics of our series $\{ x_t \}$ on a scale of length $\lambda^j = 2^{j-1}$, (with increasing scales corresponding to lower frequencies), whereas $\{ v_t^{(J)} \}$ represents the averages on a scale of length 2^J (see Gençay et al (2002)).

<u>Multi-Resolution Analysis (MRA) :</u> MRA represents a convenient way of decomposing a given series $\{ x_t \}$ into changes attributable at different scales. Let $\mathbf{q_1}$ represent the filter coefficients $\{ h_l \}$, written in reverse order i.e.

$\mathbf{q_1} = (h_N, h_{N-1}, \ldots h_1, h_0)^T.$

Let $\mathbf{q_j}$ denote the zero-padded scale j wavelet filter coefficients, obtained by j convolutions of $\mathbf{q_1}$ with itself,[iv] and let Ψ_j represent the (N/2^j x N) matrix of "circularly shifted" coefficients of $\mathbf{q_j}$ (by a factor of 2^j). We now write down the (NxN) matrix Ψ as

$$\begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \ldots \\ \ldots \\ \Psi_J \\ \vartheta_J \end{bmatrix} = \Psi \quad \text{where } \vartheta_J \text{ is a (1xN) vector with each term equal to } (1/\sqrt{N})$$

The MRA proceeds by defining the *j-th level wavelet detail* $d_{j,t}$ by

$$d_{j,t} = \Psi_j^T w_t^{(j)}, \quad j=1,2\ldots.J \tag{6}$$

where $w_t^{(j)}$ are the wavelet coefficients at the j-th scale defined in (5). We also define the *wavelet smooth* $s_{J,t}$ as

$$s_{J,t} = \vartheta_j^T v_t^{(J)} \tag{7}$$

 An MRA may now be defined by the relationship

$$x_t = \sum_{j=1}^{J} d_{j,t} + s_{J,t} \tag{8}$$

Thus each observation in the series is additively decomposed into the J *wavelet details* and the *wavelet smooth*. As we shall see shortly, it is the MRA decomposition which enables us to use wavelets for prediction.

Maximum Overlap Discrete Wavelet Transform (MODWT) : The discrete wavelet transform (DWT) defined by the wavelet coefficients (5) is often referred to as the *decimated* transform as it arises from a successive *down-sampling process* (as described, for example, in Percival & Walden (2000), p 99-104). For prediction purposes, what is needed is an *undecimated* DWT. Such an undecimated DWT is provided by the so-called *maximum overlap discrete wavelet transform* or MODWT, described in Coifman & Donoho (1995), Percival & Walden (2000) etc. The MODWT coefficients can be obtained via a pyramid algorithm, as in the case of the decimated DWT, except that no *down-sampling* is involved (so that the wavelet coefficients at each level j comprise N elements).

The MODWT possesses several advantages over the decimated DWT.
  (i)   It does not require the series length N to be dyadic. As a matter of fact, N can be arbitrary.
  (ii)   The MODWT coefficients $w_t^{(j)}$ at scale j of the signal $(x_1, x_2, ...x_m)$, m < N, are strictly the same as the first m coefficients at scale j of the signal $(x_1, x_2, ...x_N)$.
  (iii) In contrast to the DWT, the MODWT details and smooth are associated with *zero-phase filters*, thus making it straightforward to match features in the MRA with those in the original series.

Prediction by Wavelet-Based Neural Networks : The issue of prediction based on wavelets is largely an unexplored field, though Bjorn (1995), Starck *et al* (1998) and others have offered several new and promising prospects in this regard. Given our series of interest $\{x_t\}$, t = 0, 1,2….(N-1) , the 1-period ahead forecast made at t = N-1, viz. x_N^f , is based not on a vector of past observations (as in conventional AR models), but on the MRA (8) , using an un-decimated discrete wavelet transform such as the MODWT. To avoid the problem of over-parametrization, a sparse selection of the *detail* and *smooth* coefficients is required. Starck *et al* (1998) indicate a parsimonious way for selecting the coefficients[v]. In succinct terms, their method predicts x_N based on the *wavelet detail* coefficients $w_{j, N-1-2^j(k-1)}$ at scale j (j=1,2…..J where J = [log$_2$ (N)] and […] denotes "integral part of "), and the *smooth coefficients* $s_{J, N-1-2^J(m-1)}$. Here k is a positive integer, with k=1, 2….A$_j$. Thus A$_j$ is the number of wavelet details corresponding to level j, and has to be determined by the analyst. Similarly let B$_m$ denote the number of smooth coefficients to be included i.e m=1, 2….B$_m$. Since there does not seem to be any formal way available to determine A$_1$, A$_2$,….. A$_J$, B$_m$ from the data, it is usual to set all of them equal to a small number p. The experimentation by Starck *et al* (1998) seems to indicate that values of p upto 4 should be adequate. We have in our analysis used p=4 throughout, and this seemed to be reasonably adequate.

The next step in the analysis is to set up the following *artificial neural network (ANN),* which is a *feedforward network*, with one *hidden layer* and 1 *perceptron.*[vi]

$$x_N^f = \left[\Gamma_1 \left\{ \sum_{j=1}^{J} \sum_{k=1}^{A_j} \hat{a}_{j,k} \, w_{j,N-1-2^j(k-1)} + \sum_{m=1}^{B_m} \hat{b}_{J+1,m} s_{J,N-1-2^J(m-1)} \right\} \right] \qquad (9)$$

where Γ_1 is a suitable *activation function.*

3. MIXED SPECTRUM METHOD

Mixed Spectrum: Many series occurring in nature exhibit strong periodicities. The search for "hidden periodicities" is a problem with a long history, recurring time and again in several subjects such as seismology, astronomy, oceanography, acoustics and medicine. The modern treatment of this problem may be said to begin with the seminal contributions of, among others, Bartlett (1954) and Grenander & Rosenblatt (1957).

The problem has been termed in the literature the *harmonic regression* or *fixed frequency effects model.* We assume that the   series under consideration $\{x_t\}$, t = 0, 1,2….(N-1), is expressible as

$$x_t \;=\; \mu \;+\; \sum_{i=1}^{r} (A_i \cos \omega_i t + B_i \sin \omega_i t) + u_t$$
$$(t = 0,1 .... N - 1) \qquad (10)$$

where u_t is a stationary process (capable of ARMA representation) and the A_i, B_i, ω_i, μ and r (the number of *harmonics*) are unknown parameters to be estimated from the data.

It is well known that the spectrum of a stationary ARMA process is continuous, whereas that of a sinusoid consists of a sharp peak at the frequency of the sinusoid. The spectrum of a sum of k sinusoids would thus be a step function containing k jumps. This leads Priestley (1964, 1981) to the notion of a *mixed spectrum*. For a process such as x_t described in (10), the spectrum F(ω) could be decomposed as

$$F(\omega) = F_1(\omega) + F_2(\omega) \qquad (11)$$

where $F_1(\omega)$ is a discrete spectrum (corresponding to the trigonometric sum) and $F_2(\omega)$ is the continuous spectrum corresponding to the ARMA process u_t. F(ω) is then called a *mixed spectrum*.

Economic applications of such harmonic model concepts (to the best of our knowledge) have not been forthcoming, and this is surprising in view of the well-known periodic features exhibited by many economic series. We essay to remedy this deficiency in the current paper.

Priestley's P(λ) Test: Several approaches to the analysis of the model (10) have been suggested in the literature and here we develop the approach suggested by Priestley (1981), Chapter 8, as this appears to be more rigorously oriented to the underlying mathematical and statistical properties of the model than most of the signal-processing methods suggested in the engineering literature.

Priestley's approach commences by developing a test (the so-called P(λ) test) for estimating r (the number of harmonics), and then goes on to estimate the other parameters of the model.

To motivate the P(λ) test we reformulate (10) as

$$x_t = \mu + \sum_{i=1}^{r} D_i \cos(\lambda_i t + \phi_i) + u_t \tag{12}$$

where now μ, D_i, λ_i, and r are unknown parameters, the ϕ_i, are independent and rectangularly distributed on $(-\pi,\pi)$ and u_t is a stationary linear process with a continuous spectrum. It is important to note that we assume that $\{x_t\}$, has no unit roots (or the unit roots have been filtered out via successive differencing) and further that μ is independent of time.

The first step is to test the null hypothesis

$$H_0 : D_i = 0, \quad i = 1, 2 \ldots r \tag{13}$$

i.e. that harmonic terms are absent from (13).

Non-rejection of the null implies that $\{x_t\}$ is a stationary ARMA process with a purely continuous spectrum.

The P(λ) test rests on a simple intuition, *viz.* that under H_0, the correlogram of x_t will eventually decay to zero. On the other hand, if one or more D_i are non-zero, then the correlogram will exhibit sinusoidal behaviour in its tail. The great advantage of this test is that in the event of the rejection of H_0, it also suggests estimates of D_i, λ_i, etc

The analytics of the method is described in Priestley (1981, Chapter 8). Let $\hat{f}_m$ and $\hat{f}_n$ denote two "window" estimates of the spectrum of x_t, obtained using suitable truncation points m and n respectively where n>2m (m,n, < N). We next put

$$P(\lambda)=\hat{f}_n(\lambda)-\hat{f}_m(\lambda) \tag{14}$$

at the Fourier frequencies $\lambda = (2\pi j/N)$; $j = 0,1,\ldots[N/2]$

If the D_i's are not all zero, $P(\lambda)$ will have several well-defined peaks say $\omega_1 < \omega_2 < \ldots \omega_k$ These peaks are tested for significance (in the order of their occurrence), until a significant peak is found. If none of the peaks are significant, we conclude in favour of H_0. The procedure may be illustrated as follows.

Suppose we are testing the first peak at $\omega_1 = (2\pi p/N)$, $p \neq 0$, (N/2)

and let $\hat{R}_x(s)$ denote the autocorrelation of $\{x_t\}$ at lag s.
A certain test statistic J_q is now defined (see Priestley (1981), p. 631)
Let α be the chosen level of significance ;
(i) if

$$\underset{q}{max}(J_q) \leq \alpha_0, (\alpha_0 = 100\ \alpha\%\ ordinate\ of\ N(0,1))$$

then the first peak at ω_1 is deemed insignificant and we pass on to the second peak of $P(\lambda)$ at ω_2 and so on.

(ii) if
$$\underset{q}{\max}\ (J_q) > \alpha_0\ where\ \ \alpha_0 = 100\ \alpha\%\ \ ordinate\ \ \ of\ \ N(0,1)$$

then the peak at ω_1 is deemed significant and the amplitude of the corresponding harmonic term estimated by

$$\hat{D}_1^2 = 8\pi P(\hat{\omega}_1)/(n-m)....... \tag{15}$$

We now remove the effect of the harmonic by defining

$$\hat{R}_x^{(1)}(s)=\hat{R}_x(s)-0.5\hat{D}_1^2\cos(s\hat{\omega}_1) \tag{16}$$

Let ϖ_1 denote the peak selected by the above procedure. The amplitude of the corresponding harmonic term is now estimated by (15). The process is repeated using $\hat{R}_x^{(1)}(s)$. The successive iterations of $P(\lambda)$ may be denoted by $P^{(1)}(\lambda)$, $P^{(2)}(\lambda)$ etc. At the kth stage the chosen level of significance for testing the peaks in $P^{(k)}(\lambda)$ has however to be adjusted to $(\alpha/k+1)$ in view of degrees of freedom corrections. If none of the peaks at a particular stage in the process say $P^{(r)}(\lambda)$ is significant, then the procedure is terminated at this stage with r harmonics being identified at the frequencies ϖ_j, (j=1,…r).

Bhansali's Correction: Bhansali (1979) has noted that the correction formulae (15) and (16) due to Priestley are not sufficiently accurate as $\hat{R}_x^{(1)}(s)$ is unstable for large values of s. He proposes replacing $\hat{R}_x^{(1)}(s)$ by $\hat{R}_x^{(*)}(s)$, where $\hat{R}_x^{(*)}(s)$ is the autocorrelation at

lag s of the series $\{x^*_t\}$, defined by

$$x_t^* = x_t - \hat{A}_1 \cos(t\varpi_1) - \hat{B}_1 \sin(t\varpi_1) \tag{17}$$
where
$$\hat{A}_1 = (2/N)\sum_{t=1}^{N} x_t \cos(t\varpi_1)$$

$$\hat{B}_1 = (2/N)\sum_{t=1}^{N} x_t \sin(t\varpi_1)$$

Model Estimation:: Suppose by following either of the above procedures, we identify r harmonics at the frequencies ϖ_j (j=1…r)

We then estimate the following model by OLS

$$x_t = \mu + \sum_{i=1}^{r}\{A_i \cos(\hat{\varpi}_i t) + B_i \sin(\hat{\varpi}_i t)\} + u_t \tag{18}$$

The fact that u_t may be correlated is not much of a cause for concern, since Durbin (1960) has shown that for harmonic regressions of the type (18), OLS estimates of A_i and B_i are asymptotically efficient.

We have already seen that the residual term u_t will have a *continuous spectrum*, with an ARMA representation. However, for forecasting purposes, an AR model is more convenient, and Bhansali (1979) shows how an autoregressive model may be fitted to u_t by a suitable lag selection criterion. His preference is for the FPE_α criterion developed in Bhansali and Downham (1977), though the use of other criteria such as AIC, BIC, etc. is also, of course, possible.

Thus we may formulate the following three stages in fitting a *harmonic regression or fixed frequency effects* model.

    (1) We first estimate the *number* of harmonics r using Priestley's $P(\lambda)$ test.

    (2) Next, we estimate the coefficients in the model (10) (using the value of r obtained in Stage 1) by the method suggested originally by Priestley (1981), or (optionally) incorporating the Bhansali (1979) *corrections.*

    (3) Finally, we fit an AR model to the residuals from the model estimated in Stage 2, using a suitable order selection criterion.

4. Fourier-ARMA Models

It is now well recognized that linear ARMA models fall considerably short of accounting for the complex non-linear behaviour of real world financial series. Attempts to capture such complexity have run in diverse directions, with some of the important strands being the following

    (i) explicit non-linear models of the form $x_t = f(x_{t-1}.....x_{t-p}, \varepsilon_t, \varepsilon_{t-1}....\varepsilon_{t-q})$, which for different specifications of f yield the Bilinear, SETAR, STAR etc. models (see Granger & Terasvirta (1993), De Gooijer & Kumar (1992) etc).

    (ii) conditional volatility models such as ARCH, GARCH, EGARCH, GARCH-M etc (see Bollerslev et al (1992))

(iii)   asymmetric adjustment models (see Beaudry & Koop (1993), Potter (1995), Bradley & Jansen (1997) etc.)

(iv)   random-coefficients models (see Chow (1984), Nicholls & Quinn (1982) etc.). However, much of this parametric literature requires the specification of the precise nature of the underlying non-linearity. Since little *a priori* information is usually forthcoming on this aspect, such specification is often of the *ad hoc* variety.

Recently, Ludlow & Enders (2000) have suggested a method for analyzing a time series, based on the universal approximating property of Fourier series, and which does not derive from a prior specification of the exact form of non-linearity. It essentially consists in specifying an ARMA model with deterministic time-dependent coefficients, each of which can be approximated by a finite (and reasonably short) Fourier series. Following Ludlow & Enders (op. cit.), we may refer to such models as Fourier-ARMA (or F-ARMA) models. Happily as it turns out, these models are remarkably straightforward to implement, comprising the following main stages.

Step 1: Given the stationary series $\{x_t\}$, t = 0, 1,2….(N-1), we start by estimating the best fitting ARMA model (using the standard Box-Jenkins methodology)

$$x_t = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \, x_{t-i} + \sum_{j=1}^{q} \hat{\beta}_j \, \varepsilon_{t-j} + \varepsilon_t \tag{19}$$

Let the SBC (Schwarz Bayesian Criterion) value of (19) be denoted as $SBC^{(0)}$

and let $\hat{\varepsilon}_t$ denote the residuals of (19).

Step 2: The method proceeds by singling out one particular coefficient in (20) as being particularly prone to time-variability (usually α_1 or β_1 is adequate in practice).[vii] Suppose this coefficient is α_m . Let $k_j = (2\pi j/N)$, j=1,2….(N/2) and estimate the models

$$\hat{\varepsilon}_t = A_j \sin(k_j t) x_{t-m} + B_j \cos(k_j t) x_{t-m} + v_t \, , \;\; j=1,2.....(N/2) \tag{20}$$

and let $SBC^{(j)}$ denote the SBC value of (20).

Include k_j as a relevant frequency in the model iff $SBC^{(j)} < SBC^{(0)}$.

This process[viii] will usually lead to a set of relevant frequencies k_j* , j=1,2..r, (with $r \leq (N/2)$) and a set of significant coefficients from A_j* and B_j*.

Step 3: We now re-estimate our original model including the significant frequencies and coefficients from Step 2 i.e.

$$x_t = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \, x_{t-i} + \sum_{i=1}^{q} \hat{\beta}_i \, \varepsilon_{t-i} + \sum_{j=1}^{r} \left[A_j^* \sin(k_j^* t) + B_j^* \cos(k_j^* t) \right] x_{t-m} + \varepsilon_t \qquad (21)$$

Further diagnostic checks can be performed on the model (21) and in particular any of the coefficients A_j^*, B_j^* which emerge as insignificant (via standard t-tests) can be dropped.

A major advantage of this model is that it is linear in x_t rendering it particularly suitable for forecasting purposes.

5. Kalman Filter Model

Originally suggested in the engineering literature by Kalman, in the early 1960s, as a technique for estimating the state of a noisy system, the Kalman filter has emerged as a highly flexible as well as powerful tool in the econometrician's kit. In the course of econometric practice, the method has evolved over several refinements. Here we present a standard version, closely following Harvey (1989). Detailed treatments of the subject may be found in Jazwinski (1970), Kim and Nelson (1999), Durbin and Koopman (2001) etc.
The system under study is viewed as modeled by two equations:

(Transition Equation) $z_t = \Phi_t z_{t-1} + R_t \eta_t$ (22)

(Observation Equation) $x_t = C_t z_t + \varepsilon_t$ (23)

Here z_t is of dimension (kx1) and is called the *state vector* of the system at time t and x_t is the (mx1) vector of observations at time t. The stochastic disturbance terms η_t and ε_t satisfy the following assumptions

(i) $\eta_t \sim N(0, \Sigma)$ (ii) $\varepsilon_t \sim N(0,\Omega)$ (iii) $E(\eta_t, \varepsilon_t) = 0$, t=0,1…..(N-1)

The advantage of the above formulation lies precisely in the flexibility of its application.[ix]

As we are concerned with univariate applications, we deal with a much simplified version of the above in which m=1. We also assume further that R_t=I

Details of the solution in this special case, as well as more general cases, are available in Hamilton (1994), p. 373-408

6. Benchmark Models

As mentioned in Section 1, in order to assess the comparative performance of the newer group of methods (discussed in the preceding sections) we need standard benchmark models. Among the models frequently resorted to in modeling asset prices, conditional heteroscedasticity models figure prominently. Thus we selected as our benchmark models, three standard conditional heteroscedasticity models as well as a standard random-walk model. Since the literature relating to the conditional heteroscedastic models has now become fairly well-known, we need confine ourselves only to a brief recapitulation of their salient features.

Conditional Heteroscedastic Models : A specific class of models (named ARCH(q)) to deal with problems of conditional heteroscedasticity (a common occurrence in financial time series) was suggested by Engle (1982). Since then several refinements to Engle's(1982) basic model have been forthcoming on a continual basis. These refinements may be subsumed under the generic appellation of GARCH models. Following Nag and Mitra (2002), a useful taxonomy for the GARCH family is furnished by considering the following general model for the given series $\{x_t\}$

$$x_t = \beta^T y_t + \delta \ h_t^2 + \varepsilon_t \qquad (24)$$

where the y_t vector contains the independent (and also lagged dependent) variables of the model and the conditional variance h_t^2 is given as

$$h_t^2 = Var\langle \varepsilon_t | I_{t-1}\rangle \qquad (25)$$

I_{t-1} being the information set at time (t-1).

The conditional variance itself is modeled as

$$h_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \phi_j h_{t-j}^2 \qquad (26)$$

We now distinguish various types of GARCH models[x] in terms of (24) and (26)

Case1 (ARCH) : If in (24), δ=0 and p=0 in (26), we have the original ARCH(q) model of Engle (1982).

Case 2 (ARCH-M) : If in (24), δ≠0, but p = 0, q ≥ 0 in (26), we have the ARCH-M (ARCH-in-mean) model suggested in Engle et al (1987).

Case 3 (GARCH) : If in (24), δ=0, but p ≥ 0, q ≥ 0 in (26), we have the GARCH(p,q) model (accredited to Bollerslev (1986)).

Case 4 (GARCH-M) : If δ ≠ 0, and p ≥ 0, q ≥ 0, we have the GARCH-M (GARCH-in-mean) model (see Engle et al (1987))

Case 5 (EGARCH) : While the GARCH models discussed so far have been successful in capturing the phenomenon of *volatility clustering* in asset returns, they fail to account for one important observed feature of asset returns viz. that volatility tends to rise in response to "bad news" (excess returns lower than expected) and fall in response to "good news" (excess returns higher than expected). To capture these features, Nelson (1991) suggested a more general formulation in which (26) is replaced with

$$Ln(h_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{j=1}^{p} \phi_j h_{t-j}^2 \qquad (27)$$

where g(.) must be a function of both the magnitude and sign of z_t, and

$$z_t = \left(\frac{\varepsilon_t}{h_t} \right) \qquad (28)$$

Nelson (1991) suggests using the following form for g (.)

$$g(z_t) = \theta z_t + \lambda \left[|z_t| - E|z_t| \right] \qquad (29)$$

with θ and λ constants.
  In practice, (29) is the form of g(.) usually employed and for most applications the EGARCH(1,1) suffices (i.e. p=q=1 in (27)).

For the purposes of this paper, we plan to use three alternative models from the above menu viz. GARCH (1,1), GARCH-M (1,1) and EGARCH (1,1).

The final benchmark model is simply the random-walk model, which uses the latest available observation as the forecast.

7. Forecasting Results

We now turn to a presentation of our forecasting results. Our forecasting exercise applies to the 1-year, 5-year and 10-year rates of interest in India. These are identified as the yields to maturity[xi] on GoI (government of India) dated securities traded in the SGL (secondary general ledger) segment of the RBI (Reserve Bank of India).The data is published on a weekly basis (RBI Weekly Statistical Supplement available at the website address *www.rbi.org.in*). The period of analysis was selected as 6 March 1998 to 11 March 2005, incorporating a total of 363 weekly observations.[xii] In view of the unit roots evident in all the three series of interest rates, one-period differencing was resorted to.[xiii] Of the 362 observations now at our disposal, we use 300 as the *training sample* (i.e. the parameters of each model are estimated over the initial 300 observations). The remaining 62 observations (spanning 12 December 2003 to 11 February 2005) constitute the *forecasting sample*, over which we generate 1-period ahead forecasts. In computing the 1-period ahead forecasts, we incorporate successively the actual values of new observations but do not re-estimate the parameters.

Empirical Estimation of Models I-IV: We now present essential estimation details of the four *second-generation models* that we have discussed earlier.
Model I (ANN-Wavelets): Various types of wavelets were experimented with but there did not seem to be much difference in the outcomes. We therefore settled for the simplest

Haar wavelet described by the low-pass filter $(g_0, g_1) = \left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}} \right)$. So far as the

activation function is concerned, several choices are available such as the linear, sigmoidal, arc tan, radial basis etc. We experimented with the linear and the radial basis functions, and present the results for the linear case only. As already mentioned we take $A_1 = A_2 = \ldots = A_J = B_m = p$ (=4), and the ANN is solved via the *back-propagation* method , using the Widrow-Hoff (1960) learning algorithm, yielding the forecast x_N^f.
Model II (Mixed Spectum): For the mixed spectrum model, (with an incorporation of the

Bhansali correction)*,* we estimated models of the type (18), and certain broad details are

presented below in Table 1

Model III (Fourier-Non-linear ARMA): The estimates for model (21) are presented below

in Table 2.

Model IV (Kalman Filter): For this exercise, we posit an ARMA(2,2) with time-varying

coefficients.

<p style="text-align:center"><u>Table 1</u></p>

<p style="text-align:center">(Estimates for Model II)</p>

	r	Frequencies ϖ_i	Order of AR model for u_t	Coefficients of AR model for u_t
One-Year	1	$\varpi_1 = 0.3560$	2	$\phi_1 = -0.1912$ $\phi_2 = -0.0168$
Five-Year	1	$\varpi_1 = 0.3142$	1	$\phi_1 = -0.1382$
Ten-Year	3	$\varpi_1 = 0.3142$ $\varpi_2 = 1.4451$ $\varpi_3 = 2.5761$	2	$\phi_1 = -0.3376$ $\phi_2 = -0.0252$

Notes: Notation as per text. The order of AR model for u_t is selected via the Bhansali-Downham FPE_α criterion

<p style="text-align:center"><u>Table 2</u></p>

<p style="text-align:center">(Estimates for Model III)</p>

	p	q	r	Coefficients of trigonometric terms in (24)
One-Year	19	0	2	$A_1 = -0.021*$ $B_1 = -0.0177*$ $A_2 = -0.0728*$ $B_2 = -0.0645*$
Five-Year	1	0	1	$A_1 = 0.0051*$ $B_1 = 0.0175$
Ten-Year	1	0	2	$A_1 = -0.0152*$ $B_1 = 0.0161*$ $A_2 = -0.0008*$ $B_2 = -0.0152*$

Notes: (i) Notation as per equation (21)
(ii) (*) denotes significance at 5%

Forecast Evaluation (Conventional Measures): The evaluation of forecasts is of paramount importance as enabling the analyst to weed out unsatisfactory models and to isolate a set of adequate models. Both *absolute* and *relative* evaluation measures have been suggested in the literature and we use both in this paper.

The *absolute* measures that we employ in this paper are

(i)   RMSE (Root Mean Square Error) = $\sqrt{\dfrac{1}{N-T}\sum_{t=T+1}^{N}e_{t/t-1}^{2}}$

(ii)   MAXAE (Maximum Absolute Error) = $Max\{e_{t/t-1}\}_{t=T+1}^{N}$

(iii)   MAPE (Mean Absolute Percentage Error) = $\left(\dfrac{1}{N-T}\right)\sum_{t=T+1}^{N}\left[\dfrac{1}{A_t}\right]\ |e_{t/t-1}|$

(iv)   Theil's U statistic which for any model M is defined as

$$U_M = \left[\dfrac{RMSE(M)}{RMSE(RANDOM\ WALK)}\right]$$

(v)   We also introduce 2 additional accuracy measures viz. D (the *proportion* of times the model correctly forecasts the sign of the series) and TP (the *proportion* of times the model correctly forecasts a change in the sign of the series)

In the above $e_{t/t-1}$ is the one-period ahead forecast error associated with the model being tested, A_t is the absolute value of the actual observation at time t, N is the total number of observations, and T is the *training period* (see above).

Note that our forecasts all pertain to *interest rate changes*, and for the participants involved in the money market and the government securities market, losses increases more than proportionately to the error magnitudes (so that a criterion like the RMSE becomes relevant) and also a large forecasting error in either direction can impose a very heavy cost on the market participants (hence our use of the MAXAE criterion in addition to the MAPE). The U statistic is a minimal accuracy check to see that our model at least outperforms the simplest benchmark model viz. Random Walk model (Model VIII in our terminology). Very often, for money market decision makers, forecasting the direction of interest rate change becomes an important criterion as also the prediction of turning points, and hence our use of the two additional measures D and TP.

We now present details of the forecasting results for the 1-year interest rate series (see Table 3), restricting ourselves only to qualitatively summarizing the results for the 5-year and 10-year interest rates.

Table 3

(1-year Interest Rates)

(1)	RMSE (2)	% Improvement in RMSE over best benchmark model (3)	MAXAE (4)	% Improvement in MAXAE over best benchmark model (5)	MAPE (6)	% Improvement in MAPE over best benchmark model (7)	U Statistic (8)
Model I	0.1671 (2)	5.11%	0.4896 (2)	7.67%	1.3257 (2)	10.44%	0.5838
Model II	0.1750 (3)	0.62%	0.5085 (3)	4.11%	1.5760 (6)	-6.46%	0.6115
Model III	0.3556 (8)	-101.93%	0.8685 (8)	-63.78%	9.9446 (8)	-571.80%	1.2424
Model IV	0.1621 (1)	7.95%	0.4677 (1)	11.80%	0.9829 (1)	33.60%	0.5664
Model V	0.1768 (6)	-0.40%	0.5354 (6)	-0.96%	1.5166 (4)	-2.45%	0.6177
Model VI	**0.1761** **(4)**	0.00%	0.5325 (5)	-0.41%	**1.4803** **(3)**	0.00%	0.6153
Model VII	**0.1761** **(4)**	0.00%	**0.5303** **(4)**	0.00%	1.5400 (5)	-4.03%	0.6153
Model VIII	0.2862 (7)	-62.52%	0.8110 (7)	-52.93%	3.8579 (7)	-160.62%	1.00

Notes:

(i)   Figures in parentheses in Columns (2) and (4) represent the models ranked by the respective criteria (with the best model allotted Rank 1 etc.)

(ii)   The boldface entry in Columns (2) and (4) indicate the *best benchmark model*.

(iii)   Ties are given equal rank, and the subsequent rank is omitted (following the standard ranking convention)

(iv)   Column (3) gives the percentage improvement in RMSE for each model as compared to the *best benchmark model* (with negative entries corresponding to a worse performance with respect to this model). Columns (5) and (7) carry similar information with respect to the MAXAE and MAPE criteria.

Table 4

(Directional and Turning Point Statistics)

	D			TP		
	1-year	5-year	10-year	1-year	5-year	10-year
Model I	36	35	31	0	0	0
	(2)	(3)	(3)	(5)	(3)	(5)
Model II	33	23	31	2	0	11
	(3)	(7)	(3)	(3)	(3)	(2)
Model III	24	29	39	11	3	4
	(7)	(5)	(2)	(2)	(2)	(3)
Model IV	57	57	54	18	14	18
	(1)	(1)	(1)	(1)	(1)	(1)
Model V	**32**	**36**	**26**	0	0	0
	(4)	**(2)**	**(5)**	(5)	(3)	(5)
Model VI	27	30	**26**	0	0	0
	(6)	(4)	**(5)**	(5)	(3)	(5)
Model VII	**32**	26	25	0	0	0
	(4)	(6)	(7)	(5)	(3)	(5)
Model VIII	14	19	17	**1**	0	**1**
	(8)	(8)	(8)	**(4)**	(3)	**(4)**
Total	**62**	**62**	**62**	**42**	**27**	**25**

Notes : Same as (i) to (iii) of Notes to Table 3 above.

We now try to analyse the message contents of our tables. To facilitate the discussion we refer to models I to IV as the *candidate models* (we continue to refer to models V to VIII as *benchmark models)*

1. So far as the short –term (1-year) interest rate is concerned (see Table 3), two of our *candidate* models (I and IV) out-rank the best performing benchmark model, irrespective of whether the evaluation criterion is RMSE, MAPE or MAXAE. Of these Model IV (Kalman Filter) is the best, showing a marked improvement of performance (7.95% in RMSE , 11.80% in MAXAE and 33.60% in MAPE) over the best benchmark model. The wavelet-neural network model I is a fairly close second on all three criteria.The mixed spectrum model II performs well on two of the criteria (RMSE and MAXAE) but slips on the MAPE. The Fourier non-linear ARMA model is a disappointment emerging last among all the models, irrespective of the criterion used, with a performance even worse than the simple random walk (Theil's U statistic substantially over 1).

2. Analysis of the longer tenures of interest rates, reiterates the broad features of Table 3, but less sharply. Model IV emerges among the top three ranks on all the criteria in the case of 10-year rates and in two of the

three criteria (RMSE and MAXAE) in the 5-year case. Performance of models I and II , however is not clear-cut, depending upon the forecast evaluation criterion used. Model III performs poorly again.

3. The top honours among the benchmark models are evenly shared between the EGARCH (1,1) and GARCH-M models (Models VI and VII respectively).

4. We now turn to Table 4 which presents the details regarding the statistics D and TP. Because our results pertain to interest rate *changes*, the statistic D captures the accuracy of our models in the prediction of the direction of interest rate (level) movements, while TP refers to the capacity of the respective models to capture the turning points in the interest rates (levels). The performance of the benchmark models in capturing *turning points* is extremely poor, as also that of the candidate Model I. Model IV is the most successful with success rates ranging from 38% to 72%, which should be considered excellent given the well-known difficulty of accurately predicting turning points. The success of the remaining two candidate models remains fairly low.

5. In contrast to the somewhat disappointing scenario with respect to turning point prediction, the situation regarding the prediction of directional movements (as indicated by D in Table 4) is more sanguine. Model IV once again emerges a clear winner, with a success ratio of nearly 90% throughout. Model I also performs creditably, with more than 50% success ratio in all interest rate categories. Model II ranks fairly high on the 1-year and 10-year categories but quite low on the 5-year category. Model III performs well only on the 10-year interest rate. The performance of the benchmark models is markedly inferior to that of the candidate models (with isolated exceptions).

Forecast Comparisons (Formal Tests of Significance): The forecast evaluation procedures discussed so far could be faulted on the grounds that there is little formal statistical theory underpinning them.

A number of earlier attempts at formal comparisons between competing forecast models, were grounded in restrictive assumptions such as quadratic loss functions, Gaussianity and absence of serial correlation in forecast errors etc. (see e.g. the Morgan-Granger-Newbold test (Morgan (1939), Granger & Newbold (1977)). We therefore consider two modern tests, which are not tied to such restrictive assumptions viz.
    (i)   The d-test of Diebold and Mariano (1995)
    (ii)   The *forecast encompassing* test proposed by Chong and Hendry (1986)

Diebold-Mariano's d-test: Suppose we are interested in comparing 2 models A and B, whose 1-period ahead forecast errors at time t, we denote by $e^A_{t/t-1}$ and $e^B_{t/t-1}$ respectively. Define

$$d_t = g\left(e^A_{t/t-1}\right) - g\left(e^B_{t/t-1}\right) \qquad (30)$$

where g(.) is a loss function.

Suppose we have R 1-period ahead forecasts from models A and B, and construct the series $\{d_t\}^R_{t=1}$, then Diebold & Mariano (1995) show that *asymptotically* the statistic

$$S = \left[\frac{\bar{d}}{\sqrt{2\Pi \hat{f}_d(0) \big/ R}} \right] \quad \text{is distributed as N(0,1) under the null hypothesis } H_0 \text{ that}$$

models A and B are equivalent so far as forecasting ability is concerned.

(Here $\bar{d} = \left(\frac{1}{R}\right)\sum_{t=1}^R d_t$ and $\hat{f}_d(0)$ is the estimated spectrum of the series $\{d_t\}^R_{t=1}$ at the zero frequency).

For large R,

$$2\Pi \hat{f}_d(0) \approx \sum_{\tau=-(Q-1)}^{(Q-1)} \hat{\gamma}_d(\tau) \qquad (31)$$

where $\hat{\gamma}_d(\tau)$ is the autocovariance of the series $\{d_t\}^R_{t=1}$ at lag τ and Q is a suitable truncation parameter (usually $Q \approx R^{1/3}$).

We apply the d-test to effect pair-wise comparisons of the out of sample forecasting performances of our models, using a quadratic loss function. If the null hypothesis H_0 is rejected, we test which model is superior using a one-tailed test. In the interests of brevity of presentation, only the comparisons across candidate models are considered here (see Table 5). The fact that R=62, in our case implies that our sample is modest-sized and the asymptotic results need some care in interpretation.

Table 5 presents the d-statistics for the pair-wise comparisons. The entry in the cell (i,j) is interpreted as the d-statistic with the role of Model A assigned to the model in column j and Model B being taken as the model in row i. The results of the table may be summarized as follows:

(i)   For 5-year and 10-year interest rates, Model IV is unequivocally superior to the other three candidate models. For the 1-year rate the same pattern holds, except that Models I and IV are indistinguishable on the d-criterion.

(ii)   Across the range of interest rates considered, Model III is clearly inferior to the other three candidate models

(iii)   Model I fares significantly better than Model II, except for the 5-year case, where there does not seem to be a statistically significant difference between the performance of the two models.

Table 5

(Diebold-Mariano 'd' test)

	Model I	Model II	Model III	Model IV
One-Year Interest Rate				
Model I		2.2029*	5.4187**	-1.6986
Model II			5.0936**	-3.8172**
Model III				-5.4845**
Model IV				
Five-Year Interest Rate				
Model I		1.4651	7.0002**	-2.2394*
Model II			6.6596**	-3.2873**
Model III				-7.2226**
Model IV				
Ten-Year Interest Rate				
Model I		2.3467**	3.4890**	-3.0624**
Model II			3.2611**	-3.2414**
Model III				-4.1032**
Model IV				

Note: The notations (*) and (**) refer to significant values at 5% and 1% levels respectively.

Forecast Encompassing: The *forecast encompassing* test seeks to determine whether the forecast from one model (say A) incorporates all the statistically relevant information from a competitor model (say B). If this happens then A is said to *forecast encompass* B. The following model is estimated using a consistent estimator

$$x_t = \phi_A x_t^A + \phi_B x_t^B + \varepsilon_t, \ t = T+1, \ ....N \qquad (32)$$

where x_t is the actual value of the series, x_t^A, x_t^B, are the forecasts from models A and B, and {T+1…N} is the period over which the forecasts are made.

We now have the following three possibilities

(i) $(\phi_A, \phi_B) = (1,0)$ then A *forecast encompasses* B

(ii) $(\phi_A, \phi_B) = (0,1)$ then B *forecast encompasses* A

(ii) neither of the above is true, in which case neither model *forecast encompasses* the other.

The results of the forecasting encompassing exercise are reproduced in Table 6.

Table 6

(Forecast Encompassing Test)

	Model I	Model II	Model III	Model IV
One-Year Interest Rate				
Model I		(N,N)	(Y,N)	(N,N)
Model II			(Y,N)	(N,N)
Model III				(N,N)
Model IV				
Five-Year Interest Rate				
Model I		(Y,N)	(Y,N)	(N,N)
Model II			(Y,N)	(N,N)
Model III				(N,N)
Model IV				
Ten-Year Interest Rate				
Model I		(Y,N)	(Y,N)	(N,N)
Model II			(Y,N)	(N,N)
Model III				(N,N)
Model IV				

The entry in the cell (i,j) of each interest rate panel panel is to be interpreted as follows. If Model i forecast encompasses Model j (for the interest rate under consideration) , the first entry in the cell (i,j) is put as Y (otherwise N) and if Model j forecast encompasses Model i, the second entry is put as Y (otherwise N) . The striking feature about Table 6 is that the qualitative features are uniform across the three interest rates considered (except for one solitary instance) and these common features are listed below

(i) Model I forecast encompasses Models II and III (but not IV) for the 5-year and 10-year interest rates, but in the one-year case, it only forecast encompasses Model III

(ii) Model II forecast encompasses Model III (but not IV)

(iii)   Model IV neither forecast encompasses any other model, nor is it forecast encompassed by any other model.

8. CONCLUSIONS

In view of their pronounced volatility in a deregulated environment, forecasting interest rates constitutes a formidable challenge. Linear models have, by and large, turned in disappointing results, falling considerably below the aspirations of market players and regulators, alike. What seems to be indicated is the need for methods which can satisfactorily account for the myriad departures from standard assumptions (such as stationarity, Gaussianity, linearity etc.) exhibited by interest rates in practice. A wide variety of GARCH models has now become established, as a particularly convenient framework for asset price forecasting, and any new method needs to be assessed with respect to the GARCH models as benchmarks. In this paper, four alternative methods are examined (Models I to IV—see Section 1 for a description of their salient features), which have emerged in recent years but have not so far been applied extensively in the interest forecasting context.

Our four *candidate* models are compared with a benchmark set comprising three varieties of GARCH models and a random-walk model. A number of comparison criteria were deployed for evaluating model forecasting performance. While some variation in model performance was in evidence across the comparison criteria as well as the three interest rates categories, quite a few conclusions displayed a fair amount of robustness. Two such conclusions are the remarkably superior performance of Model IV and the rather disappointing performance of Model III. Models I and II perform fairly well, with the former holding a significant edge over the latter. While Model IV almost always out-performs the benchmark models, Model I (and to a lesser extent Model II) do so creditably often. Only Model III rarely rises above the benchmark performance and quite often slips below even the random walk model.

The search for superior forecasting models is an endless one. This article has tried to project the potentiality of some newer models for forecasting volatile series such as interest rates. Our best performing model (Kalman Filter), however, being a well-established method in econometrics, hardly needs any special advocacy now. Model III, in spite of its intuitive appeal, needs considerable elaboration and refinement before it can become a serious contender in the forecasting race[xiv]. Our analysis points to Models I and II as newer alternative forecasting methodologies, that might hold interesting promise for the future.

REFERENCES

Bartlett, M.S. (1954). Problemes de l'analyse spectrale des series temporelles stationnaires, *Publications de l'Institute Statistique, Universite de Paris,* 3, 119-134

Beaudry, P. & Koop, G. (1993). Do recessions permanently change output ?, *Journal of Monetary Economics*, 31, 149-163

Bhansali, R.J. (1979). A mixed spectrum analysis of the lynx data, *Journal of the Royal Statistical Society, Series A,* 142, 199-209.

Bhansali, R.J. & Downham, D.Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion, *Biometrika,* 64, 547-551

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, Oxford : Oxford University Press

Bjorn, V. (1995). Multiresolution methods for financial time series prediction, in *Proceedings of the IEEE/IAFE 1995 Conference on Computational Intelligence for Financial Engineering*, 97,

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

Bollerslev, T., Chou, R. & Kroner, K. (1992). ARCH modeling in Finance :A review of the theory and empirical evidence, *Journal of Econometrics,* 52, 5-59

Bradley, M & Jansen, D. (1997). Nonlinear business cycle dynamics : Cross-country evidence on the persistence of aggregate shocks, *Economic Inquiry,* 35, 495-509

Chong, Y.Y. & Hendry, D.F. (1986). Econometric evaluation of linear macroeconomic models, *Review of Economic Studies,* 53, 671-690

Chow, G. (1984). Random and changing coefficient models. In : Z. Griliches & M.D. Intrilligator (Eds.), *Handbook of Econometrics II,* (pp. 1213-1243). Amsterdam : North-Holland

Coifman, R.R. & Donoho, D.L. (1995). Translation invariant de-noising. In : A. Antoniadis & G. Oppenheim (Eds.), *Wavelets and Statistics, Lecture Notes in Statistics,* 103 (pp. 125-150). New York : Springer-Verlag

Daubechies, I. (1992). *Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics,* 61, Philadelphia : SIAM

De Gooijer ,J.G. & Kumar, K. (1992). Some recent developments in non-linear time series : Modeling, testing and forecasting. *International Journal of Forecasting,* 8, 135-156

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

Dua, P. and Pandit, B.L. (2992). Interest rate determination in India: Domestic and external factors, *Journal of Policy Modelling*, 24, 853-875.

Dua, P. Raje, N and Sahoo, S. (2003). *Interest rate modeling and forecasting in India*, *Technical Report, DRG-DEAP, Reserve Bank of India*.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society Series B*, 22, 139-153.

Durbin, J. and Koopman, S.J. (2001). *Time Series Analysis by State Space Models*, Oxford: Oxford University Press.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50, 987-1008.

Engle, R.F., Lilien, D.M., and Robins, P.R. (1987). Estimating time-varying risk premia in the term structure: The ARCH-M model. *Econometrica*, 55, 391-407.

Engle, R.F. and Ng. V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48, 1749-1778.

Gabor, D. (1946). Theory of communication. *Journal of the IEE*, 93, 429-457.

Gençay, R. and Qi, M. (2001). Pricing and hedging derivative securities with neural networks and Bayesian regularization, early stopping and bagging. *IEEE Transactions on Neural Networks*, 12, 726-734.

Gençay, R., Selçuz, F. and Whitcher, B. (2002). *An Introduction to Wavelets and other Filtering Methods in Finance and Economics*, London: Academic Press.

Granger, C.W.J. and Newbold, P. (1977). *Forecasting Economic Time Series*, Orlando, Fl: Academic Press.

Granger, C.W.J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. New York: John Wiley & Sons.

Haar, A. (1910). Zur Theorie der orthogonalen funkionesysteme. *Mathematiische Annalen*, 69, 331-371.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Harvey, A.C. (1989). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Heutschel, L. (1991). *The absolute value GARCH model and the volatility of U.S. returns*. Technical Report, Princeton University.

James, J. and Webber, N. (2000). *Interest Rate Modelling*, New York: John Wiley and Sons.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Kane, E.J. (1984). *Microeconomic and Macroeconomic Origins of Financial Innovation*. Boston: Kluwer – Nijhoff.

Kim, C.J. and Nelson, C. (1999). *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge: MIT Press.

Kuan, C.M. and White, H. (1994). Artificial neural networks: An econometric perspective (with discussion). *Econometric Reviews*, 13, 1-91.

Ludlow, J. and Enders, W. (2000). Estimating non-linear ARMA models using Fourier Coefficients. *International Journal of Forecasting*, 16, 333-347.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11(7), 674-693.

Melard, G. (1985). An example of the evolutionary spectrum theory. *Journal of Time Series Analysis*, 6, 81-90.

Morgan, W.A. (1939). A test for significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, 31, 13-19.

Nachane, D. (2004). Time-frequency analyses for nonstationary time series. *Journal of Quantitative Economics*, (New Series) 2(2), 41-57.

Nachane, D. and Ray, D. (1993). Modelling exchange rate dynamics: New perspectives from the frequency domain. *Journal of Forecasting*, 12(5), 379-394.

Nag, A.K. and Mitra, A. (2002). Time series modeling with genetic neural networks: Case studies of some important Indian economic and financial series. *Statistics and Applications*, 4(1), 37-58.

Nelson, D.B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347-370.

Nicholls, D. & Quinn, B. (1982). *Random Coefficient Autoregressive Models : An Introduction. Lecture Notes In Statistics,* 115. New York : Springer-Verlag

Percival,D.B. & Walden, A.T. (2000). *Wavelet Methods For Time Series Analysis.* Cambridge (U.K.) : Cambridge University Press.

Potter,S. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics,* 10, 109-125

Priestley, M.B. (1964). Estimation of the spectral density function in the presence of harmonic components. *Journal of the Royal Statistical Society, Series B,* 26, 123-132

Priestley, M.B. (1981). *Spectral Analysis And Time Series.* London : Academic Press.

Rabemanajara, R. & Zakoian, J.M. (1993). Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics,* 8, 31-49

Ramsey, J.B., Uskinov, D. & Zaslavsky, G.M. (1995). An analysis of US stock price behaviour using wavelets. *Fractals*, 3, 377-389

Ramsey, J.B. & Zhang, Z. (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance*, 4, 341-372

Refenes, A. & Azema-Barac, M. (1994). Neural networks applications in financial asset management. *Neural Computation Applications*, 2, 13-39.

Starck, J.L., Murtagh, F. & Bijaoni, A. (1998). *Image Processing And Data Analysis : The Multi-Scale Approach*. Cambridge : Cambridge University Press.

Swanson, N. & White, H. (1995). A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business and Economic Statistics,* 13, 266-275

Widrow, B. & Hoff, M.E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record, New York,* 4, 96-104

Zurbenko, I.G. (1991). Spectral analysis of non-stationary time series. *International Statistical Review*, 59, 163-174

ENDNOTES

---

[i] In fairness, spectral analysts have also been much concerned with time-frequency resolution, the work on time-varying spectra by Priestley (1981), Gabor (1946), Melard (1985), Zurbenko (1991) and others, being prime examples. But it is now generally agreed that wavelets are a far more convenient way of representing time-varying complex data patterns than time-varying spectral analysis (see Percival & Walden (2000))

[ii] A *high-pass (low-pass) filter* preserves the high frequency (low frequency) properties of the underlying series, while discarding the low frequency (high-frequency) features.

[iii] The Haar wavelets are the earliest and simplest form of wavelets, and still retain their popularity in applied work.

[iv] The method for obtaining $\mathbf{q_j}$ is given in Gencay et al (2002), p. 121.

[v] Starck *et al* (1998) also prove that the method is efficient, in the sense of minimizing the MSE (mean square error).

[vi] The technical definitions can be recovered from any standard text such as Bishop (1995) or Kuan & White (1994).

[vii] The method can be generalized to more than one coefficient without much difficulty, but this is likely to lead to over-parametrization of the final model. Hence, working with a single coefficient would usually lead to more efficient results.

[viii] It could happen that for none of the j, it is true that $\text{SBC}^{(j)} < \text{SBC}^{(0)}$, in which case no asymmetries seem to be present to any significant extent on account of α_m, and either our choice of α_m as the non-linear coefficient is wrong or a linear model might be the appropriate one.

If the coefficient in question is β_m , the procedure is essentially the same except that (21) is replaced with

$$\hat{\varepsilon}_t = A_j \sin(k_j t)\hat{\varepsilon}_{t-m} + B_j \cos(k_j t)\hat{\varepsilon}_{t-m} + v_t$$

[ix] Thus if C_t is interpreted as a vector of independent variables and $\Phi_t = R_t = I$, the system (31), (32) is a *random coefficient regression equation*. Similarly, defining Φ_t and R_t as matrices of autoregressive and moving average coefficients respectively, the system may be viewed as an ARMA model (with time varying coefficients) and measurement errors.

[x] Apart from the models considered below, we have the *quadratic* GARCH (QGARCH) model (Engle & Ng (1993)), the *absolute* GARCH (AGARCH) model (Heutschel (1991)) and the *threshold* ARCH (TARCH) model (Rabemanjara and Zakoian (1993)) and several others.

[xi] The calculation of this yield is described in RBI Bulletin (June 2005), p.S-543

[xii] Jan 1998 witnessed mayhem in the Indian government securities market, in the aftermath of the Asian crisis, when yields rose to unprecedented levels. To avoid unnecessary influence of this extreme situation on our results, we thought it best to commence the analysis from a later period, when the yields had returned to their normal levels.

[xiii] Except for the wavelet-based neural network model, all the other methods considered here require the underlying series to be stationary. In the interest of uniformity, we use first differences throughout.

[xiv] Of course, the Kalman filter is a well-established method in econometrics, so it hardly needs any special advocacy now.