CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

Economics Bulletin

Volume 31, Issue 3

Bayesian Factor Selection in Dynamic Term Structure Models

Marcio Laurini
Ibmec Business School

Abstract

This paper discusses Bayesian procedures for factor selection in dynamic term structure models through simulation methods based on Markov Chain Monte Carlo. The number of factors, besides influencing the fitting and prediction of observed yields, is also relevant to features such as the imposition of no-arbitrage conditions. We present a methodology for selecting the best specification in the Nelson-Siegel class of models using Reversible Jump MCMC.

The author wishes to acknowledge the valuable comments of an anonymous referee.
Citation: Marcio Laurini, (2011) "Bayesian Factor Selection in Dynamic Term Structure Models", *Economics Bulletin*, Vol. 31 no.3 pp. 2167-2176.
Submitted: Apr 17 2011. **Published:** July 25, 2011.

1 Introduction and Methodology

Selecting the appropriate specification in econometric models is fundamental to guarantee the expected econometric properties, such as consistency, absence of bias and efficiency. The functional form and variables included in the specification chosen by the researcher also has important practical consequences, such as the predictive performance of the model, and in models used in financial applications, the estimated specification can affect the asset prices derived from the model.

In the analysis of dynamic econometric models for the term structure of interest rates, the functional form chosen, besides the influence on the properties of estimators and forecasts, may also have important consequences on the validity of no-arbitrage conditions. The functional form chosen is directly associated with the consistency with no-arbitrage in the sense defined in Filipovic (1999) and Filipovic (2001). In the widely used Nelson-Siegel family (Nelson and Siegel (1987)) of models is possible to show that original Nelson-Siegel specification is not consistent with no-arbitrage, but the addition of an additional slope factor, leading to the model proposed by Svensson (1994), allows a specification consistent with no-arbitrage (although limited in the empirical fit), and the overall consistency is achieved by an additional latent factor, leading to the five factor model of Björk and Christensen (1999).

These conditions were then reformulated in the affine formulation for the Nelson-Siegel family proposed in Christensen et al. (2009), Christensen et al. (2010) and Joslin et al. (2011). In this formulation consistence with no-arbitrage is obtained when the model contains three or five factors, but not in other specifications. The validity of no-arbitrage conditions is related to the matching of each slope factor with a associated curvature. This way, the imposition of no-arbitrage conditions depends on the validity of appropriate specification, related to the number of factors chosen. It is also important to note that the arbitrage free formulation of Christensen et al. (2009) and Christensen et al. (2010), consistence with the no-arbitrage is obtained by adding an correction factor in the Nelson-Siegel family. For example in the formulation of the model with five factors the arbitrage free curve is given by:

$$y(t,T) = X_t^1 + \frac{1-e^{-\lambda_1(T-t)}}{\lambda_1(T-t)}X_t^2 + \frac{1-e^{-\lambda_2(T-t)}}{\lambda_2(T-t)}X_t^3 + \tag{1}$$

$$\left[\frac{1-e^{-\lambda_1(T-t)}}{\lambda_1(T-t)} - e^{-\lambda_1(T-t)}\right]X_t^4 + \left[\frac{1-e^{-\lambda_2(T-t)}}{\lambda_2(T-t)} - e^{-\lambda_2(T-t)}\right]X_t^5 - \frac{C(t,T)}{T-t}$$

with $y(t,T)$ denoting the observed yields, $(T-t)$ the time to maturity, $X_t^1, X_t^2, X_t^3, X_t^4, X_t^5$ the five latent factors and λ_1, λ_2 the decay factors of the Nelson-Siegel model. In this specification correction for no-arbitrage is given by the term $C(t,T)$, which is basically a function[1] of the parameters λ_1 and λ_2. Because of this term $C(t,T)$ the procedure of

[1]The analytical form of this term is quite extensive, and therefore omitted from the article, but can be found in Christensen et al. (2009) and Christensen et al. (2010).

model selection is crucial in the analysis of consistency with no-arbitrage. For example the model with three factors (including X_t^1, X_t^1 and X_t^4) is arbitrage-free, but the models with only one factor (X_t^1) and four factors (X_t^1, X_t^2, X_t^4 and X_t^5) are not consistent with no-arbitrage.

Note that the usual procedures of model selection are hampered by the nonlinear formulation that contains common parameters in the definition of the latent factors. Each decay parameter λ_i affects both the associated slope and curvature factors. Besides the trivial case of misspecification raised by the omission of a significant latent factor, note that the factor analysis with a redundant (e.g. correct model contains four factors and the estimated model contains five factors), due to the nonlinearity of the model, consistency of the estimation is no longer valid, as it would in the specification of a linear regression model. This problem hinders the use of traditional methods of model selection such as information criteria, which are only consistent (BIC by example, e.g. Claeskens and Hjort (2008)) in the selection of linear models, property which is not generally valid for nonlinear models. Another difficulty in using information criteria is that the validity of these procedures is only asymptotic and finite sample properties may not be optimal.

Because of these difficulties, this paper proposes the use of Bayesian methods based on Reversible Jump Markov Chain Monte Carlo (RJMCMC). This methodology generalizes the sampling framework of Markov Chain Monte Carlo (e.g. Gamerman and Lopes (2006), Robert and Casella (2005)) to perform the estimation of models whose dimension is variable. In this situation we can adequately address the problem of selection of variables and models, and perform procedures such as model averaging. A Bayesian model of variable dimension[2] can be defined as an indexed collection of models:

$$\mathcal{M}_k = \{p(\bullet|\theta_k); \theta_k \in \Theta_k\} \tag{2}$$

associated with a collection of priors on the parameters of these models, denoted by $p_k(\theta_k)$. It is also necessary define a prior on the index of the model - $\varrho(k), k = 1, \ldots, K$. We unified the notation for the structure of priors as $p_k(k, \theta_k) = \varrho(k)p_k(\theta_k)$. This is a proper density defined with respect to Lebesgue measure on the union of the spaces $\Theta = \cup_k \{k\} \times \Theta_k$. With this structure we can obtain the posterior distribution of parameters via Bayes' theorem:

$$p(\mathcal{M}_i|x) = \frac{\varrho(i) \int_{\Theta_i} p_i(x|\theta_i)p_i(\theta_i)d\theta_i}{\sum_j \varrho(j) \int_{\Theta_j} p_j(x|\theta_j)p_j(\theta_j)d\theta_j}. \tag{3}$$

We can choose the model with the highest $p(\mathcal{M}_i|x)$, or then perform a procedure of model averaging:

$$\sum_j p(\mathcal{M}_i|x) \int_{\Theta_j} p_j(x|\theta_j)p_j(\theta_j)d\theta_j \tag{4}$$

[2]We follow the definition adopted in Robert and Casella (2005).

with x denoting the observed sample and the above equation as a predictive density. Although this structure is a Bayesian complete model, there are several computational difficulties. Methods as Gibbs sampling can not be directly used in this context, since it is necessary to define how transitions between models $\mathcal{M}_i$ will be made[3].

The methodology known as Reversible Jump MCMC, proposed by Green (1995), is based on the determination of a reversible transition kernel between models, i.e., a kernel satisfying $\int_A \int_B K(x, dy)p(x)dx = \int_A \int_B K(y, dx)p(y)dy$ for $A, B \subset \theta$ and p representing the invariant density of the model. The kernel represents the transition between different models using the following decomposition:

$$K(x, B) = \sum_m \int_B \rho_m(x, y')q_m(x, dy') + \omega(x)\mathbb{I}_B(x) \tag{5}$$

with q_m denoting the transition function for the model $\mathcal{M}_m$, with transition probabilities ρ_m. We define the probability of not transition as $\omega(x) = 1 - \sum_m (q_m, p_m)(x, \Theta_m)$. Under some technical conditions we can define the probability of transition inside the usual Metropolis-Hastings procedure:

$$\rho_m(x, y) = min\left\{1, \frac{g_m(y, x)}{g_m(x, y)}\right\} \tag{6}$$

with $g_m(.)$ denoting the instrumental density.

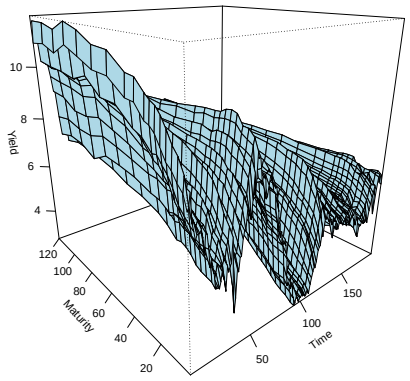
The algorithm proposed by Green (1995) solves the computational problem of the change (jump) between the different models considered through supplementation of the spaces Θ_{k_1} and Θ_{k_2} with artificial spaces, creating a bijection between the models. For example, this solves the problem of distinct parameter's dimension between each model considered, and also allows to create the necessary Jacobians of these transformations, by the use of instrumental models with a fixed dimension, usually taken by the size of the largest model considered. The details of computational implementation used in this work can be found in Lunn et al. (2009), which discusses the necessary changes in Gibbs and Metropolis-Hastings sampling procedures to implement the Reversible Jump MCMC of Green (1995).

To implement the Reversible Jump MCMC procedure for model selection in the family of Nelson-Siegel models, we perform some modifications on the procedures proposed in Green (1995) and Lunn et al. (2009), using some special features of this family of models. Note that conditional to the decay parameters λ_i, the specification given by equation 1 is a linear model. The second modification is the possibility of to analyze the average behavior of the relevant factors, thus avoiding direct estimation of latent states, as in Diebold and Li (2006), Christensen et al. (2009) and Christensen et al. (2010). This simplification allows performing model selection through the methodology of Reversible Jump MCMC developed for selection of linear regression models.

In first stage of the estimation of each sub-model, we estimate the parameters λ_i for each proposed model by Metropolis-Hastings sampling. Conditional to the estimation these

[3]See Robert and Casella (2005) for a discussion of problems involved.

Figure 1: Fama-Bliss Database



parameters, the model becomes conditionally linear and we can perform the Reversible Jump algorithm to the estimation of loadings $\bar{X}^1, \bar{X}^2, \bar{X}^3, \bar{X}^4$ and $\bar{X}^5$. The selection of these loadings is performed assuming that these parameters are constant throughout the sample, and so we analyze the influence of each factor by the average effect. Note that this assumption is not restrictive, being based only on the assumption of existence of a an invariant distribution for these factors.

Another exploited property is to assume that the $\bar{X}^1$ is present in all specifications. Because this factor has a level interpretation in the yield curve, it is always present in all specifications, and thus the model selection procedure may be restricted to the combinations between factors $\bar{X}^2, \bar{X}^3, \bar{X}^4, \bar{X}^5$, reducing the space of models analyzed. Another important point is that given the interpretation of slope and curvature of the other latent factors, they are approximately orthogonal, which avoids the transformation of regressors to orthogonal coordinates, as discussed in Robert and Casella (2005).

2 Database and Results

This study examines the same database of Diebold and Li (2006), Christensen et al. (2009) and Christensen et al. (2010). The base is composed of unsmoothed forward rates of U.S. Treasuries obtained by the procedure of Fama and Bliss (1987), with fixed maturities of 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, using monthly observations for the period 1985:01 - 2000:12, containing 192 observations of this curve. Figure 1 shows the database used. Descriptive statistics and other properties can be found in Diebold and Li (2006).

To complete the Bayesian specification, we define the set of priors for the parameters of interest. We assume that the priors for the decay parameters are $\lambda_1 \sim logNormal(-2.337901, 1)$,

Table 1: Posterior Model Probabilities $p(\mathcal{M}_i|x)$

model structure	posterior prob.	cumulative prob.	log marginal likelihood
11010	9.0901E-5	0.00009	827.568
11011	9.5454E-4	0.00104	1144.065
10110	0.001909	0.01622	1326.232
10111	0.013272	0.002954	1336.614
11100	0.032681	0.048909	703.9478
11101	0.139181	0.188090	1260.590
11110	0.156227	0.344318	1169.585
11111	0.655681	1.000000	1350.823

$\lambda_2 \sim logNormal(-0.1768565, 1)$, consistent with the values estimated in Diebold and Li (2006), Christensen et al. (2009) and Christensen et al. (2010). For the factors $\bar{X}^2, \bar{X}^3, \bar{X}^4, \bar{X}^5$ we assume a multivariate gamma-normal distribution $p(\tau, \bar{X}|k) = Ga(\tau|a, b) \times MVN(\bar{X}|\mu, \tau^{-1}I_k)$ where $\mu = 0$ and I_k is an identity matrix of dimension k, with the dimension of the sub-model being evaluated, which is equivalent to imposing a same precision for all elements of the multivariate normal distribution. In this problem we assume a diffuse prior precision $\tau^{-1} = .0001$. The hyperparameters a and b of Gamma prior are given by the .75 and .75. The last relevant prior is that defines the number of factors included in each specification. For the reported results, we use $\varrho(k) \sim Binomial(.6, 4)$.

Regarding the set of priors used, we conducted some sensitivity analysis. The results basically remain unchanged in relation to the priors for the $\bar{X}^i$, precisions and decay parameters λ_i. The posterior probabilities of each sub-model are influenced by the prior value for the number of factors, but the overall results do not change relevantly, since even with a low probability for the number of factors ($\varrho(k)$=.2), the posterior probabilities of all the latent factors are still relevant, with minimum probability of .24 for each term, thus further justifying the inclusion of all factors. To obtain the posterior distributions, we run the MCMC procedure using 50,000 initial iterations (burn-in), and over 500,000 simulations to construct the posterior distributions.

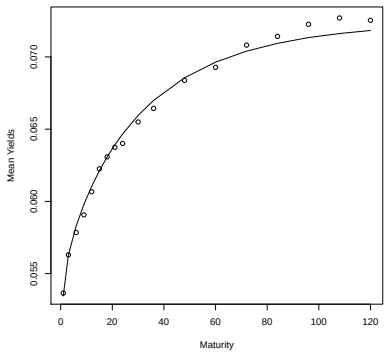
The posterior probabilities of each analyzed model are placed in Table 1. The column model structure shows the factors included in each specification. One denotes that the latent factor X^i is included, and zero that the factor is not included in the specification. As an example, in the first line we have the posterior probability of the model with level, slope and curvature factors, according to the ordinance adopted in equation 1, corresponding to the original Nelson-Siegel model. In the next column we find cumulative probability of all models analyzed. We also show in this table the values of the marginal log-likelihood for each sub-model, calculated using the modified harmonic mean approach of Raftery et al. (2007).

The results support a complete specification for this dataset, indicating that the model with five factors obtains the highest posterior probability among all estimated specifications. The marginal log-likelihood, as expected, favors the model with five latent factors,

Table 2: Marginal Probabilities

Factor	marginal prob.
2	0.9850
3	0.9957
4	0.8213
5	0.8209

Figure 2: Model Fit - Bayesian Model Averaging



but has a different result in relation to sub-models with three and four latent factors, favoring the sub-models 10110 and 10111, while the RJMCMC methodology indicates a high posterior probability to sub-model 11110, a result more consistent with the practical applications of these methods, that always include the first slope factor in the model.

Another way to analyze this result is through the marginal probabilities of each model, shown in Table 2. We obtain high marginal probabilities for all latent factors in the Nelson-Siegel family, supporting the use of models with five factors proposed in Björk and Christensen (1999) and Christensen et al. (2009), and consistent with no-arbitrage, an important financial result. Also we can see evidence in favor of a greater number of factors in the posterior distribution of parameter $\varrho(k)$, which indicates the number of factors included in each model. The posterior distribution of parameter $\varrho(k)$ has posterior mean of 3.67, supporting the largest number of factors.

To illustrate the application of Reversible Jump MCMC methodology as a Bayesian model averaging procedure, we illustrate the model fit obtained by weighting all models analyzed by the estimated posterior probabilities, according to equation 4. Figure 2 show the average yields over time (marked with circles in figure), compared to the fit obtained by model averaging, in the continuous line. The results shows that weighted model provides an excellent fit for the average yields observed in this sample.

3 Conclusions

In this paper we introduce the use of Reversible Jump Markov Chain Monte Carlo

methods for model and factor selection in dynamic term structure models. From the special structure of Nelson-Siegel family, it is possible to adapt procedures used in the selection of linear models to explore all possible specifications derived from this family of models.

These specifications have important consequences on fitting, prediction and consistency with no-arbitrage, and the Reversible Jump MCMC procedure allows overcome the limitations in the usual procedures for model selection, such as only asymptotic consistency for linear models.

Other methodologies that could be used in this context are the procedures for calculating posterior probabilities of models using Gibbs Sampling, as Carlin and Chib (1995), Condgon (2007) and Dellaportas et al. (2002), and the Stochastic Search algorithm of George and McCulloch (1993), with the observation that these procedures are intended for use in generalized linear models, while the specifications analyzed in this article are formulated as nonlinear regressions.

Also important in this context are the algorithms for selection of the number of factors proposed by Lopes and West (2004) and Frühwirth-Schnatter and Lopes (2010) in traditional factor analysis, and especially the work of Frühwirth-Schnatter and Wagner (2010), which discusses procedures for selection of factors in state-space models. In particular the formulation of a regression model with time-varying parameters, which is analogous to the dynamic Nelson-Siegel model, is pointed in Frühwirth-Schnatter and Wagner (2010) as a possible development of this methodology. A comparative analysis with these methods and other procedures of Bayesian model selection and averaging is a promising topic for future study, particularly in relation to the procedure of Frühwirth-Schnatter and Wagner (2010), which allows the estimation of the full vector of latent states.

References

Björk, T. and B. J. Christensen (1999). Interest rate dynamics and consistent forward rate dynamics. *Mathematical Finance 9*, 323–348.

Carlin, B. and S. Chib (1995). Bayesian model choice via markov chain monte carlo methods. *Journal of Royal Statistical Association Series B 157*, 473–484.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2009). An arbitrage-free generalized nelson-siegel term structure model. *Econometrics Journal 12*, 33–64.

Christensen, J. H., F. X. Diebold, and G. D. Rudebusch (2010). The affine arbitrage free class of nelson-siegel term structure models. *Journal of Econometrics forthcoming*.

Claeskens, C. and N. L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge University Press.

Condgon, P. (2007). Model weights for model choice and averaging. *Statistical Methodology 4*, 143–157.

Dellaportas, P., J. Foster, and I. Ntzoufras (2002). On bayesian model and variable selection using mcmc. *Statistics and Computing 12*, 27–36.

Diebold, F. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal Of Econometrics 130*, 337–364.

Fama, E. and R. R. Bliss (1987, September). The information in long-maturity forward rates. *American Economic Review 77*, 680–691.

Filipovic, D. (1999). A note on the Nelson-Siegel family. *Mathematical Finance 9(4)*, 349–359.

Filipovic, D. (2001). *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*. Springer-Verlag.

Frühwirth-Schnatter, S. and H. F. Lopes (2010). Parsimonious bayesian factor analysis when the number of factors is unknown. Unpublished Working Paper, Booth Business.

Frühwirth-Schnatter, S. and H. Wagner (2010). Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics 154*(1), 85–100.

Gamerman, D. and H. Lopes (2006). *Stochastic Simulation for Bayesian Inference*. Chapman and Hall.

George, E. and R. McCulloch (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association 88*, 881–889.

Green, P. (1995). Reversible jump mcmc computation and bayesian model determination. *Biometrika 82*(4), 711–732.

Joslin, S., K. Singleton, and H. Zhu (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies, Forthcoming*.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 41?67.

Lunn, D. J., N. Best, and J. C. Whittaker (2009, December). Generic reversible jump mcmc using graphical models. *Statistics and Computing 19*, 395–408.

Nelson, C. R. and A. F. Siegel (1987). Parsimonous modelling of yield curves. *Journal of Business 60*(4), 473–489.

Raftery, A., M. Newton, J. Satagopan, and P. Krivostcky (2007). *Bayesian Statistics 8*, Chapter Estimating the Integrated likelihood via posterior model simulation using the harmonic mean identity (with discussion), pp. 1–45.

Robert, C. P. and G. Casella (2005). *Monte Carlo Statistical Methods*. Springer.

Svensson, L. E. O. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. *NBER Working Paper* (4871).