TEXTO PARA DISCUSSÃO N° 1245a

COMPARING MODELS FOR FORECASTING THE YIELD CURVE

Marco S. Matsumura
Ajax R. B. Moreira

Rio de Janeiro, dezembro de 2006

TEXTO PARA DISCUSSÃO N° 1245a

COMPARING MODELS FOR FORECASTING THE YIELD CURVE

Marco S. Matsumura*
Ajax R. B. Moreira**

Rio de Janeiro, dezembro de 2006

* Da Diretoria de Estudos Macroeconômicos do Ipea. Email: marcom@ipea.gov.br.

** Da Diretoria de Estudos Macroeconômicos do Ipea.

Governo Federal

Ministério do Planejamento, Orçamento e Gestão

Ministro – Paulo Bernardo Silva
Secretário-Executivo – João Bernardo de Azevedo Bringel

ipea Instituto de Pesquisa Econômica Aplicada

Fundação pública vinculada ao Ministério do Planejamento, Orçamento e Gestão, o Ipea fornece suporte técnico e institucional às ações governamentais, possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro, e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

URL: http:/www.ipea.gov.br

Ouvidoria: http:/www.ipea.gov.br/ouvidoria

ISSN 1415-4765

JEL C13, C32, C53, E43, E44, E47, G12, G13

TEXTO PARA DISCUSSÃO

Uma publicação que tem o objetivo de divulgar resultados de estudos desenvolvidos, direta ou indiretamente, pelo Ipea e trabalhos que, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

SINOPSE

A evolução das diversas maturidades das taxas de juros está relacionada e pode ser descrita por um número reduzido de variáveis latentes comuns. Os modelos de taxas de juros multivariados da literatura de finanças utilizam esta propriedade, assim como os modelos de fator comum da literatura de séries temporais, e modelos de decomposição da curva de juros. Cada um desses modelos tem vantagens e desvantagens, sendo uma questão empírica avaliar o desempenho dessas abordagens. Esse exercício compara a resposta de quatro modelos alternativos para a curva de juros, em três mercados diferentes: juros domésticos brasileiros, juros soberanos externos brasileiros, e juros domésticos dos Estados Unidos.

ABSTRACT

The evolution of the yields of different maturities is related and can be described by a reduced number of commom latent factors. Multifactor interest rate models of the finance literature, common factor models of the time series literature and others use this property. Each model has advantages and disadvantages, and it is an empirical matter to evaluate the performance of the approaches. This exercise compares 4 alternative models for the term structure using 3 different markets: the Brazilian domestic and sovereign market and the US market.

SUMMARY

Comparing Models for Forecasting the Yield Curve

Marco S. Matsumura*
Ajax R. B. Moreira†

November 2006

Abstract

The evolution of the yields of different maturities are related and can be described by a reduced number of commom latent factors. Multifactor interest rate models of the finance literature, common factor models of the time series literature and others use this property. Each model has advantages and disadvantages, and it is an empirical matter to evaluate the performance of the approaches. This exercise compares 4 alternative models for the term structure using 3 different markets: the Brazilian domestic and sovereign market and the US market.

1 Introduction

Studies after Litterman and Scheinkman (1991) documented that the evolution of the yield curve could be represented by the path of up to 3 latent factors which summarize the yield curve and somehow represent the state of the economy. The intertemporal dependence among the factors describe in a parsimonious way the movements of the yield curve. The yields are given by weighted sums of the state factors. This summarizes the multifactor interest rate models.

The weights can be specified according to approaches that solely emphasize the adherence to data, or that contain no arbitrage restrictions, or which specify a certain shape for the yield curve. Each of the approaches pertain to a different literature. The one that only takes into account the fitting is the common factor model (CF), a standard model in the multivariate time series literature (Harvey, 1989, West and Harrisson, 1997). One of the many models imposing no arbitrage restrictions is the affine model (NA) of Duffie and Kan (1996). Others assume that the yield curve can be described with components with a given shape, for

*Corresponding Author. Instituto de Pesquisa Economica Aplicada, Rio de Janeiro, RJ, Brazil. Email: marcom@ipea.gov.br. Address: Av. Presidente Antônio Carlos, 51 - 17 andar, Sala 1715, 20020-010. Tel: 55 21 3515-8533. Fax: 55 21 3515-8615.

†Instituto de Pesquisa Economica Aplicada, Rio de Janeiro, RJ, Brazil.

example using 1) Legendre Polynomials, Almeida (1998, LP) or 2) the functions proposed by Nelson and Siegel (1987, NS).

Those models possess different characteristics. The NS and LP have less parameters to be estimated, but impose shape restrictions that may not be realistic, and require a number of factors to represent the yield curve that may not be compatible to the number of stochastic sources. The NA model uses a particular rule for the fluctuation of the risk premium and short rate - they are linearly dependent of the state variables -, is more flexible with respect to the format of the curve, has less parameters that have to be estimated than CF model, but some of its parameters, those of the premia, introduce nonlinear characteristics that make it more difficult to estimate. Finally, the CF model is more flexible than the previous ones, easier to estimate, but contains much more parameters. However, this may not be an important deficiency in case the available data has daily frequency.

The model that imposes no arbitrage restrictions is conceptually superior to a purely functional model. It has less scope than a general equilibrium model, but uses less restrictive hypothesis and is more numerically tractable. However, the affine characterization of the model comes from assumptions on the format of the short rate and of the risk premium that may not fit for the Brazilian market, which, until recently, was too concentrated on the short end of the curve. Besides, it is only empirically that it will be possible to verify if the local market is sufficiently ample and liquid to guarantee no arbitrage or if the premia is affine with the state variables.

All the models assumed that the evolution of the yield curve can be described with a reduced number - up to 3 - of latent variables. The CF model is a descriptive representation of the yield curve and can adjust with more flexibility the empirical particularities of the yield curve. Hence it will be used as the benchmark model.

Each model has advantages and disadvantages. It is an empirical matter to evaluate which one has the best forecasting performance. To this end, 3 yield curves will be analyzed: 1) the Brazilian domestic market, given by the Brazilian Futures (BM&F) `DIxPRE` swaps, 2) FED's constant maturity zero-coupon rates extracted from US treasury bonds, and 3) Bloomberg's Brazilian sovereign constant maturity zero-coupon rates extracted from Republic bonds and Brady bonds.

The models were estimated using Monte Carlo Markov Chain - a Bayesian approach (see Gamerman, 1997, and Johannes and Polson, 2003). It constructs samples of the distributions of the estimators and of associated statistics, which permit the construction of performance criteria which take into account the joint effect of the estimator's uncertainty.

The focus of this text is to compare the model's capacity to explain and forecast the yield curve, observing that each one has a different number of parameters. This will be achieved by using 3 largely used criteria: 1) Posterior predictive loss, Gelfand and Ghosh (1998), Banerjee et al (2004), 2) DIC, a generalization of the AIC proposed by Spiegelhalter et al (2002), and 3) a measure proposed by Theil which provides a direct indication of the relevance of the model for forecasting. The next sections present the models, the method of estimation and the obtained results. The last one concludes.

2 Term Structure Models

Let Y be the vector of the n selected interest rate yields, θ the vector of the $p < n$ latent monetary factors that describes the economy. It is assumed that the path of Y is given by the sum of the effect of the state variables, $B\theta$, and of independent errors u. The monetary factors follow a multivariate mean-reverting process with correlated innovations. The weights B price bonds of different maturities with respect to the instantaneous interest rate r, which is the first component of the vector Y. The criterion used to obtain the weights determines the models. Thus,

$$Y_t = A(.) + B(.)\theta_t + \sigma e_t,\ e_t \sim N(0, I_n), \tag{1}$$

$$\theta_t = \mu + \phi\theta_{t-1} + \Sigma u_t,\ u_t \sim N(0, I_p). \tag{2}$$

2.1 Common Factor Model

The model (1, 2) is estimated with unrestricted A and B.

$$Y_t = A + B\theta_t + \sigma e_t,\ e_t \sim N(0, I_n) \tag{3}$$

$$\theta_t = \mu + \phi\theta_{t-1} + \Sigma u_t,\ u_t \sim N(0, I_p) \tag{4}$$

The model as shown is sub-identified. Any nonsingular matrix L applied to the state factors θ will transform it to an equivalent model. Thus, the monetary factors are undetermined. It can be identified in many ways. Here, we adopt the same identification as proposed by Dai and Singleton (2000) for affine models, which consists of $E(\theta) = 0, V(\theta) = I$, upper triangular ϕ.

2.2 Modified Legendre Model

At each moment, the yield curve can be seen as a function that relates an interest rate to a maturity. This function changes along the time. The Legendre polynomials constitute a base for the space of functions, and the variation of the curves is represented by the alteration of the relative importance of the components of the base. Almeida et al (1998) used this representation to describe the evolution of the yield curve at each instant. Incorporating to this construction an equation for the transition of the components, we have a specification for the model (1), (2) where the matrix of weights B are components of Legendre polynomials. It is denoted as (LP). More specifically:

$$y_t^n = \theta_{1t} + \theta_{2t}x + \theta_{3t}(3x^2 - 1)/2 + \theta_{4t}(5x^2 - 3x)/2 + e_{nt}\ ,\ \text{or} \tag{5}$$

$$y_t^n = \theta_{1t} + B_{2n}\theta_{2t} + B_{3n}\theta_{3t} + B_{4n}\theta_{4t} + e_{nt}, \tag{6}$$

$$x = 2n/n^* - 1, \text{ where } n^* \text{ is the greatest maturity.} \tag{7}$$

In this construction, the first component describes the level of the rates, the second describes the inclination, and the third the curvature. The changes of yield curve are described by up to 4 stochastic components. Empirical evaluation will indicate the relative importance of the components and the number of latent factors necessary to represent the yield curve in each market.

2.3 Modified Diebold and Li Model

Another well know model of the yield curve is Nelson and Siegel (1987). This decomposition was used by Diebold and Li (2005), testing its forecasting performance on government bonds. Incorporating a transition equation, we obtain the third model:

$$y_t^n = \theta_{1t} + \left(\frac{1-e^{-\lambda n}}{\lambda n}\right)\theta_{2t} + \left(\frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right)\theta_{3t} + e_{nt}\ , \text{ or} \tag{8}$$

$$y_t^n = \theta_{1t} + B_{2n}(\lambda)\theta_{2t} + B_{3n}(\lambda)\theta_{3t} + e_{nt}. \tag{9}$$

As before, the components describe geometric properties of the curve, this time with 3 components. Again, only empirically it is possible to evaluate the importance of the components and the compatibility between the number of stochastic sources and the number of components that must be used to describe the yield curve.

2.4 Affine No Arbitrage Model

The weights of the matrix B determine the relation between the yields of the different maturities and the short rate. The condition that there is no arbitrage at each instant among the rates implies in restrictions on the components of the matrix B. In particular, in the affine model, the risk premia λ_t and the short rate r_t are linearly dependent on the state variables θ_t.

Following Ang and Piazzesi (2003), we derive the pricing equation. The price at time t of an asset V_t that pays no dividend is

$$V_t = E^{\mathbb{Q}}[\exp(-r_t)V_{t+1}|\mathbb{F}_t]. \tag{10}$$

Under no arbitrage, there exists a martingale measure $\mathbb{Q} \sim \mathbb{P}$, the objective measure, and $\mathbb{F}_t$ is the filtration. The short rate and the risk premium are affine functions of the state vector $X_t \in \mathbb{R}^p$, that is, $r_t = \delta_0 + \delta_1 X_t$ and $\lambda_t = \lambda_0 + \lambda_1 X_t$, where the dynamics of the state vector is a multifactor vector autoregression

$$X_t = \mu + \phi X_{t-1} + \Sigma\epsilon_t. \tag{11}$$

Note that using a time-varying risk premium improves the adherence to data. Denote by ξ_t the Radon-Nikodym derivative $\frac{dQ}{dP} = \xi_t$. A discrete time "version" of Girsanov theorem is assumed setting $\xi_{t+1} = \xi_t \exp(-\frac{1}{2}\lambda_t \cdot \lambda_t - \lambda_t\epsilon_{t+1})$, where $\{\epsilon_t\}$ are independent normal errors. Then, the Pricing Kernel is $m_t = \exp(-r_t)\frac{\xi_{t+1}}{\xi_t}$, so that the price of a zero-coupon bond maturing n+1 periods ahead is $p_t^{n+1} = E[m_{t+1}p_{t+1}^n]$. Using induction, it can be seen that the price of bond will be exponential affine:

$$p_t^n = \exp(\alpha_n + \beta_n X_t), \tag{12}$$

where:

$$\begin{aligned} \alpha_1 &= -\delta_0,\ \beta_1 = -\delta_1, \\ \alpha_{n+1} &= -\delta_0 + \alpha_n + (\mu^\intercal - \lambda_0^\intercal \Sigma)\beta_n + \tfrac{1}{2}\beta_n^\intercal \Sigma^\intercal \Sigma \beta_n, \\ \beta_{n+1} &= -\delta_1 + (\phi - \lambda_1^\intercal \Sigma)\beta_n. \end{aligned} \tag{13}$$

Then $Y_t^n = -\log p_t^n/n = A_n + B_n X_t$, where $A_n = -\alpha_n/n$ and $B_n = -\beta_n/n$.Thus,

$$y_t^n = A_n + B_n X_t, \tag{14}$$

where A_n and B_n depend on parameters $\Psi = (\delta_0, \delta_1, \mu, \phi, \sigma, \mu^*, \phi^*, \Sigma)$, where $\mu^* = \mu - \Sigma\lambda_0$ and $\phi^* = \phi - \lambda_1^\intercal \Sigma$.

Here as in the CF model, the matrix B is estimated, which means that it must be identified. We use the same identifying restrictions as Dai and Singleton (2000).

3 Inference

The last section showed that the models only differ in the specification of the matrices A, B which relates the yield curve to the latent factors. In general, the model can be defined as

$$Y_t = A(\Psi) + B(\Psi)\theta_t + \sigma e_t,\ e_t \sim N(0, I_n) \tag{15}$$

$$\theta_t = \mu + \phi\theta_{t-1} + \Sigma u_t,\ u_t \sim N(0, I_p) \tag{16}$$

where $\Psi = (\mu, \phi, \sigma, \zeta, \theta)$ and the definition of ζ depend on the model and is summarized below.

CF	LP	NS	NA
$\zeta = (A, B)$	ζ is empty	$\zeta = \lambda$	$\zeta = (\delta_0, \delta_1, \mu^*, \phi^*, \Sigma)$

The likelihood $L(\Psi) = p(Y|\Psi) = p(Y|\theta, \Psi)p(\theta|\Psi)p(\Psi)$, where we assume a non informative prior $p(\Psi) = 1$, and

$$p(Y|\theta, \Psi) = \prod_t p(Y_t|\theta_t, \psi) = -1/2\left[T\sum_i \log(\sigma_i^2) + \sum_t\sum_i (u_{it}^2/\sigma_i^2)\right], \tag{17}$$

$$p(\theta|\Psi) = \prod_t p(\theta_t|\theta_{t-1}, \psi) = -1/2\left[T\sum_i \log(|\Sigma|) + \sum_t (e_t^\intercal (\Sigma^\intercal\Sigma)^{-1} e_t\right], \tag{18}$$

$$u_{it} = Y_{it} - A_i(\delta_0, \Sigma, \mu^*, \phi^*) - B_i(\delta_1, \phi^*)X_t, \quad e_t = \theta_t - \mu - \phi\theta_{t-1} \tag{19}$$

The distribution of the parameters,

$$p(\theta|Y, M_t, \psi) \propto p(Y|\theta, M_t, \Psi)p(\theta|M_t, \Psi)p(\Psi), \tag{20}$$

cannot be derived analytically, but the Clifford-Hammersley theorem guarantees that the recursive sampling of subsets of parameters, obtained from the complete conditional distributions, converges to the joint distribution. The subsets are chosen in a convenient way such that the subproblems have, when possible, analytical solutions and known complete conditional distributions, as is the case of subproblems 1-3 bellow. These problems correspond to, respectively, an estimation of a VAR model, the variance of known random variables, and the extraction of unobservable factors from a multivariate dynamic model. The subproblem (4) relative to ζ does not have known expression and its distribution will be derived utilizing the Metropolis-Hastings rejection method (Gamerman, 2001, and Johannes and Polson, 2003), with a proposal obtained from a normal distribution, centered on the value of the previous iteration, and with an arbitrarily fixed variance such that the acceptance rate is in the interval [0.3, 0.8]. The distributions calculated in each step of the algorithm are:

1. $(\mu^w, \phi^w) \sim p(\mu, \phi|\sigma^w, \zeta^w, \theta^w)$,
2. $\sigma^w \sim p(\sigma|\mu^w, \phi^w, \zeta^w, \theta^w)$,
3. $\theta^w \sim p(\theta|\mu^w, \phi^w, \zeta^w, \sigma^w)$,
4. $\zeta_i^w \sim p(\zeta_i|\zeta_{-i}^w, \mu, \phi, \Sigma, \sigma, \theta)$,

We have:

Subproblem1: $p(\mu, \phi|\sigma^w, \zeta^w, \theta^w) \sim N((X^\intercal X)^{-1}X^\intercal X^*, (X^\intercal X)^{-1} \otimes \Sigma)$, where $X = (\theta_1^w, ..., \theta_{T-1}^w)^\intercal$, $X^* = (\theta_2^w, ..., \theta_T^w)^\intercal$.

Subproblem2: $p(\sigma|\mu, \phi, \zeta, \theta) \sim \mathcal{IG}(diag(e^\intercal e))$, where $e = Y - A - BX$, and $\mathcal{IG}$ is the inverse gamma distribution.

Subproblem3: $p(\theta|\mu, \phi, \sigma, \zeta) = \prod_t p(\theta_t|\mu, \phi, \sigma, \zeta)$, where $p(\theta_t|\mu, \phi, \sigma, \zeta) = p(\theta_t|D_T) \sim N(h_t, H_t)$ is the FFBS algorithm defined in the Appendix.

The subproblems 1-3 are common to all models. The subproblem 4 depends on the definition of ζ. In the case of the LP there is no ζ, and so subproblem 4 is not defined.

In the case of the CF model, $\zeta = (A, B)$ is estimated without restrictions. Subproblem 4 becomes

$$(\zeta|\mu, \phi, \sigma, \theta) = (A, B|\mu, \phi, \sigma, \theta) = N((\theta^{w\intercal}\theta^w)^{-1}\theta^{w\intercal}Y, (\theta^{w\intercal}\theta^w)^{-1} \otimes \sigma^2). \quad (21)$$

In the models NA and NS, the parameter ζ do not have known conditional distribution. In this case, its distribution will be obtained through a rejection method - Metropolis-Hastings - where the proposal is sampled from a normal distribution centered on the value of the previous iteration, with arbitrarily fixed variance such that the acceptance ratio lies in the interval $[0.3, 0.8]$: $p(\zeta_i|\zeta_{i-1}, \mu, \phi, \sigma, \theta) \sim N(\xi_i^k, c)$ and accepts if $p(Y|\xi_i^k) - p(\theta|\xi_i^{k-1}) > u$, $u \sim U(0, 1)$.

3.1 Performance Criteria

The models under investigation have a different number of parameters, and hence they must be compared emphasizing forecasting performance or adherence to data. Gelfand and Ghosh (1998) proposed the minimum posterior predictive loss (PPL) criterion emphasizing forecasting performance. Spiegelhalter (2002) proposed the DIC criterion emphasizing adherence. Besides those measures, we will calculate Theil's U statistical measure, which consists of normalizing the MSE of out-of-sample forecasts and of in-sample adherence with respect to corresponding measures using random walks.

3.1.1 Minimum posterior predictive loss

For each point of the distribution of the estimators $\Psi^w \sim (\Psi|Y)$ there corresponds a forecasting for the yield curve $Y|\ \Psi^w$. Gelfand and Ghosh (1998) proposes a loss function penalizing the expected error $E(Y|\Psi^w) - Y$ and the variance of the forecasts $Y|\Psi^w - E(Y|\Psi^w)$. In our case, the target variable is multivariate, so that we take the mean of the expected losses calculated for each of the maturities. In other words, the criterion is:

$$\text{PPL} = \sum_i \sum_t \left(Y_t^i - E(Y_t^i|\Omega)\right)^2 + \sum_i \sum_t \frac{1}{N_w} \sum_w \left(E(Y_t^i|\Psi^w) - E(Y_t^i|\Omega)\right)^2, \tag{22}$$

3.1.2 Divergence of Information Criterion (DIC)

Spiegelhalter (2002) proposed a generalization of the AIC criterion based on the distribution of the divergence $D(\Psi) = -2\log L(\Psi)$:

$$\text{DIC} = E(D(\Psi)) - p_d = 2E(D(\Psi)) - D(E(\Psi)), \tag{23}$$

where $p_d = E(D(\Psi)) - D(E(\Psi))$ measures the equivalent number of parameters in the model, $E(D(\Psi))$ is the mean of the divergences taken in the posterior distribution of the estimators and $D(E(\Psi))$ is the divergence calculated at the mean point of the posterior distribution of the estimators.

Banerjee et al (2004) claims that LLP and DIC evaluate the fitting and penalize the degree of complexity of the models, but that the DIC takes into account the likelihood on the space of the parameters and PPL on the predictive space. Thus, when the main interest lies is forecasting, the PPL is to be preferred, whereas when the capacity of the model to explain the data is more interesting, DIC should be used.

3.1.3 Theil's U

When the processes under study have high persistency, the simplest representation, the random walk, frequently adheres to data. This is our case, in which hardly the yield curve suffers abrupt changes. Hence, a direct criterion to evaluate the results of the model is to compare it to the results of the random walk with the same set of information. Theil proposed to make this comparison taking the ratio between the standard deviation of the errors of the one-step forecasts and the standard deviation of the first difference of the variable. In our case, this result is calculated for each maturity, and when necessary it is summarized by the mean value along all maturities. The formula is:

$$\text{Theil-U} = \left(\frac{\sum_t (Y_t - \widehat{Y}_{t|t-21})^2}{\sum_t (Y_t - Y_{t-21})^2} \right)^{\frac{1}{2}}. \tag{24}$$

4 Results

The performance of the models is evaluated in 3 markets having distinct features. The first one is the `DIxPRE` Swap contract traded in the Brazilian futures market (BM&F) for many maturities, and is used as an approximation to the term structure of public bonds traded in the domestic market. The second one is the market of Sovereign bonds issued by the Brazilian Treasury. We use data treated by Bloomberg which provides constant maturity zero-coupon bonds. It is smaller and less liquid than the domestic government bonds market, and, being a sovereign bond depends more directly on the effects of fluctuations of the perception of risk that agents have about the capacity of the Brazilian government to honour those bonds, that is, on credit risk. Finally, the third market is that of the US Treasury bonds. The Federal Reserve provides free of coupon and constant maturity rates.

The features of the interest rates market and the availability of data motivated our choice of analyzing the markets with daily frequency. This was particularly important for the Brazilian domestic market, which, like other emerging markets, has specific characteristics. It is conditioned to the credit risk of the public debt, to the higher volatility of the rates due to macroeconomic instability, to the vulnerability of the exchange rate, and finally, to interventions of the monetary authorities. In January 1999, Brazil adopted the floating exchange rate regime, which changed the mechanism of the formation of the domestic and foreign interest rates. Consequently, the sample used in the estimations was [01/1999, 09/2005]. The lower temporal dimension was partly compensated with the use of daily data.

In the Brazilian sovereign market, the data is available starting in March 1998 and ending in July 2005. For convenience, we analyze the US Treasury for the same period.

The degree of linear dependence among the maturities of the 3 markets was measured calculating the proportion of the variance that is explained by the first 3 principal components of the correlation matrix. The table below suggests that the markets have at most 3 sources of independent stochastic variance.

	First	Second	Third
Swap	90.5	9.0	0.5
Sovereign	95.6	3.6	0.3
FED	95.9	3.6	0.5

In the NA model, the time has 2 dimensions, the historical time in which data is collected, and the maturity time of the yield curve. The periodicity of the former, τ_1, is daily in our exercise, while of the latter, τ_2, which gives the interval between maturities, is monthly. It is necessary to adjust the variance of the innovations for the difference in the periodicities[1]. Note that the transition equations for the $\mathbb{P}$-measure and $\mathbb{Q}$-measure are defined independently.

The forecasting horizon of interest not necessarily coincides with the periodicity of the data. When we are interested in the forecasting horizon $h > \tau_1$, which is $Y_{t|t-h} = \phi^h Y_{t-h} = \phi^* Y_{t-h}$, the model can be estimated maximizing the likelihood of the h-steps forecasting. In the case of a first order autoregression, this is done choosing a lag of size h. The results for different values of h are not necessarily equal, for in each case the likelihood is specific for the horizon of interest.

The previous table suggests that the 3 sets of data have at least 2 stochastic sources, and so the models will have two latent factors. The tables 1-3 show the performance of the 4 models, all estimated for the forecasting horizon of 21 days, the mean number of commercial days in a month, according to 3 proposed criteria. In case of the models that do not have no arbitrage restrictions, versions with 1 and 3 factors were estimated to evaluate the sensibility of the results to this parameter. In the case of 1 factor, the NS and LP are identical, and so the results are presented jointly.

Note that for the DIC, PPL and Theil-U, the lower the value the better.

Table1. Forecasting performance in the Swap market.

#F	1		2				3		
	CF	NS/LP	NA	CF	NS	LP	CF	NS	LP
DIC	-30.0	-28.6	-38.2	-43.7	-41.4	-41.6	-53.6	-22.7	-51.0
PPL	0.190	0.333	0.079	0.060	0.215	0.087	0.053	8.367	0.083

Table2. Forecasting performance in the Brazilian sovereign bonds market.

#F	1		2				3		
	CF	NS/LP	NA	CF	NS	LP	CF	NS	LP
DIC	-65.0	-63.3	-77.6	-86.4	-85.1	-82.1	-101.6	-95.9	-74.0
PPL	0.175	0.200	0.133	0.122	0.131	0.180	0.120	0.140	1.759

Table3. Forecasting performance in the US Treasury market.

[1] When the periodicity of the innovations is τ_1 and the no arbitrage equations are defined for the periodicity τ_2 then if $V = \Sigma\Sigma^\intercal$, $k = int(\tau_2/\tau_1)$, then $V_{\tau_2} = V_{\tau_1} + \phi^\intercal V \phi + \phi^{2\intercal} V \phi^2 + ... + \phi^{k\intercal} V \phi^k$.

#F	1		2				3		
	CF	NS/LP	NA	CF	NS	LP	CF	NS	LP
DIC	-84.1	-77.4	-102	-110	-104	-100	-128	-118	-122
PPL	0.60	2.50	0.27	0.25	0.33	0.79	0.21	0.44	0.25

Tables 1-3 show that:

- The performance of the CF model is always superior to the other models, for the any number of factors and for the 3 sets of data.
- The models with 3 factors are in general better than those with 2 factors, but the improvement is much less when comparing with the 1 factor models. This suggests that in the 3 cases there are at least 2 factors, and that in some cases a third may be possible to estimate.
- The performance of the NA model is worse than the CF according to 2 criteria for the 3 markets. But is similar to the best model when we consider the PPL criterion.
- The NS is better than the LP or not depending on the market and on the criterion.

The comparison between the mean squared error (MSE) of the model and of the random walk (rw), the Theil's U statistic, is a measure of practical utility. Models with errors greater than that of the random walk are useless for forecasting. This error depends on the forecasting horizon. In mean-reversion processes, the error of the rw tends to grow faster than that of the model with respect to greater horizons. Tables 4-5 present results of the model CF with 2 factors and for 3 markets: 1) the PPL criterion, 2) the mean of the MSE for the forecasting errors of the model for all maturities in the indicated horizon, 3) the mean of Theil's U statistics computed in-sample, and finally 4) the value of the out-of-sample Theil-U for all maturities (last 30 observations).

Values of out-of-sample Theil-U's less than 1 are highlighted because it means the model presented better forecasting performance than the rw.

Table 4. Effect of the forecasting horizon. Swaps.

lags	10*PPL	10*MSE	In-sample TU
1	0.09	0.03	1.41
5	0.21	0.09	1.06
10	0.33	0.18	0.99
21	0.51	0.31	0.94

	Out-of-sample Theil U								
lags	1	2	3	6	9	12	18	24	36
1	4.52	1.62	**0.99**	1.58	1.62	1.82	1.51	1.01	3.93
5	1.51	**0.96**	**0.93**	1.08	**0.99**	1.03	1.04	1.08	2.20
10	**0.81**	**0.70**	**0.75**	1.11	**0.96**	**0.91**	**0.97**	1.13	1.84
21	**0.52**	**0.59**	**0.58**	**0.94**	1.14	1.06	1.11	1.40	2.32

The results indicate that:

- When the horizon is 20 times greater, the MSE rises 10 times, less than what would happen with a rw.
- For horizons above 5 days the predictions of the model are better than those of the random walk.

Table 5. Effect of the forecasting horizon. Brazilian Sovereign bonds.

lags	10*PPL	10*MSE	In-Sample TU
1	0.22	0.06	0.99
5	0.28	0.19	1.08
10	0.70	0.31	1.03
21	1.15	0.53	0.99

	Out-of-sample Theil U								
lags	1	6	12	24	36	60	84	120	240
1	1.36	**0.81**	**0.62**	**0.99**	2.53	4.40	4.03	5.73	5.77
5	1.31	**0.83**	1.57	**0.91**	1.01	1.80	1.31	5.74	6.23
10	1.16	**0.70**	1.04	**0.66**	1.10	2.20	1.77	5.60	6.17
21	1.36	**0.81**	**0.62**	**0.99**	2.53	4.40	4.03	5.73	5.77

The results show that:

- When horizon is 20 times greater, the MSE is 9 times greater.
- Better forecasts are not linked to the horizon, but are linked to maturities [6,24] months.

Table 6. Effect of the forecasting horizon. US Treasury.

lags	100*PPL	100*MSE	In-Sample TU
1	0.075	0.015	1.03
5	0.125	0.030	1.10
10	0.168	0.048	1.21
21	0.251	0.086	1.85

	Out-of-sample Theil U								
lags	1	6	12	24	36	60	84	120	240
1	**0.60**	1.20	2.20	4.04	3.49	1.70	**0.94**	1.51	4.20
5	**0.51**	1.12	2.76	4.02	3.23	1.78	**0.85**	1.01	3.57
10	1.49	1.09	3.84	5.21	4.22	2.17	**0.92**	1.12	4.08
21	5.87	**0.98**	7.06	11.89	9.81	4.44	**0.98**	2.02	9.61

The results show that the model presents a superior forecasting performance with respect to the random walk only for the short rate and for horizons up to 5 days.

5 Conclusion

We analyzed with daily data 3 yield curves, the Brazilian domestic market interest rate swaps, the Brazilian sovereign bonds and US Treasury bonds, equipped with 4 models: the common factor model of the time series literature, the affine no arbitrage model of the Finance literature, and 2 models that geometrically decompose the yield curve, Nelson-Siegel and Legendre polynomials, modified to include dynamic effects of the latent components.

It resulted that the common factor model, in spite of having a much greater number of parameters, had better performance according to two criteria, the posterior predictive loss and DIC, related to the predictive and explanatory power of the model, respectively. Also, the affine model presented inferior but comparable results. This may be attributed to the complexity of estimation of the risk premia.

The common factor model was used to evaluate the effect of the forecasting horizon on the forecasting performance in the 3 markets. Depending on the market, the model tend to have better results compared to the random walk for longer horizons.

An immediate extension of this work is the incorporation of macro state variables as Ang and Piazzesi (2003), and evaluate the predictive performance of the models with and without no arbitrage.

6 References

1. Almeida CIR, Duarte AM, Fernandes CA. Decomposing and Simulating the Movements of Term Structures in Emerging Eurobonds Markets, Journal of Fixed Income 1998; 1, 21-31.

2. Ang A, Piazzesi M. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics 2003; 50, 745-787.

3. Banerjee S, Carlin B, Gelfand A. Hierarchical Modeling and Analysis. for Spatial Data. CRC Press/Chapman Hall; 2004.

4. Dai Q, Singleton K. Specification analysis of term structure of interest rates. Journal of Finance 2000; 55, 1943-78.

5. Diebold FX, Li C. Forecasting the term structure of government bond yields. Journal of Econometrics 2006; 130, 337-364.

6. Duffie D, Kan R. A yield-factor model of interest rates. Mathematical Finance 1996; 6, 379-406.

7. Gamerman D. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman & Hall: London; 1997.

8. Gelfand A, Ghosh S. Model choice: a minimum posteriori predictive loss approach. Biometrika 1998; 85 1-11.

9. Harvey AC, Forecasting, structural time series models and the Kalman filter. Cambridge University Press: Cambridge; 1989.

10. Johannes M, Polson N. MCMC methods for continuous-time financial econometrics. Working paper 2003.

11. Litterman R, Scheinkman J. Common factors affecting bond returns. Journal Fixed Income 1991; 1, 51-61.

12. Nelson CR, Siegel AF, Parsimonious modeling of yield curves, Journal of Business 1987, Vol. 60, No. 4, 473-89.

13. Rudebusch G, Wu T. A macro-finance model of the term structure, monetary policy and the economy. Working Paper 2003; Federal Reserve Bank of San Francisco.

14. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian Measures of Model Complexity and fit (with discussion). Journal of the Royal Statistical Society Series B 2002, 64, 583–639.

15. West M, Harrison PJ. Bayesian forecasting and dynamic models, Springer:New York; 1997.

A Appendix: Kalman Filter and FFBS algorithm

We present the Kalman Filter and the FFBS algorithm of the Dynamic Linear Model (DLM) in which part of the components are observed (M). Defining

$$\begin{array}{l} Y_t = A + B_M M_t + B_\theta \theta_t + e_t,\ e_t \sim N(0, \mathbb{I}_\sigma), \\ M_t = \mu_M + \phi_{MM} M_{t-1} + \phi_{M\theta} \theta_{t-1} + u_t^M, \\ \theta_t = \mu_\theta + \phi_{\theta M} M_{t-1} + \phi_{\theta\theta} \theta_{t-1} + u_t^\theta, \end{array} \tag{25}$$

we obtain the linear dynamic model

$$\begin{array}{l} Y_t = y_t + F\theta_t + e_t,\ e_t \sim N(0, I\sigma), \\ \theta_t = x_t + G\theta_t + u_t,\ u_t \sim N(0, W), \\ \text{where } y_t = A + B_M M_t, \\ x_t = \mu_\theta + \phi_{\theta M} M_{t-1}, \\ F = B_\theta, G = \phi_{\theta\theta}. \end{array} \tag{26}$$

that can be estimated as follows:

$$
\begin{array}{l}
\text{Given: } (\theta_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1}). \\
\text{Prior: } (\theta_t|D_{t-1}) \sim N(a_t, R_t), \\
\text{where } a_t = Gm_{t-1} \quad R_t = GC_{t-1}G^T + W. \\
\text{Forecast: } (Y_t|D_{t-1}) \sim N(f_t, Q_t), \\
\text{where } f_t = Fa_t \quad Q_t = FR_tF^T + \sigma. \\
\text{Posteriori: } (\theta_t|D_t) \sim N(m_t, C_t), \\
\text{where } m_t = a_t + A_t(Y_t - f_t), C_t = R_t - A_tQ_tA_t^T, A_t = R_tFQ_t^{-1}.
\end{array}
\tag{27}
$$

Once the conditional distribution of $(\theta_t|D_t)$ $t = 1..T$ is obtained, the FFBS algorithm permits one to obtain a sample of $(\theta_t|D_T)$.

$$
\begin{array}{l}
\text{Given: } (\theta_T|D_T) \sim N(m_T, C_T). \\
(\theta_t|\theta_{t+1}) \sim N(h_t, H_t), \\
\text{where } h_t = m_t + B_t(\theta_{t+1} - a_{t+1}) \quad H_t = C_t - B_tR_{t+1}B_t^T \quad B_t = C_tG^TR_{t+1}^{-1}.
\end{array}
\tag{28}
$$

EDITORIAL

Coordenação
Iranilde Rego

Supervisão
Marcos Hecksher

Revisão
Lucia Duarte Moreira
Alejandro Sainz de Vicuña
Eliezer Moreira
Elisabete de Carvalho Soares
Míriam Nunes da Fonseca
Tamara Sender

Editoração
Roberto das Chagas Campos
Alessandra Cerqueira Mattos
Camila Guimarães Simas
Carlos Henrique Santos Vianna
Leandro Daniel Ingelmo (estagiário)

Brasília
SBS – Quadra 1 – Bloco J – Ed. BNDES,
9º andar – 70076-900 – Brasília – DF
Fone: (61) 3315-5090
Fax: (61) 3315-5314
Correio eletrônico: editbsb@ipea.gov.br

Rio de Janeiro
Av. Nilo Peçanha, 50, 6º andar — Grupo 609
20044-900 – Rio de Janeiro – RJ
Fone: (21) 3515-8433 – 3515-8426
Fax (21) 3515-8402
Correio eletrônico: editrj@ipea.gov.br

COMITÊ EDITORIAL

Secretário-Executivo
Marco Aurélio Dias Pires

SBS – Quadra 1 – Bloco J – Ed. BNDES,
9º andar – sala 908
70076-900 – Brasília – DF
Fone: (61) 3315-5406
Correio eletrônico: madp@ipea.gov.br

Tiragem: 136 exemplares