FORECASTING LONG TERM INTEREST RATE: AN ECONOMETRIC EXERCISE FOR INDIA

Kapil Singh

Indian Institute of Management
Prabandh Nagar, Off Sitapur Road
Lucknow-226013, India
Email: kapil.iiml@gmail.com

and

Rudra Sensarma
Indian Institute of Management
Prabandh Nagar, Off Sitapur Road
Lucknow-226013, India
Email: rsensarma@gmail.com

Version: February 2006

FORECASTING LONG TERM INTEREST RATE: AN ECONOMETRIC EXERCISE FOR INDIA

Abstract

The objective of this paper is to develop a robust model for the forecasting of long term interest rates in India. The variable under study is yield on 10-year government Securities (GSEC10). The paper takes into account models available in the literature for developed as well as developing countries including India. Based on various alternative specifications, we have arrived on a parsimonious specification which is based on macroeconomic theories as well as is useful for generating forecasts. We employ multivariate time-series techniques to develop the model and find that GSEC10 is co-integrated with many macroeconomic variables. Of these, Stock Index, Real Effective Exchange Rate and Money Supply are the most important ones. Based on this, our estimated model is used to generate out-of-sample forecasts. Subsequently the forecasts are compared with naïve models such as random walk and it is found that our model yields superior forecasts.

Keywords: interest rate, vector error correction model.

JEL Classification: G10, G14

1. Introduction

Interest rates are one of the key indicators of the state of an economy. The movement in interest rates affects banks, the corporate world, the Central bank, and even the public. It also influences several other variables like exchange rate, foreign investment and stock markets. Hence it is of primary importance to develop a model to forecast at least the direction in which the interest rates are likely to move. Accordingly this paper investigates the role of domestic and external factors in determining long term interest rates in India. We find that a parsimonious multivariate model based on various macroeconomic variables culled from the literature outperforms a naïve univariate model.

Such forecasts may be especially useful for banks whose business is based on interest spread, as well to other agents in the economy. An accurate model can tell the banks about the future expected inflows and outflows during each of the time periods and hence ensure a proper match between the same. For corporates, interest rates forecast can be a useful determinant of the decision to pick between equity and debt or to choose between a fixed or a floating rate. For example in case the interest rates are expected to fall it makes sense to borrow floating rate loan. Similarly it may come useful in structuring interest rate swap deals. For an investor, the value of a portfolio in the debt market is dependent on the interest rate. Hence it would be of advantage to any investor who can have a good forecasting model at her disposal. For the policy maker (Central bank and the Government) it is important to understand which are the important factors in determining interest rates and what would be the implications of bringing a policy change on the interest rates. In view of the above it is of immense importance to all the agents in the economy to predict the interest rates correctly.

Empirical modelling of interest rates has a long history. Fauvel, Paquet and Zimmermann (1999) present a survey of various empirical models in this context and opine that VAR and VECM enable an integrated treatment of various interest rates, including both their short-term dynamics and any existing long-run relationships. According to them, the forecasting performance of these multivariate models is also usually better than those of univariate models. They conclude that usually long-run equilibrium relationships exist between a macroeconomic system's variables and the VECM can be superior to the VAR in forecasting interest rates over longer horizons. A recent paper by Butter and Jansen (2004) attempted to forecast 10-year German bond yields using a multivariate time-series approach. They found that interest rates are co-integrated with various macroeconomic factors, e.g. business cycle indicators and yield on foreign bonds. Specifically for India, Dua and Pandit (2002) have identified various domestic and external factors determining interest rates in India. They studied behaviour of short term and long term interest rates. However their modelling of long term interest rates was limited to 12-month Treasury bills. Dua, Raje and Sahoo (2004) compared various models including Vector Error Correction Model (VECM) and Vector Auto Regression (VAR) and found that Bayesian VAR performs best for long term interest rates. However this study used data for the period April 1997 to September 2002 which was a period of continuously falling interest rates. However the interest rate scenario changed thereafter and the rates started hardening from the second half of 2003.

In light of the above, the present paper uses an updated data-set (which includes the recent period of rising interest rates) to model long term interest rate (10-year government bond yields) in India based on macroeconomic variables. After identifying the most important variables that determine long term interest rate, an attempt has been made to generate out-of-sample forecasts. The model is compared in terms of forecasting accuracy with other naive models. It is shown that our model outperforms naïve models in terms of accuracy of forecasts. The rest of the paper proceeds as follows. In section 2 recent trends in long term interest rate in India has been reviewed. Section 3 presents an overview of extant literature on interest rate modelling. Section 4 introduces the data used in the paper. Section 5 discusses the econometric methodology adopted in the paper. Section 6 presents the results and section 7 compares the forecast from our model with some naive models. Finally section 7 concludes the paper.

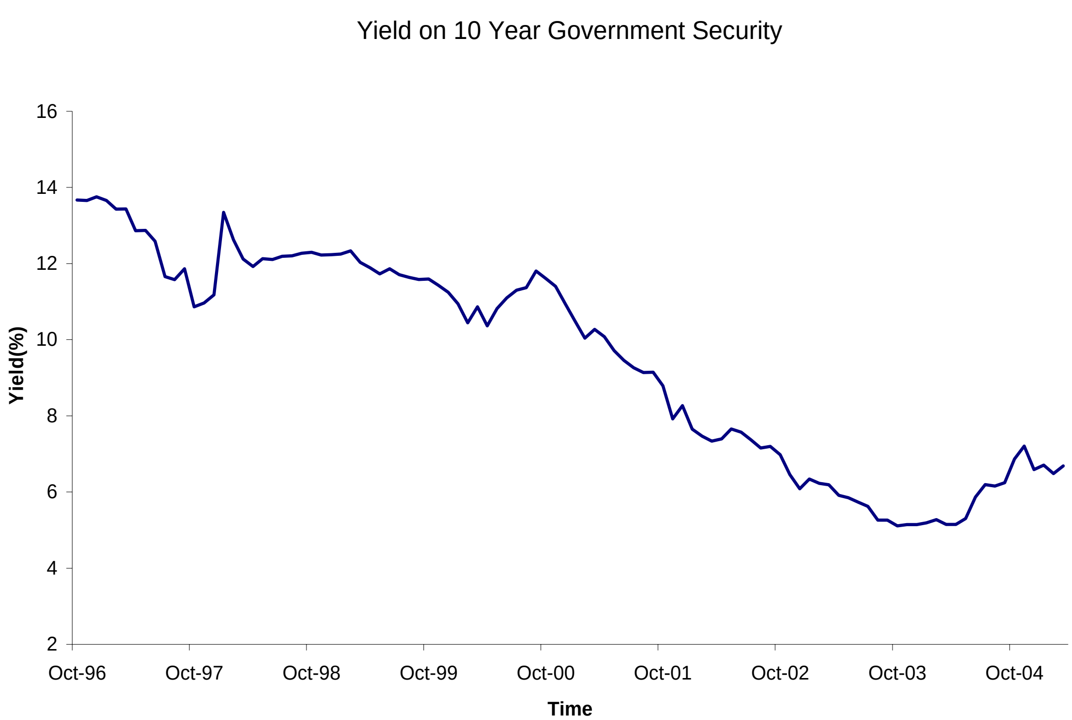
2. Interest Rates in India

Interest rates in India were highly regulated before the liberalization of the economy in 1991. Since then sequential deregulation has been followed and interest rates are now largely market determined, e.g. banks are not forced to lend at a particular interest rate except for priority sector loans. In line with macroeconomic reforms, the government securities market has been developed with a view to creating an efficient and deep market for these instruments. Figure 1 presents the data on long term interest rate in India, specifically the yield on 10-year government securities. The figure clearly shows that interest rate levels have come down over the last 10 years, as the Reserve Bank of India

(RBI) continued to follow an easy monetary policy to enable the economy to recover from the industrial slowdown of the mid-nineties. .

However the figure shows that the rates have started picking up after hitting rock bottom in 2003. This was caused by inflationary threats arising out of domestic factors as well as rising international oil prices. Moreover, the financial markets were witnessing surplus liquidity caused by persistent inflows of foreign capital into the Indian economy and consequent intervention of the Central Bank in the forex market in order to prevent appreciation of the domestic currency. The RBI was forced to review its interest rate outlook and raise the policy rates slowly. As a result, the interest rates in the economy firmed up. This might or might not be a short term phenomenon like it was during the 1999-2000 period. The future trend depends on a variety of factors, e.g. RBI's view on inflationary threat. In the present paper, an attempt has been made to investigate the direction of movement of long term interest rate based on the behaviour of various macroeconomic variables.

Figure 1: Interest rates trend in India for the past 10 years



3. Literature Review

While there are a number of interest rate modelling exercises available in the literature, we review a few of the recent papers here. Fauvel, Paquet and Zimmermann (1999) provide an exhaustive survey of the empirical literature. Bidarkota (1998) set up a univariate unobserved components model for the realized real interest rates in the U.S. and a bivariate model for the nominal rate and inflation which imposes co-integration restrictions between them. He concluded that the error-correction model provides more accurate one-period ahead forecasts of the real rate within the estimation sample whereas the unobserved components model yields forecasts with smaller forecast variances.

A recent application of interest rate modelling can be found in Butter and Jansen (2004) who have conducted an analysis of long term yields on government paper in Germany. In their model they have made an attempt to include macroeconomic variables to forecast interest rates. They used quarterly time series data from 1982 to 2001 to develop an equation for predicting yield on German 10-year government paper. While they use a variety of econometric specifications, in most of the models studied by them they find that other interest rates are important in explaining the interest rates. They find that the important variables in the modelling exercise are 3-month German Libor, Government Balance, US 10-year yield, German IFO business activity indicator, Japanese 10-year yields and Real Effective Exchange rate. The problem with these models is that they only confirm integration of world markets but cannot be used for forecasting based on macroeconomic trends. Moreover for the calculation of quarterly government balance the yearly balance has been divided by four which distorts the data into the form of a step function which may not be an appropriate proxy for quarterly government balance.

There have been a few Indian exercises in interest rate modeling. Dua and Pandit (2002) studied behavior of short-term and long term interest rates (viz. commercial paper rate, 3-month Treasury bill rate, 12-month Treasury bill rate). Their empirical results indicate the existence of a co-integrating relationship between real interest rates, real government expenditure, real money supply, foreign interest rates and the forward premium. The estimations also show that movements in interest rates are Granger caused by both domestic and external factors.

Dua, Raje, Sahoo (2004) have developed a model which explains 10-year government security yields in India. After this they have tried to make predictions for the next nine months. They establish both VECM as well as VAR models for the prediction of interest rates and find that VAR models outperform VECM models. They attribute this to a smaller sample size. We feel that while it may be due to their short sample period, i.e. April 1997 to December 2001, during which the interest rates were continuously on a decline. The explanatory variables used by them are inflation rate (year-on-year), bank rate, yield spread, credit, foreign interest rate (6-months Libor), and forward premium (6-months).

Based on the various models developed in the empirical literature we selected certain macroeconomic variables for our analysis. The variables and their impact on the interest rates are being discussed below: The sign in bracket indicates the direction of movement of interest rates consequent upon a positive movement in the particular variable

- **(+) Spot Oil Prices**: It is expected that as the Spot Oil prices go up the import bill will increase meaning that dollars flow out of the Indian economy. This will result in Rupee falling and the interest rates going up based on interest rates parity theory.
- **(+) Wholesale Price Index:** as per the Fisher's interest rate theory investors want to be compensated for inflation so that they get a constant real rate of return. Hence the interest rates will rise.
- **(+) Index of Industrial production:** This is a surrogate measure of the GDP. Since this figure is reported monthly it was used in place of GDP. Higher industrial production (which coincides with higher production in other sectors) would mean a higher expected demand for goods leading to higher demand for money and hence higher interest rates.

- **(-) Real Effective Exchange Rate:** As the exchange rate rises the interest rate parity theory implies that interest rates should fall,
- **(+) Yield on 10-year U.S. government security:** The Indian markets if assumed to be integrated with global markets should move in a similar direction. To prevent financial capital from moving out of India, when the U.S. Gsec10 rises, Indian interest rates would also rise.
- **(+) Bombay Stock Exchange Index:** As the stock market prices rise money will flow into the stock market and demand for bonds will reduce. A lower demand for bonds will result in lower prices which would imply higher interest rates.
- **(-) Money supply:** As the money supply in the economy will increase, easy availability of credit will be associated with lower interest rates.

4. Data

For our analysis we considered monthly data from October 1996 to March 2005. Then out-of-sample forecasts were made for a period of eight months from April 2005 to November 2005. The variables considered were yield on 10-year Government of India Securities (GSEC10), Spot Oil Prices (SpotOil), Wholesale Price Index (WPI), Index of Industrial production (IIP), Real Effective Exchange Rate (REER), yield on 10-year U.S. government securities (USGsec10), Bombay Stock Exchange Index (BSE100) and money supply (M3). The data for the SpotOil and USGSEC10 was obtained from Economic research website of Federal Reserve Bank of St. Louis. All other data was obtained from the Handbook of Statistics on Indian Economy published by the Reserve bank of India available on its website. A graph of the normalized variables being considered by us is shown in the Appendix.

5. Econometric Methodology

All our time-series variables exhibit non-stationarity, i.e. they are not mean reverting. We denote them as I(1) variables, i.e. they require differencing once to make them stationary or the order of integration is one. To estimate long-run relations among non-stationary variables, the vector error correction model (VECM) is used. It not only statistically specifies the short-run dynamics of each variable in the system; it does so in a framework that anchors these dynamics in a manner consistent with long-run equilibrium relationships suggested by economic theory. For instance, such relationships have been shown to exist empirically across interest rates of different maturities (through a term structure long-run equilibrium), across comparable interest rates between two countries (through along-run international equilibrium condition), and across interest rates of different securities of a given maturity. [1]

Empirically, the literature on forecasting tends to support the superiority of the VECMs for longer-horizon forecasting (Fauvel, Paquet and Zimmermann, 1999). Other issues in practice relates to the determination of the appropriate number of unit roots, and of the dimension of the cointegrating space. Over-differencing tends to worsen the quality of short-term forecasts, while under-specifying the number of unit roots affect adversely the longer-term forecasts.

The approach to be followed for modelling is as follows:

- If the I(1) series do not have a long term relationship (i.e. not cointegrated), then we can estimate unrestricted VAR in first differences
- If they are cointegrated, we have to study cointegrated VAR or VECM

If a linear combination of I(1) variables is stationary, then the variables are said to be cointegrated.Two time series are said to be co- integrated of order d,b, denoted as CI(d,b) if :

- They are both integrated of order d

[1] Many authors have studied the cointegrating properties of interest rates and used them for forecasting. Hall, Anderson, and Granger (1992) found 10 cointegrating or long-run relationships underlying the dynamics between the 1- to 11-month US Treasury bills.

- But there is some linear combination of them that is integrated of order b (b<d)

Suppose x_t and y_t , are both I(1). If there exists a constant A, such that $z_t = x_t - A y_t$ is I(0), then xt and yt are said to be cointegrated

$\Delta Y_t = (\phi - 1) Y_{t-1} + e_t$

Generalizing this to a vector framework:

$\Delta Y_t = (A - I) Y_{t-1} + e_t$

If A-I has zero rank, i.e. a null matrix, then the model can not talk of relationship among the y variables Therefore, test for the presence of non-zero rank of A-I or the presence of non-zero eigen values of A-I. Non-zero ranks or non-zero eigen values gives the number of cointegrating vectors.

The Granger representation theorem says that any co integrating relationship can be expressed as a VECM.

The Equation for the VECM is as shown below:

$$\Delta X_t = a_0 + \sum_{i=1}^{k} a_{1i}\Delta X_{t-i} + \sum_{j=1}^{k} b_{1j}\Delta Y_{t-j} + \alpha_1 * ECT_{1,t-1} + u_{1t}, where, ECT_{1,t-1} = (X_{t-1} - \beta_1 Y_{t-1})$$

$$\Delta Y_t = a_0 + \sum_{i=1}^{k} a_{2i}\Delta X_{t-i} + \sum_{j=1}^{k} b_{2j}\Delta Y_{t-j} + \alpha_2 * ECT_{2,t-1} + u_{2t}, where, ECT_{2,t-1} = (Y_{t-1} - \beta_2 X_{t-1})$$

The larger the parameter alpha, the faster is adjustment of X to the previous period's deviation from long-run equilibrium. The idea is that a part of the disequilibrium from one period is corrected in the next period. Since the ECM terms and α_1 and α_2 cannot at the same time be equal to zero (following Granger Representation Theorem), there must exist long-term causality in at least one direction. Therefore α_1 and α_2 indicate long-term causality and a's and b's indicate short-term causality

6. Results and Discussion

Firstly all the series were checked for stationarity using the Augmented Dickey Fuller Test. It was found that all the above series are non-stationary.[2] Subsequently all the series were tested for the existence of Cointegration using Johansen's test. Based on rank test and trace statistics it was found that the above series are co-integrated with rank=1.[3] To find out the appropriate lag for testing cointegration, unrestricted VAR was first run on the differenced series with the lag values taken from p=1 to p=4. It was found that the Information Criteria values were minimum for p=1. Hence this was selected as the appropriate lag (see Table 1).

Table 1: Information criteria for unrestricted VAR

p(lag)=	**1**	**2**	**3**	**4**
AICC	18.34574	18.78045	20.02859	21.31639
HQC	18.96097	19.65727	20.78478	21.39769
AIC	18.20627	18.22335	18.66359	18.58106
SBC	20.07051	21.76638	23.90626	25.54465
FPEC	81008136	84310095	$1.39*10^8$	$1.46*10^8$

Since a high correlation coefficient of 0.8 was found between SpotOil and WPI, the cointegration test was run using each of these two variables one at a time as well as keeping both of them together and then the Information Criteria values were compared. The Information Criteria was less for the model containing only WPI and hence it was selected and SpotOil was dropped from the analysis. Also we tried an alternative specification with de-seasonalized IIP, but no improvement in Information Criteria was observed. The Information Criteria Values after dropping SpotOil are given in Table 2.

[2] Results are not reported due to lack of space, but are available on request.

[3] Results are not reported due to lack of space, but are available on request.

Table 2: Information criteria after dropping Spot Oil

P(lag)=	1	2	3	4
AICC	16.74461	16.96526	17.93085	18.74104
HQC	17.23621	17.70175	18.67527	19.16565
AIC	16.64922	16.59467	17.04196	16.99983
SBC	18.09919	19.3301	21.07881	22.35441
FPEC	17048089	16364705	$2.66*10^7$	$2.74*10^7$

Subsequently a Co-integrating relationship was estimated for this complete model. The coefficients of the variables in the ECM as well as their p-values are being reproduced in Table 3.

Table 3: ECM Coefficient Estimates for Complete Model

Variable	Estimate	Standard Error	t-statistic	Pr > |t|
Constant	-0.01971	0.11657	-0.16908	0.557
GSEC10(t-1)	0.00138	0.00308	0.448052	0.000
WPI(t-1)	-0.00256	0.00571	-0.44834	0.513
IIP(t-1)	-0.00056	0.00126	-0.44444	0.896
REER(t-1)	0.00297	0.00663	0.447964	0.245
USGSEC10(t-1)	0.01318	0.02946	0.447386	0.836
BSE100(t-1)	0.00001	0.00002	0.500000	0.309
M3(t-1)	0.00012	0.00027	0.444444	0.205

It can be seen above that most of the p-values are quite high indicating that the coefficients in the VECM based on the complete model are statistically insignificant.

Hence some of the variables were dropped in order of their p-values and the equation was re-estimated till an equation with minimum Information Criteria was arrived at. We refer to this specification as the parsimonious model. The coefficients of the variables in the ECM for the parsimonious model and their p-values are being reproduced in Table 4.

Table 4: Estimates and significance levels of variables for Parsimonious Model

Variable	**Estimate**	**Standard Error**	**t-statistic**	**Pr > |t|**
Constant	-1.04529	0.41634	-2.51	0.014
GSEC10(t-1)	-0.02572	0.01093	-2.35	0.021
REER(t-1)	0.01995	0.00848	2.35	0.021
BSE100(t-1)	0.0001	0.00004	2.5	0.014
M3(t-1)	-0.00024	0.0001	-2.4	0.018

The VECM for this parsimonious model is given as:

$$\Delta GSec10_t = \text{-}1.04529 - 0.02572\,\Delta GSec10_{t-1} + 0.01995\,\Delta REER_{t-1} + 0.0001\Delta BSE100_{t-1} - 0.00024\,\Delta n3_{t-1}$$
$$- 0.02572 * ECT_{1,t-1}, where, ECT_{1,t-1} = (GSec10_{t-1} - 0.77553\,\mathrm{REER}_{t-1} - 0.00398\ \mathrm{BSE100}_{t-1}$$
$$+ 0.00932\ \mathrm{n3}_{t-1})$$

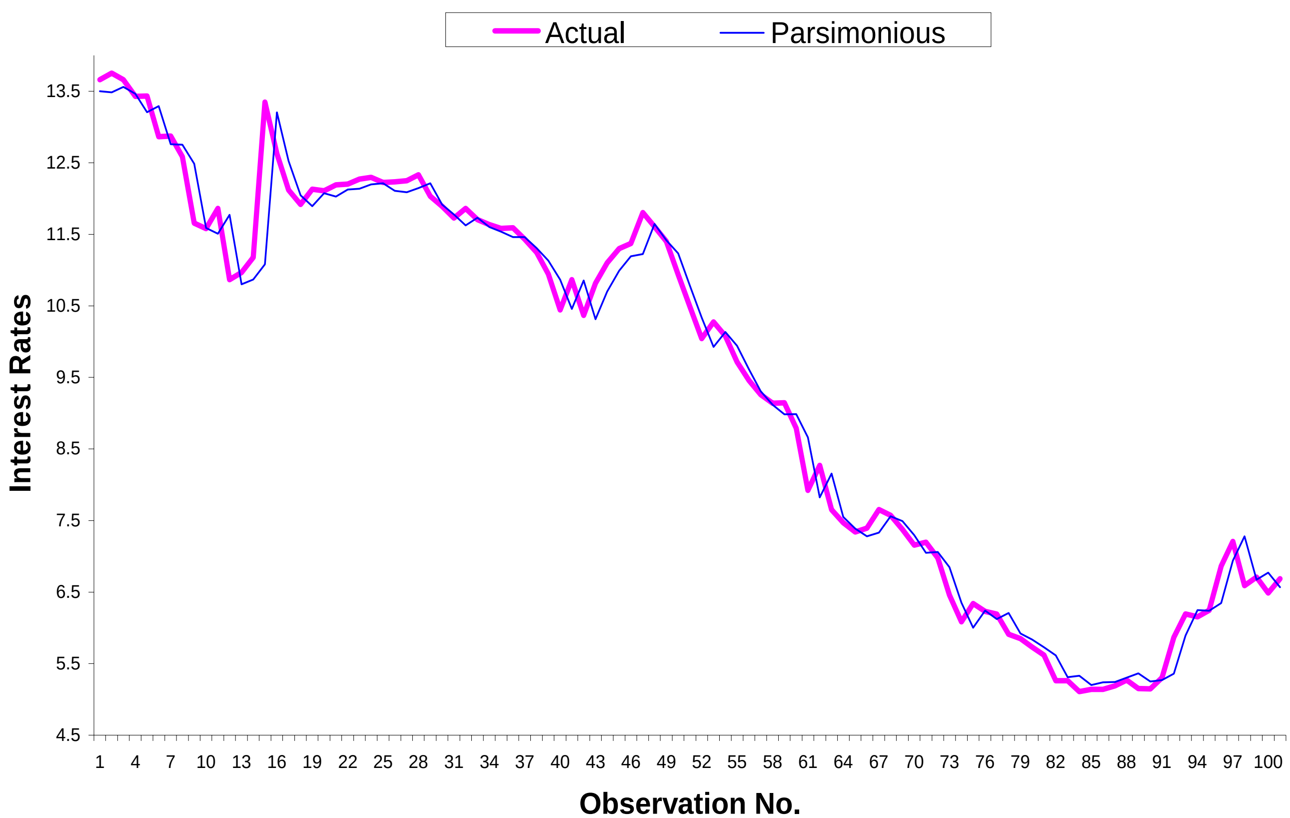
It can be seen from Table 4 that the coefficients are statistically significant at conventional levels for all the variables and hence we can say that all the variables in the above equation are important determinants of long term interest rate. R – Square of the estimated model is 0.98. The other model diagnostics are reported in the appendix. It can also be seen that the signs of all the variable coefficients are as per theoretical prediction discussed earlier. Optimal lag for the parsimonious model is selected based on Information Criteria as reported in Table 5. It can be seen that the Information Criteria for this model are much less than the complete model estimated earlier and hence this also indicates that the parsimonious specification may be a better fit.

Table 5: Information Criteria values for Parsimonious Model

p=	**1**	**2**	**3**	**4**
AICC	13.63213	13.72278	13.94033	14.10265
HQC	13.82115	14.03114	14.33304	14.53689
AIC	13.61151	13.65157	13.78153	13.8114
SBC	14.12935	14.58943	15.14462	15.60505
FPEC	815723.1	850450.9	972504.8	1010114

Fitted values of the interest rate from the parsimonious model are shown in Figure 2. It can be seen from the figure that the actual values are very close to the fitted values throughout the sample period.

Figure2. Fit Equation and actual interest rates, period, Oct 1996-Mar 2005



We also conducted exogeneity test to see whether the explanatory variables as identified above Granger cause GSEC10. The test conducted is a test of weak exogeneity of each variable and the results are reported in Table 6. The null hypothesis of weak exogeneity is rejected for GSEC10 but cannot be rejected for the other variables clearly indicating that that GSEC10 is Granger caused by REER, BSE100 and M3.

Table 6: Weak Exogeneity Test

Variable	**Chi-Square**	**Pr > ChiSq**
GSEC10	6.99	0.0722
REER	0.34	0.5573
BSE100	1.1	0.2944
M3	1.61	0.204

6. Comparison of Forecast Accuracy

To enable a comparison of our parsimonious model with naïve time series models which contain only the lagged value of interest rate, the ARIMA(p,d,q) was proposed as a benchmark. This would enable us to judge whether or not it is important to use additional information for the prediction of interest rates. The p, d and q values for the above ARIMA model were found to be 0, 1 and 0 using the minimum Information Criteria indicating that long term interest rates may follow a random walk.

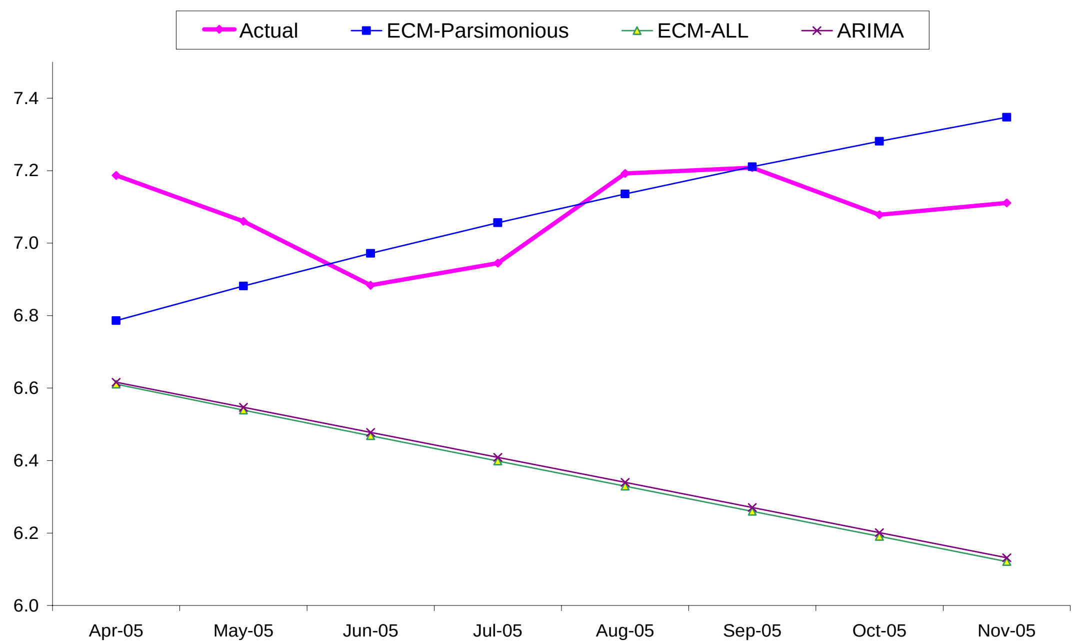
Based on the two benchmark models, viz. random walk and the complete multivariate specification, as well as on our parsimonious model, forecasts were generated for a period of eight months from April 2005 to November 2005. The forecasts were compared with the actual data for this period and the Root Mean Prediction Square Errors (RMPSE) for the three models are compared in Table 7.

Table 7: Comparison of ARIMA, VECM-Complete and VECM-Parsimonious

Model	**RMPSE**
Random Walk	0.7395
VECM-Complete	0.7487
VECM-Parsimonious	0.1975

It can be seen that the parsimonious model performs much better than the other two models in terms of forecasting accuracy. A graphical comparison of the forecasts generated by the three models as compared with the actual data is shown in figure 3.

Figure 3: Forecasts based on ARIMA(0,1,0) (i.e. Random Walk), ECM-Complete and ECM-Parsimonious compared with actual GSec-10



Once again it can be seen from the figure, the parsimonious model is able to predict the future trend correctly. Hence it can be accepted as the superior model.

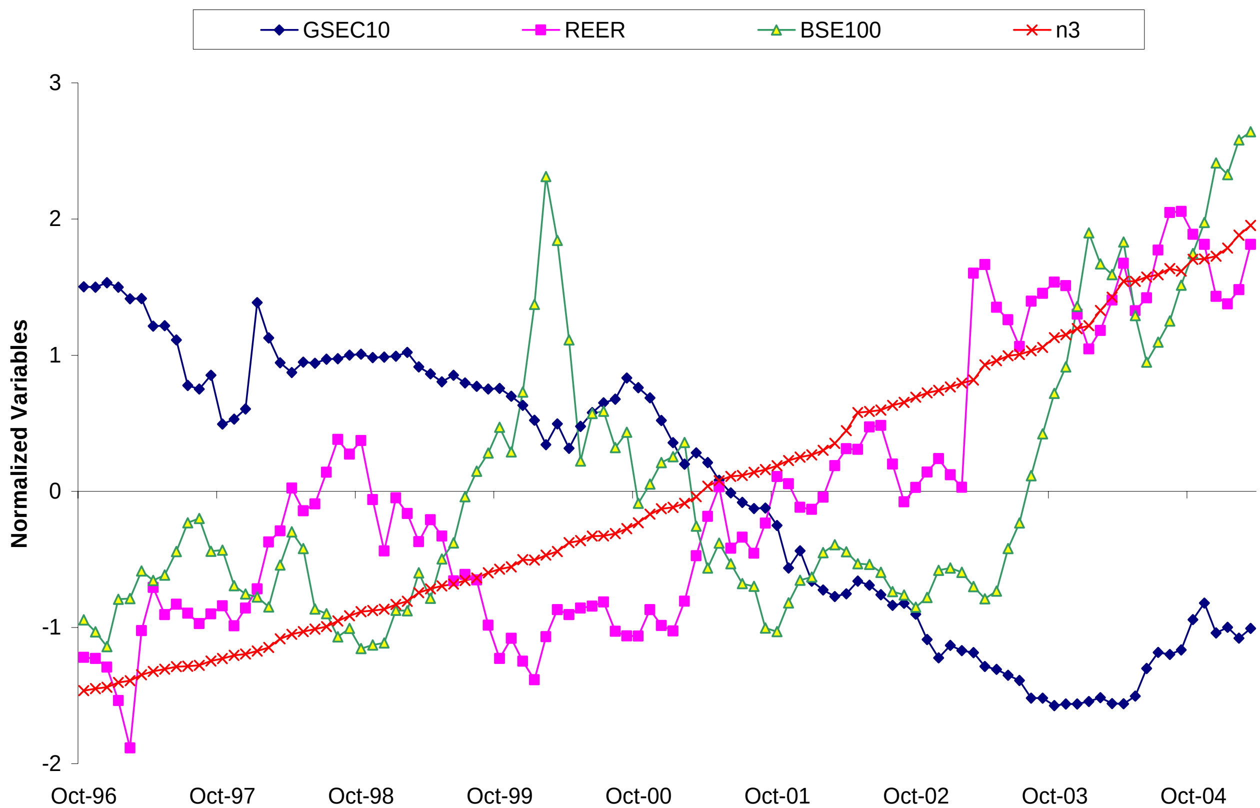
7. Conclusion

In this paper we try to model long term interest rates in India for which we select the yield on 10-year government securities as our variable of interest. Using a vector error correction representation, we try to identify the important variables which can be used for prediction of the interest rate. The analysis reveals that the long term interest rate is co-integrated with various variables which are macroeconomic in nature. Accordingly a complete VECM based on all variables and then a parsimonious VECM are developed and the important variables for predicting long term interest rate are determined.

It is concluded that Money Supply, Stock Index and the Real Effective Exchange rate are the three important variables which help predict the long term interest rates with a high degree of accuracy. A comparative analysis of this parsimonious model with other benchmark or naive models like random walk and the complete multivariate model shows that the parsimonious model can give much better results for forecasting long term interest rates. While our results are able to identify the important determinants of future long term interest rates and our parsimonious model can be used for generating forecasts, it remains to be seen how such a model would perform when compared against non linear techniques such as Artificial Neural Networks. We propose to take this issue up in our future research.

Appendix

1. Normalized Input Variables for parsimonious model



2. Model Diagnostics:

Variable	Durbin Watson	Chi-Square	Pr > ChiSq	F Value	Pr > F
GSEC10	2.05559	630.41	<.0001	0.08	0.7748

Note: The second column contains the Durbin-Watson test statistics; the third and fourth columns show the Jarque-Bera normality test statistics and the corresponding p-value; the fifith and sixth columns show the F-value for Engle's ARCH test and the corresponding p-value. These diagnostic tests indicate that the ECM for GSEC10 does not suffer from problems of auto-correlation, non-normality and ARCH in the error term.

References

Bidarkota, Prasad V. (1998), "The comparative forecast performance of univariate and multivariate models: an application to real interest rate forecasting", *International Journal of Forecasting*, 14(4), 457-468

Caporale, Guglielmo Maria and Nikitas Pittis (1997), "Domestic and external factors in interest rate determination", *Applied Financial Economics*, 7(5): 465-471

Den Butter, Frank A. G. and Pieter W. Jansen (2004), "An empirical analysis of the German long-term interest rate", *Applied Financial Economics*, 14: 731–741

Dua, Pami and B.L. Pandit (2002), "Interest rate determination in India: domestic and external factors", *Journal of Policy Modeling*, 24: 853-875

Dua, Pami, Nishita Raje, Satyananda Sahoo (2004), "Interest Rate Modeling and Forecasting in India", Occasional paper no.3, Centre for Development Economics, Delhi School of Economics

Fauvel, Yvon, Alain Paquet and Christian Zimmermann (1999), "A Survey on Interest Rate Forecasting", Working paper no. 87, CREFÉ