Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Estimating term structure of interest rates: neural network vs one factor parametric models

F. Abid & M. B. Salah
Faculty of Economics and Busines, Sfax, Tunisia

Abstract

The aim of this paper is twofold; first we concentrate on the work of Vasicek (1977) and Cox, Ingersoll and Ross (1985). We examine and test empirically each model and discuss its performance in predicting the term structure of interest rates using a parametric estimating approach GMM (Generalized Moments Method). Second we estimate the term structure of interest rate dynamics using a nonparametric approach ANN (Artificial Neural Network). Two neural network models are performed. The first model uses spreads between interest rates of 10 different maturities as the only explanatory variable of interest rate changes. The second model introduces two factors, spreads and interest rates' levels. Using historical U.S. Treasury bill rates and Treasury bond yields, we compare the ability of each model to predict the term structure of interest rates. Data are daily and cover the period from 3 January 1995 to 29 December 2000. Results suggest that neural network, Vasicek (1977) and Cox, Ingersoll and Ross (1985) models generate different yield curves. Neural network models outperform the parametric standard models. The most successful forecast is obtained with a two factors neural network model.

1 Introduction

Modeling and predicting the term structure of interest rates have attracted considerable attention in the literature. Based on restrictive hypothesis concerning the number of state variables and their dynamics, Vasicek [9] and Cox *et al* [3] develop one-factor parametric models that attempt to capture the particular feature of the observed yield curve movements. Although, the one factor term structure of interest rate models are recognized to be analytically tractable and easy to implement, empirical findings strongly suggest that interest rate dynamics specified by those models conflict with observed interest rates' behavior. In more recent works, Ait Sahalia [1], Stanton [7], Jiang [4], Cottrell *et al* [2], Wood and Dasgupta [10] and Tappinen [8] propose nonparametric models with no restriction on the functional form of the process generating the structure of interest rates. Estimating non-parametrically the yield curve may avoid the misspecification of parametric model and the discrete approximation errors, and may lead to a better fitting to the observed term structure of interest rates. This paper tries to add evidence about the ability of the nonparametric approach, developing one and two factors neural network models to predict the term structure of interest rates. For comparison purposes we use GMM to estimate Vasicek [9] and Cox *et al* [3] models.

2 Data description

Term structure of interest rate data, are provided by the Federal Reserve statistical release. Data describe historical changes of American Treasury bill rates with maturities 6 and 12 months as well as American bond yields with maturities 2, 3, 5, 7, 10 and 30 year. Empirical analysis covers the period from 3 January 1995 to 29 December 2000 providing 1484 observations in total. For parameter estimates and predicting purposes the entire period is divided in two sub-periods. The first one, from 3 January 1995 to April 1998, is used to generate neural network inputs. The second sub-period, from May 1998 to December 2000, reproduces the observed yield curve, used as desired output for neural network models.

American 3-month Treasury bill rates, which are supposed to capture the volatility of short-term interest rates, are used to estimate the spot rate process. Data are in daily basis and cover the period from 1st May 1998 to 29 December 2000, providing 670 observations in total.

3 Research design and methodology

As mentioned above, the empirical study focuses on estimating Vasicek [9] and Cox *et al* [3] one-factor models using GMM and constructing neural network models to predict yield curve dynamics.

In Vasicek [9] work, term structure of interest rate is modeled using short term interest rate as the unique state variable. Interest rates are assumed to change according to a stochastic mean reverting process that can be described by the following equation:

$$dr = \beta \left(\alpha - r\right) dt + \sigma\, dz \tag{1}$$

where dz is a standard Brownian motion process and r, the current level of interest rate. Parameter α is the long-run average and coefficient β is the speed of adjustment of interest rate toward its long run average level. As the process is mean reverting, interest rate is pulled back over the time to some long-run average level, at a rate β. When r is high, mean reversion tends to cause it to have a negative drift; when r is low mean reversion tends to cause it to have positive drift. Vasicek's term structure of interest rates is extracted from the following equation:

$$R(r,t,T)=\alpha+\frac{\lambda\sigma}{\alpha}-\frac{\sigma^2}{2\alpha^2}+\frac{\left(1-e^{-\beta(T-t)}\right)}{\beta(T-t)}\left(r(t)-\alpha-\frac{\lambda\sigma}{\beta}+\frac{\sigma^2}{2\beta^2}\right)+\frac{\sigma^2}{4\beta^3(T-t)}\left(1-e^{-\beta(T-t)}\right)^2 \tag{2}$$

In Cox *et al* [3] model interest rates are assumed to change according to a mean reverting squared root process described by the following expression:

$$dr = \beta(\alpha - r)dt + \sigma\sqrt{r}dz \tag{3}$$

The square root term on the volatility ensures that the interest rate will not fall below zero. The Cox *et al* term structure is extracted from the following equation:

$$R(r,t,T)=\frac{B(T-t)}{T-t}r(t)-\frac{\left(\tilde{\beta}+\gamma\right)}{2\beta\alpha}\frac{In\left(A(T-t)\right)}{(T-t)} \tag{4}$$

with

$$A(t,T)=\left(\frac{2\gamma\, e^{\left(\tilde{\beta}+\gamma\right)(T-t)/2}}{(\tilde{\beta}+\gamma)\left(e^{\gamma(T-t)}-1\right)+2\gamma}\right)^{\left(\tilde{\beta}+\gamma\right)\left(\alpha+\frac{\lambda\sigma}{\beta}-\frac{\sigma^2}{2\beta^2}\right)}$$

$$B(t,T)=\frac{2\left(e^{(T-t)}\right)}{\left(\tilde{\beta}+\gamma\right)\left(e^{\gamma(T-t)}-1\right)+2\gamma}$$

$$\gamma=\left(\tilde{\beta}+2\sigma^2\right)^{1/2}$$

$$\tilde{\beta}=\beta+\lambda\sigma$$

Vasicek [9] and Cox *et al* [3] interest rate term structure models are function of parameters α, β, σ and λ. Hence, estimating term structure of interest rates, come down to estimate parameters that define the short term interest rate dynamics process, as well as the market price of interest rate risk. Using a discrete-time econometric specification, we estimate parameters of the continuous riskless interest rate process by the following specification:

$$r_{t+1} - r_t = \beta(\alpha - r_t) + \varepsilon_{t+1} \tag{5}$$

$$E[\varepsilon_{t+1}] = 0\,,\ E\left[\varepsilon_{t+1}^2\right] = \sigma^2 r^\gamma \tag{6}$$

where γ takes 0 for Vasicek model [9] and 1 for Cox *et al* model [3]. The discrete time specification has the advantage of allowing variance of interest rate changes to depend directly on interest rate's level in a way compatible with continuous-time process. Discrete time approximation error is of second order importance if

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

changes in r are measured over short time periods. This approximation is estimated using GMM of Hansen [11]:

Define θ to be the parameters' vector with elements (α,β,σ^2) and given $\varepsilon_{t+1} = r_{t+1} - r_t - \beta(\alpha - r_t)$, estimators of these parameters are obtained from the first and second moment conditions in the vector $f(\theta)$:

$$f(\theta)=\begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} r_t \\ \varepsilon_{t+1}^2 - \sigma^2 r_t^{\gamma} \Delta t \\ \left(\varepsilon_{t+1}^2 - \sigma^2 r_t^{\gamma} \Delta t\right) r_t \end{bmatrix} \tag{7}$$

In order to estimate parameters α, β and σ, we define two instrumental variables; a constant term and, r_t. With four orthogonalilty restrictions and three parameters to estimate, the system is over identified.

The market price of interest rate risk is estimated by minimizing the sum of squared deviations across maturities, between the historical yield curve and the yields generated by Vasicek [9] and Cox *et al* [3] model (the average yield curve observed in the period from 1st May to December 2000). Having estimated parameters describing the interest rate process and the market price of risk, we investigate implications on the term structure dynamics.

The neural network approach is used to predict interest rates' changes three-month forward. We estimate term structure of interest rates with two different neural network specifications. The first model uses the same approach as in Yan Tappinen [8]. According to expectation theory, spreads between, 12 and 6-month Treasury bill rates, 3 and 2-year, 7 and 5-year, 10 and 7-year, 30 and 10-year bond yields, are used to predict 6 month, 2, 5, 7 and 10-year interest rates changes three months later (expectation theory stipulates that spreads between long and short term maturities predict future interest rate changes). The setting of expectation theory is broadened to use the whole historical interest rate term structure to estimate future yield curve. In other words, to predict 6-month interest rate changes, not only 12 month-6 month spread is used but also all spreads between interest rates of different maturities. Hence, the input's vector contains spreads and the output's vector is composed of interest rate changes.

$$\Delta R^{m}_{t+3,t} = f\left(R_t^{12m} - R_t^{6m}, R_t^{3y} - R_t^{2y}, R_t^{7y} - R_t^{5y}, R_t^{10y} - R_t^{7y}, R_t^{30y} - R_t^{10y}\right) \tag{8}$$

This first neural network model (NNM1) contains 5 nodes in the input layer and 5 nodes in the output layer. Function f in expression (8) is a non-linear arctangent bipolar function.

Following Wood and Dasgupta [10] work, the second neural network model (NNM2), would refer not only to information provided by interest rates' spreads, but also to that contained in interest rates' levels. Litterman and Scheinkman [6] argue that presence of interest rate level as well as spread is necessary to deduct the intertemporal changes in the term structure of interest rates. Thus, The

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

second neural network model (NNM2) would contain ten nodes in the input layer (5 interest rate levels plus 5 spreads).

$$\Delta R^{m}_{t+3,t}=f\left(R^{12m}_t-R^{6m}_t,R^{3y}_t-R^{2y}_t,R^{7y}_t-R^{5y}_t,R^{10y}_t-R^{7y}_t,R^{30y}_t-R^{10y}_t,R^{6m}_t,R^{2y}_t,R^{5y}_t,R^{7y}_t,R^{10y}_t\right) \tag{9}$$

Using a multilayer perceptron architecture and arctangent bipolar function ($f(x) = \arctan(2x/3.14159)$) as non-linear transfer function, we extract the term structure of interest rates produced by each neural network model.

Forecasting error ($E = \frac{1}{n}\sum_{j=1}^{n}\left(d_j - y_j\right)^2$ Mean Square Error) for each observation is determined by the difference between the desired values d_j and the computed values y_j. To perform the neural network model, we divide our database in two sets. For the training set, used to estimate the model, observations range from 2 February 1998 to 29 December 2000. As the predictive horizon is three months, inputs are drawn from the period from 2 February1998 to 29 September 2000 and the outputs, with a lag of three months, should belong to the period from 1st May 1998 to 29 December 2000. For the test set, used to evaluate the forecasting performance of the model, observations are drawn from the period from 3 January 1995 to 31 December 1997. The inputs are from the period from 3 January 1995 to 30 September 1997 and the outputs are from the period from 3 April 1995 to 31 December 1997.

4 Empirical findings and result analysis

Estimation of parametric models is carried out using the GMM technique of Hansen [11].

Table 1 shows parameter estimates of Vasicek [9] and Cox *et al* [3] models. The market prices of risk λ are estimated by minimizing the sum of squared deviations of the model implied yield curve from the average yield curve of the sample. Their values are respectively of, 3.66 and 2.76.

Table 1: Parameter estimates of the spot rate process (numbers in parentheses are standard deviations of the estimates).

Parameters	Vasicek process	CIR process
α	0.05416 (0.00352)	0.05410 (0.00309)
β	0.00582 (0.00303)	0.00662 (0.00303)
σ^2	2.5789 E-7 (0.0005)	4.6587E-6 (0.0009)

From eqns (2) and (4) we deduct the term structure dynamics of Vasicek [9] and Cox *et al* [3] models. Parametric and observed yield curves are then compared to those produced by neural network models.

Among a large number of architecture experimentation, two neural network models that lead to the smallest forecasting error have been retained. Selection is based on the following criteria: (1) the best percentage of training and test correct classification, (2) the minimum difference between training and test correct classification, (3) the simplest neural network structure. According to these criteria, a preliminary selection is made to retain the best iteration for each architecture, then a definitive selection is carried out to choose the best architecture.

For first neural network model, the architecture composed with three hidden layers, one node in each one is selected.

For the second neural network model, the best architecture is composed with one hidden layer and two nodes. We note that while the neural network model with ten inputs (NNM2), one hidden layer is sufficient to identify the complex pattern in data, the neural network model with five inputs (NNM1) needs three hidden layers. Thus, the addition of five inputs to the neural network architecture has allowed identifying the complexity in data with a lesser number of hidden layers.

The two neural network models (NNM1 and NNM2) provide estimates of interest rate changes for a three-month horizon and allow extracting the level reached by each interest rate three month forward. The neural network model with five inputs (NNM1) is less performing than that of ten inputs (NNM2). Performance is measured here by the degree of matching the direction and the level of observed interest rates.

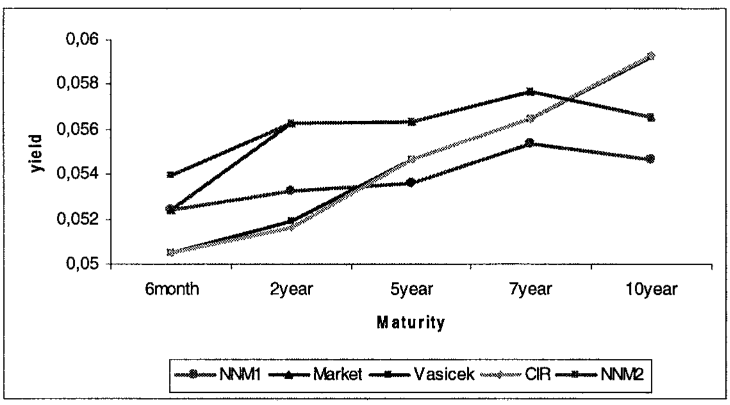

Figure 1: Estimates of parametric and neural network models' yield curves [3].

Figure 1 exhibits that neural network, Vasicek [9] and Cox *et al* [3] models generate different yield curves. The term structure predicted by the two neural network models, tend to be a good fit to the historical yield curve. Inspection of the historical American term structure reveals that, the yield curve tends to be upward sloping at the shorter end (6 month-2 year), relatively flat for maturities between 2-year and 5-year, again upward sloping for maturities between 5-year and 7-year and finally downward sloping for maturities that exceed 7-year. This is in general consistent with yield curves generated by neural network models. In

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

contrast to five inputs' neural network model that matches only the shape of the yield curve, ten inputs' neural network model replicates nearly perfectly the shape and the level of the observed interest rates. This confirms Litterman and Scheinkman [6] conclusion that the presence of interest rate's level as well as spread is necessary to deduct the intertemporal interest rate term structure changes. The resulting yield curve of Cox *et al* model [3] is very close to that of Vasicek [9]. Failure of parametric models to reproduce the historical term structure of interest rates is due to the restrictive functional forms imposed on the dynamics of term structure. This may be a common reason to explain why neural network approach that allow for a maximal flexibility in fitting the data, outperforms the one factor parametric model in modeling a complex pattern as the interest rate term structure dynamics.

5 Conclusion

This paper has presented the results of specifying a neural network model of term structure of interest rates for treasury bonds and estimating mono-factor Vasicek [9] and Cox *et al* [3] models using GMM. The one factor neural network model is less performing than two factors one, but more performing than one-factor parametric models. Interest rates' level has a strong effect on the changes of bond rates and the slope of the yield curve.

References

[1] Ait-Sahalia, Y; Nonparametric pricing of interest rate derivative securities. *Econometrica,* **64**, pp. 527–560, 1996.

[2] Cottrell, M & Debodt, E &. Gregoire P; Simulation de l'évolution de la structure par terme des taux d'intérêt test d'une approche non-paramétrique. SAMOS, university of Paris I, working paper, 1996.

[3] Cox, J. C & Ingersoll, J. E, & Ross S. A; A Theory of the term structure of interest rates. *Econometrica*, **53**, pp. 385–407, 1985.

[4] Jiang, J. G; Nonparametric modeling of US interest rate term structure dynamics and the implications on the prices of derivatives securities. *Journal of Financial and Quantitative Analysis*, **4**, pp. 565–499, 1998.

[5] Jiang, J. G & Knight, J. L; A nonparametric approach to the estimate of diffusion processes, with an application to a short-term interest rate model, *Econometric Theory*, **13**, pp. 615–645, 1997.

[6] Litterman, R & Scheinkman J; Common factors affecting bond returns. *Journal of Fixed Income*, **1**, pp. 54-61, 1991.

[7] Stanton, R; A Nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance*, **52**, pp. 1973–2002, 1997.

[8] Tappinen J., Interest rate forecasting with neural networks, Government Institute for Economic Research, working paper n° 170, 1998.

[9] Vasicek, O. A., An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, pp. 177–188, 1977.

[10] Wood, D & Dasgupta, B; Modeling the term structure of sterling interbank interest rates. *Proc of the 3rd Int. Conf on Neural Network in the capital markets*. Eds Refenes & Apostolos, John Wiley & Sons, pp. 6–49, 1995.

[11] Hansen, L, Large simple properties of generalized method of moments estimator, *Econometrica*, **50**, pp. 1029–1054, 1982.