Nelson and Siegel, no–arbitrage and risk premium*

François Le Grand†

February 16, 2008

Abstract

In this article, we plug no-arbitrage constraints into the standard Nelson and Siegel model. These no-arbitrage constraints do not impair the tractability and the parsimony, which have made the standard Nelson and Siegel so popular. The resulting model outperforms significantly the standard model on the relevant aspects: (i) the fit of the yield curve, (ii) the rejection of the expectation hypothesis and (iii) out-of-sample forecasts. This also produces better results in portfolio management. We illustrate this in a simple mean-variance framework: One-month returns with our model are more than 2 points greater than with standard Nelson and Siegel.

1 Introduction

Nelson and Siegel (1987) models are popular amongst financial experts, notably asset and debt managers or central bankers. As reported in a recent article of the Bank for International Settlements (BIS Monetary and Economic Department 2005), most of central banks use Nelson and Siegel approaches. Two reasons explain this popularity: Its empirical performances and its simplicity. First, the Nelson and Siegel model performs well in fitting the yield curve as well as in out-of-sample forecasting. The replication of historical data is a central feature for models in general and the Nelson and Siegel representation does it well. Concerning out-of-sample forecasts, few models perform as well as Nelson and Siegel (Diebold and Li 2006). Second, Nelson and Siegel is a parsimonious and tractable framework, since the zero coupon simply expresses yield as the affine transformation of a given 3 dimensional

*__Acknowledgments:__ I am specially grateful to Gabrielle Demange and Laurence Boone for helpful discussions. I am also indebted to Andras Fulop, Michel Juillard, Guy Laroque for valuable suggestions. I also thank Nicolas Cœurdacier, Stéphane Guibaud, and the participants at the Federation PSE seminar, at the ADRES seminar and at the 12th International Conference on Computing in Economics and Finance for their comments.

†Paris School of Economics and Barclays Capital, Paris. Email: `legrand@pse.ens.fr` and `francois.legrand@barclayscapital.com`. Postal Address: PSE, 48 boulevard Jourdan, 75014 Paris, France and Barclays Capital, 21 boulevard de la Madeleine, 75001 Paris, France. Phone Number: +331 43 13 63 07 and +331 44 58 32 64

Markovian process. These three components interpret respectively as the level, the slope and the curvature of the yield curve.

One of the main drawbacks of Nelson and Siegel is the lack of theoretical foundations. The model only consists in an analytical expression for zero coupon yields. Whereas a large part of interest rate models refers to the no-arbitrage theory, Nelson and Siegel does not. There are several reasons for imposing no-arbitrage restrictions in interest rate models. First, from a financial point of view, no-arbitrage appears as a reasonable assumption for bonds markets, which are deep and liquid. Second, no-arbitrage is a preliminary for the derivatives pricing (vanilla caps and floors for the simplest ones). Finally, a no-arbitrage framework provides an explicit definition of the term structure risk premium, which reflects the relative price of different maturities across time and is of major interest for asset and debt management.

In this model, we introduce explicit no-arbitrage conditions in the standard Nelson and Siegel representation and we compare empirically this model to the standard one. One of the most striking differences is the spread between term structure risk premia, which is on average 15% higher in the standard model. In order to assess the consequences for portfolio management, we compare the two models for the same investment strategy and the monthly average return (during 5 years) with no-arbitrage is 2 points greater than without.

The two main insights of this article are (i) deriving properly a no-arbitrage Nelson and Siegel representation in a general 3 factor framework and therefore linking it to a theoretical background and (ii) investigating the empirical comparison with the standard model through several aspects: Yield curve fitting, risk premium modeling, variance decomposition and out-of-sample forecast. This paper is an extension of an article of Diebold, Piazzesi and Rudebush (2005). They were the first to be interested in a no-arbitrage Nelson and Siegel representation and in reconciling the views of two branches in affine yield curve modeling: The Nelson and Siegel and the no-arbitrage ones. Their no-arbitrage model consists in two independent mean reverting factors and it admits a Nelson and Siegel representation if one constraints the mean reverting speeds of both factors. We go further in two directions. First, we derive exact analytical expressions in a three factor framework. Whereas they make assumptions regarding the nullity of Jensen terms, we choose to keep complete expressions. It weakens the analytical simplicity but we earn better empirical performances. Second, we investigate the empirical performances of both models. To the best of our knowledge, this paper is the first empirical study of a three factor no-arbitrage model with a Nelson and Siegel representation.

We estimate both models using US monthly zero coupon prices from Feb. 1971 to Dec. 2000. We use the McCulloch and Kwon (1993) dataset from February 1971 to February 1991, as well as Bliss dataset from 1991 to Dec. 2000. We focus on the sixteen following maturities: 1, 3, 6, 9, 12, 15, 24, 36, ..., 120 months. We use the Kalman filter technique to estimate the likelihood of models, since it provides an interesting trade-off between the cross sectional and the times series constraints. At each period, the model has indeed to fit the yield curve, i.e. the sixteen yields of the dataset, whose maturities vary between 1 month and 10 years. Our choice avoids notably to suppose that certain yields are priced without errors on the whole sample – meaning stochastic singularities –, which is for example the case in the Chen and Scott (1993) estimation method.

Globally, the no-arbitrage Nelson and Siegel model performs significantly better than the original one. We regroup comparison results in three fields: Practical, financial and finally economic ones.

From a practical point of view, empirical costs (including computational and tractability ones) are in both models analogous. The estimation of the no-arbitrage model is not more complex than the standard Nelson and Siegel one. In fact, the number of free parameters is in both models strictly identical. We address the lack of parsimony, which is one of the usual drawbacks of factor term structure modeling.

From a financial point of view, the no-arbitrage model (i) allows improving portfolio management results and (ii) exhibits a very good fit on yield curves, as the standard version. The aspect is particularly important for asset or debt managers, who are currently using a standard Nelson and Siegel model to determine their portfolio allocations. Our model improves average one month return of more than 2 points (in a yearly basis). Regarding the second point, bond pricing errors are very low and are, on absolute average, less than 10 basis points.

From an economic point of view, our model performs better in out-of-sample forecasting than the random walk benchmark as well as than the standard Nelson and Siegel model. The information contained in today's yield curve regarding the future is better used in our model.

In the rest of the article, NS refers to the standard Nelson and Siegel representation and NANS to the no-arbitrage one. $x_t^{(\text{NS})}$ (resp. $x_t^{(\text{NANS})}$) refers to a variable modeled with NS (resp. NANS). In order to lighten the notations, we skip the superscript when there is no doubt.

2 The model: A three factor approach

We explain how we impose a Nelson and Siegel representation to a three factor term structure model that results in a no-arbitrage Nelson and Siegel model. Two reasons motivate our choice of three factors. First, the three factor case is quite general: It replicates the level, the slope and the curvature of the yield curve. Second, a third factor is useful in modeling properly the risk premium. This is the first direction in which we extend the work of Diebold, Piazzesi and Rudebush (2005), who studied a model based on two independent mean-reverting processes. Before deriving the NANS model, we present the standard Nelson and Siegel representation and affine term structure models.

2.1 Standard Nelson and Siegel representation

The Nelson and Siegel model provides a very simple expression for the zero coupon yields. The expression at time t of a zero coupon yield $y_t^{(\tau)(\text{NS})}$ maturating τ periods later is the following:

$$y_t^{(\tau)(\text{NS})} = f_t^{(1)} + f_t^{(2)} \frac{1 - \exp(-\lambda\tau)}{\lambda\tau} + f_t^{(3)} e^{-\lambda\tau} \tag{1}$$

A principal component analysis run on changes of time-series yields points out that three factors are often sufficient to explain most of the variance (more than 97 %). Litterman and Scheinkman (1991) first characterize these factors and call them respectively: 'Level', 'slope' and 'curvature'. The Nelson and Siegel model mimics relatively well this pattern. The zero coupon yield is indeed the weighted sum of three factors $\left(f_t^{(i)}\right)_{t\geq 0}$. The three weights determine the cross sectional fit, whereas the three factors determine the times series properties. To clarify the interpretation, it is useful to refer to the representation proposed by Diebold and Li (2006). The zero coupon yield expression becomes:

$$y_t^{(\tau)(\text{NS})} = f_t^{(1)} + (f_t^{(2)} + f_t^{(3)}) \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - f_t^{(3)} \left[\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right] \tag{2}$$

The interpretation becomes straightforward. The first factor $f^{(1)}$ is the level of the yield curve and affects all maturities in a uniform way. The sum of the second and third factors $f^{(2)} + f^{(3)}$ is the (opposite of the) slope and concerns more the short end of the curve than the long end. Finally, the opposite of the third factor $-f^{(3)}$ is the curvature and is particularly important for the 'middle' of the curve.

We do not chose the Diebold and Li (2006) representation because: (i) We keep the original definition of the Nelson and Siegel representation and (ii) computation of the no-arbitrage Nelson and Siegel representation is more tractable.

The factors $\left(f_t^{(i)}\right)_{t\geq 0}$ characterizing the time series behavior of the yield curve, are Gaussian :

$$d\begin{bmatrix} f_t^{(1)} \\ f_t^{(2)} \\ f_t^{(3)} \end{bmatrix} = K^{(\text{NS})}\left(\begin{bmatrix} \theta_1^{(\text{NS})} \\ \theta_2^{(\text{NS})} \\ \theta_3^{(\text{NS})} \end{bmatrix} - \begin{bmatrix} f_t^{(1)} \\ f_t^{(2)} \\ f_t^{(3)} \end{bmatrix}\right) dt + \Sigma^{(\text{NS})} dW_t^{(\text{NS})} \tag{3}$$

$$\Sigma^{(\text{NS})} = diag\left(\sigma_1^{(\text{NS})}, \sigma_2^{(\text{NS})}, \sigma_3^{(\text{NS})}\right) \qquad K^{(\text{NS})} = \left(\kappa_{ij}^{(\text{NS})}\right)_{i,j=1,2,3} \tag{4}$$

The cross sectional behavior needs estimating the parameter λ, which expresses as the inverse of a maturity. The time series properties imply to estimate a VAR(1) process of dimension 3.

2.2 Factor Term Structure model

We use a general three factor model, which generalizes the Diebold, Piazzesi and Rudebush framework. It is compatible with the Dai and Singleton (2000) classification, and according to their notations, it belongs to the $\mathbb{A}_0(3)$ family: The volatilities of the three factors are not stochastic, but constant. Moreover, the factor dynamics is Gaussian, as in the standard Nelson and Siegel representation, and the affine risk premium is unconstrained.

The short rate r_t is the sum of three latent and time-varying processes $\left(x_t^{(i)(\text{NANS})}\right)_{i=1,2,3}$, which do not have any financial or economic signification. The vector $X_t^{(\text{NANS})} = \left[x_t^{(1)(\text{NANS})}\ x_t^{(2)(\text{NANS})}\ x_t^{(3)(\text{NANS})}\right]^\top$ follows a Gaussian diffusion under the historical probability $\mathbb{P}$. $\widetilde{W}_t$ is a $\mathbb{R}^3$ Brownian motion under $\mathbb{P}$.

$$r_t = x_t^{(1)(\text{NANS})} + x_t^{(2)(\text{NANS})} + x_t^{(3)(\text{NANS})} = [1\ 1\ 1].\ X_t^{(\text{NANS})} \tag{5}$$

$$dX_t^{(\text{NANS})} = -\tilde{K}^{(\text{NANS})}\, X_t^{(\text{NANS})} dt + \Sigma^{(\text{NANS})} d\widetilde{W}_t \tag{6}$$

$$\Sigma^{(\text{NANS})} = diag\left(\sigma_1^{(\text{NANS})}, \sigma_2^{(\text{NANS})}, \sigma_3^{(\text{NANS})}\right)$$

$$\tilde{K}^{(\text{NANS})} = \left(\tilde{\kappa}_{ij}^{(\text{NANS})}\right)_{i,j=1,2,3}$$

The no-arbitrage hypothesis implies the existence of a risk-neutral probability $\mathbb{Q}$. Coefficients with $\tilde{}$ refer to coefficients under the historical probability $\mathbb{P}$, whereas coefficients without it refer to coefficients under the risk-neutral one $\mathbb{Q}$. According to the no-arbitrage pricing theory developed by Harrison and Kreps (1979), Harrison and Pliska (1981) and Kreps (1981), asset prices equal the expectation under this probability $\mathbb{Q}$ of their discounted payoffs. Remarking that the discount factor is $\exp(-\int_t^{t+\tau} r_s ds)$, the price $B(t,\tau)$ at date t of a zero coupon maturating in τ periods is: $B(t,\tau) = \mathbb{E}^{\mathbb{Q}}\left[\exp(-\int_t^{t+\tau} r_s ds)\right]$.

The change of probability from the historical one to risk neural one allows defining the instantaneous remuneration of the investor for taking the risk $d\widetilde{W}_t$. It builds up the so-called instantaneous risk premium that we note $\Sigma^{-1}\mu_t$. This means that, as soon as one compensates the risk adverse investor for taking this risk, he behaves as if he were in the risk neutral world. Girsanov has formulated this in a theorem stating that the process $W_t = \widetilde{W}_t - \Sigma^{-1}\int_0^t \mu_s ds$ is a Brownian motion under $\mathbb{Q}$.

According to the affine term structure literature, we suppose that the risk premium μ is affine: $\mu_t = \left(\theta_i^{(\text{NANS})}\right)_{i=1,2,3} + \left(\kappa_{i,j}^{(\text{NANS})}\right)_{i,j=1,2,3} X_t^{(\text{NANS})}$[1]. The two main advantages of this hypothesis are (i) to keep risk neutral diffusions and historical ones analogous and (ii) to get affine expressions for zero coupon yields. In our case, it is a sufficient condition to have Gaussian diffusions under both probabilities.

Under the risk neutral measure, the dynamics of the vector $X_t^{(\text{NANS})}$ is the following:

$$dX_t^{(\text{NANS})} = \left(\theta^{(\text{NANS})} - (\tilde{K}^{(\text{NANS})} - \kappa^{(\text{NANS})})\, X_t^{(\text{NANS})}\right) dt + \Sigma^{(\text{NANS})} dW_t$$

We define from now on $K^{(\text{NANS})} = \tilde{K}^{(\text{NANS})} - \kappa^{(\text{NANS})}$, which is the ‘mean reverting’ matrix under $\mathbb{Q}$. Using standard calculus, we derive the yield $y_t^{(\tau)(\text{NANS})}$ at date t of a zero coupon of maturity τ.

$$y_t^{(\tau)(\text{NANS})} = -\frac{1}{\tau}\left(\alpha(\tau) + \beta(\tau) X_t^{(\text{NANS})}\right) \tag{7}$$

Following Duffie and Kan (1996), the coefficients $\alpha(\tau)$ and $\beta(\tau)$ are defined through the two following ordinary differential equations with boundary conditions $\alpha(0) = 0$ and $\beta(0) = 0_3$.

$$\alpha'(\tau) = \theta^{(\text{NANS})\top}\,\beta(\tau) + \frac{1}{2}\sum_{i=1}^{3} \sigma_i^{(\text{NANS})2} \beta_i(\tau)^2 \tag{8}$$

$$\beta'(\tau) = -K^{(\text{NANS})\top}\,\beta(\tau) - [1\ 1\ 1]^\top \tag{9}$$

The coefficients $\alpha(\tau)$ and $\beta(\tau)$ determine the cross sectional behavior of the model and depend essentially on the risk neutral dynamics of factors, i.e. on parameters $\theta^{(\text{NANS})}$ and $K^{(\text{NANS})}$. The time series properties of the term structure rely on the time-varying risk premium $\kappa^{(\text{NANS})}$.

2.3 Imposing a Nelson and Siegel representation of a factor term structure model

In this section, we derive a Nelson and Siegel representation of the preceding factor term structure model. Because both models offer an affine representation for the zero coupon yield, the easiest solution consists in imposing identical factor weights for both cases. Comparing the equations (1) and (7), the equality of factor loadings in both models means simply that:

[1] These notations are homogeneous to NANS ones.

$$\beta(\tau) = -\begin{bmatrix} \tau & \frac{(1-e^{-\lambda\tau})}{\lambda} & \tau\, e^{-\lambda\tau} \end{bmatrix}^{\top} \tag{10}$$

Since an ordinary differential equation (9) defines $\beta(\tau)$, imposing $\beta(\tau)$ also constraints coefficients of (9) and thus $K^{(\text{NANS})}$. After some calculation, we find the following expression for the matrix $K^{(\text{NANS})}$:

$$K^{(\text{NANS})} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & \lambda \\ 0 & 0 & \lambda \end{bmatrix} \tag{11}$$

Imposing a Nelson and Siegel representation for the factor term structure model constraints the factor risk-neutral dynamics of $X_t^{(\text{NANS})}$. This result extends the result of Diebold, Piazzesi, and Rudebush (2005) in a three factor framework. The mean reverting speeds of their two independent factors are respectively equal to 0 and λ, which is also the 2×2 upper left submatrix of $K^{(\text{NANS})}$ in (11). Using (8), it is straightforward to derive $\alpha'(\tau)$ and then $\alpha(\tau)$, which is the uniform deformation of the yield expression. In the equation (7), we plug the expressions of $\beta(\tau)$ and $\alpha(\tau)$ to get the NANS expression of $y_t^{(\tau)(\text{NANS})}$:

$$y_t^{(\tau)(\text{NANS})} = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \frac{1-e^{-2\lambda\tau}}{2\lambda\tau} \\ e^{-\lambda\tau} \\ (1+\lambda\tau)\, e^{-2\lambda\tau} \\ \lambda\tau \\ \lambda^2\tau^2 \end{bmatrix}^{\top} \cdot \begin{bmatrix} \frac{\theta_2^{\text{NANS}}}{\lambda} - \frac{\sigma_2^2}{2\lambda^2} \\ \frac{\theta_3^{\text{NANS}} - \theta_2^{\text{NANS}}}{\lambda} + \frac{\sigma_2^2}{\lambda^2} \\ -\frac{2\sigma_2^2+\sigma_3^2}{4\lambda^2} \\ -\frac{\theta_3^{\text{NANS}}}{\lambda} \\ \frac{\sigma_3^2}{4\lambda^2} \\ \frac{\theta_1^{\text{NANS}}}{2\lambda} \\ -\frac{\sigma_1^2}{6\lambda^2} \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ e^{-\lambda\tau} \end{bmatrix} \cdot \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix} \tag{12}$$

$$dX_t^{(\text{NANS})} = -(K^{(\text{NANS})} + \kappa^{(\text{NANS})})\, X_t^{(\text{NANS})}\, dt + \Sigma^{(\text{NANS})}\, d\widetilde{W}_t \tag{13}$$

Diebold, Piazzesi, and Rudebush (2005) assume in their model that $\alpha(\tau)$ is close to 0, which allows them to derive an exact Nelson and Siegel representation. We do not follow them for two reasons: (i) Imposing $\alpha(\tau) = 0$ means strictly speaking that $\sigma_i = \theta^{\text{NANS}} = 0$ for $i = 1, 2, 3$ and deterministic factors. (ii) This assumption implies to neglect the Jensen terms, which are not null especially for long maturities.

The yield expression (12) in the no-arbitrage Nelson and Siegel model is the sum of two terms:

(i) A Nelson and Siegel like expression and (ii) an expression gathering Jensen terms and constant risk premium. The last term mainly reflects the no-arbitrage constraints. More precisely, the terms in $\sigma_i^{(\text{NANS})2}$, $i = 1, 2, 3$ are Jensen terms and therefore reflect the fundamental risk of the yield curve. Second, terms in $\theta^{(\text{NANS})}$ reflect the constant risk premium of the yield curve. These terms deform the yield curve in a uniform way, which does not depend on time t: The cross-sectional shape of the curve is different, but time-series properties remain unchanged. Moreover, the functional form in τ of those additional terms are consistent with the extension of Nelson and Siegel by Björk and Christensen (1999).

To complete the interpretation, one can argue that factor dynamics are a priori different in both models, because in the NANS model the arbitrage free dynamics is constrained through the equation (11). Nevertheless, since the risk premium μ is unconstrained, the factor dynamics under $\mathbb{P}$ is as flexible as in the NS model. The no-arbitrage conditions do not limit the dynamics of the model.

3 Empirical Estimations

We estimate the model using Kalman filter and maximum likelihood on US data from 1971 to 2000.

3.1 Data and the estimation method

3.1.1 Data

We use McCulloch and Kwon (1993) monthly nominal zero coupon prices from August 1971 to February 1991 with maturities of 1, 3, 6, 9, 12, 15, 18, 24, 36,. . . 120 months. It is an extension of McCulloch U.S. Treasury term structure data appearing in the Handbook of Monetary Economics (1990). H.C. Kwon collected the yields after 1983. For data from March 1991 to December 2000, we use Bliss data.

McCulloch and Kwon extract zero coupon yields from US data using cubic splines. In a few words, the technique is the following. They suppose that the zero coupon curve is a function of unknown parameters (here coefficients of a cubic spline function). They then express the prices of all traded debt securities using these parameters, which are estimated by minimizing the mean square errors between modeled prices and actual ones. They finally compute zero coupon yields for all maturities, which builds the dataset. These data are estimates of the true yields and are therefore affected with measurement errors.

3.1.2 Estimation method

For the estimation, we do not use the Chen and Scott (1993) technique, but the Kalman filter and the maximum likelihood estimation, as in Dai and Philippon (2004). This technique offers a natural trade-off between times-series properties of factors (i.e. $x_t^{(i)(\text{NANS})}$ and $f_t^{(i)}$ for the both models) and the cross sectional fit. In appendix, we give a detailed description of the estimation procedure.

The Chen and Scott technique relies on the assumption that certain yields are perfectly priced and do not suffer from any measurement error. Using these yields, one computes the dynamics of the latent factors. Assuming then that other yields are priced with IID errors, one can express the likelihood and, through maximization, estimate the other parameters. The main issue with this technique concerns the selection of the perfectly priced yields. To the best of our knowledge, there is no criterion to select them.

That is why we have decided to use the Kalman filter, which supposes that all yields are priced with IID errors. The selection issue and the stochastic singularity disappear. This choice is moreover consistent with the estimation of zero coupon yields, which are measured with errors.

3.2 Estimation of the standard NS model

In order to express the state-space system, we need some notations and definitions. $\tau_1, \ldots, \tau_N$ are the maturities available in the dataset, i.e. 1, 3, 6, 12,...144 and 156 months ($N = 16$). $Y^{(\text{NS})}(t)$ is the vector gathering zero coupon yields of different maturities. We note $\varepsilon_t^{(\text{NS})}$ the 16×1 vector of measurement errors. Their covariance matrix $\Omega_\varepsilon^{(\text{NS})} = diag(s_1^{(\text{NS})}, \ldots, s_{16}^{(\text{NS})})$ is supposed to be diagonal, for two reasons: (i) Measurement errors are likely to be independent from each other, and (ii) these errors can be seen as an econometric tool to overcome the preceding selection problem and independent shocks are sufficient.

We define $X_t^{(\text{NS})}$ as the demeaned factors: $X_t^{(\text{NS})} = \left[f_t^{(1)} - \mathbb{E}f_t^{(1)} \quad f_t^{(2)} - \mathbb{E}f_t^{(2)} \quad f_t^{(3)} - \mathbb{E}f_t^{(3)}\right]^\top$. The vector of average factors is noted $\left[\theta_1^{(\text{NS})}\ \theta_2^{(\text{NS})}\ \theta_3^{(\text{NS})}\right] = \left[\mathbb{E}f_t^{(1)}\ \mathbb{E}f_t^{(2)}\ \mathbb{E}f_t^{(3)}\right]$. η_t is the vector of shocks affecting $X_t^{(\text{NS})}$. The covariance matrix of η_t is by construction diagonal: $\Sigma^{(\text{NS})} = diag(\sigma_1^{(\text{NS})}, \sigma_2^{(\text{NS})}, \sigma_3^{(\text{NS})})$. The factor dynamics (3) becomes after discretization (I_3 is the 3×3 identity matrix):

$$X_t^{(\text{NS})} = (I_3 - K^{(\text{NS})})X_{t-1}^{(\text{NS})} + \eta_t^{(\text{NS})}$$

$$Y_t^{(\text{NS})} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda^{(\text{NS})}\tau_i}}{\lambda^{(\text{NS})}\tau_i} & e^{-\lambda^{(\text{NS})}\tau_i} \end{bmatrix}_{i=1\dots N} (X_t^{(\text{NS})} + \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}) + \varepsilon_t^{(\text{NS})}$$

We define now $\Lambda^{(\text{NS})} = (I_3 - K^{(\text{NS})})$ and $B^{(\text{NS})} = \begin{bmatrix} 1 & \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} & e^{-\lambda\tau_i} \end{bmatrix}_{i=1\dots N}$ the $N \times 3$ matrix of factors weights. Finally $A^{(\text{NS})}$ is the constant deformation of the yield curve: $B^{(\text{NS})} \begin{bmatrix} \theta_1^{(\text{NS})} & \theta_2^{(\text{NS})} & \theta_3^{(\text{NS})} \end{bmatrix}$.

We have to estimate 32 parameters: $\Theta^{(\text{NS})} = \{\lambda^{(\text{NS})}, \kappa_{ij}^{(\text{NS})}, \theta_i, \sigma_i^{(\text{NS})}, s_n^{(\text{NS})}\}_{i,j=1,2,3 \;\; n=1,\dots,16}$.

With preceding notations, the factor dynamics and yield expressions simplify to the following state-space system:

$$Y_t^{(\text{NS})} = A^{(\text{NS})} + B^{(\text{NS})} X_t^{(\text{NS})} + \varepsilon_t^{(\text{NS})} \tag{14}$$

$$X_t^{(\text{NS})} = \Lambda^{(\text{NS})} X_{t-1}^{(\text{NS})} + \eta_t^{(\text{NS})} \tag{15}$$

The model is now written under the standard state-space system form, for example described in Hamilton's book (1994). Kalman filter techniques help to compute the best forecast $\hat{Y}_{t+1|t}^{(\text{NS})}$ of $Y_{t+1}^{(\text{NS})}$ at time t as a function of $X_t^{(\text{NS})}$. One can therefore determine the distribution of $Y_{t+1|t}^{(\text{NS})}$ knowing $X_t^{(\text{NS})}$ and then compute recursively the total likelihood. The likelihood maximization enables to estimate all parameters. In appendix, we detail further the estimation procedure through the Kalman filter.

Two algorithms are combined to modify the standard Newton-Raphson method, used to maximize the likelihood: (i) Berndt-Hall-Hall-Hausman (1974) algorithm and (ii) Levenberg-Marquardt algorithms (Levenberg 1944) (Marquardt 1963). The first algorithm avoids computing second order derivatives and decreases the number of mathematical operations in each loop. The second algorithm modifies the step between each recursion and optimizes the convergence speed toward the optimal value.

The table (Tab. 1) gathers estimation results of the parameter values with their associated standard errors. Standard errors remain weak and almost all coefficients are accurately estimated. The last line, *Log Lik*, computes the log-likelihood of the model, which is equal to 28034. Results have been checked out with *Dynare* (Juillard 2006). *Dynare*, which is a free collection of MatLab routines allowing solving, estimating and simulating non-linear and stochastic models with forward-looking variables. *Dynare* results do not differ significantly from ours.

3.3 Estimation of the NANS model

The estimation procedure for the no-arbitrage model is very close to the one described above. Only minor modifications are necessary to account for no-arbitrage restrictions. Zero coupon yields expressions (12) and factors dynamics (13) can be treated as in the standard Nelson and Siegel case, regarding the discretization and parameterization. Using analogous notations, we get the following state-space system:

$$Y_t^{(\text{NANS})} = A^{(\text{NANS})} + B^{(\text{NANS})}\, X_t^{(\text{NANS})} + \varepsilon_t^{(\text{NANS})} \tag{16}$$

$$X_t^{(\text{NANS})} = \Lambda^{(\text{NANS})}\, X_{t-1}^{(\text{NANS})} + \eta_t^{(\text{NANS})} \tag{17}$$

We estimate again 32 parameters: $\Theta^{(\text{NANS})} = \left\{\lambda^{(\text{NANS})}, \sigma_i^{(\text{NANS})}, \theta_i^{(\text{NANS})}, \kappa_{ij}^{(\text{NANS})}, s_n^{(\text{NANS})}\right\}_{i,j=1,2,3\ \ n=1,\ldots,16}$, as in the standard NS model. No-arbitrage conditions decrease the number of independent parameters of the factors term structure model and make it compatible with the one of the standard Nelson and Siegel model. Both models are equally parsimonious, which cancels out one of the traditional drawbacks of factor term structure models. The large number of parameters in factors term structure models usually leads to poorly significant parameters, which deters their empirical performances, and notably the out-of-sample forecast.

We compute recursively the likelihood and its maximization leads to the parameters gathered in table (Tab. 1). As in the standard model estimation, almost all parameters are significant. The log-likelihood is equal to 28281 in this model.

4 Comparisons of both models

In this section, we quantify to what extent no-arbitrage conditions improve the modeling of the interest rate term structure. The standard NS model overestimates the term structure risk premium at the height of 15% for 5 and 10 years maturities. Trading implications for asset and debt managers are of important magnitude and reach several tens of basis points for realized returns on the last 5 years of the dataset.

More precisely, the NANS model outperforms the standard NS for each of the following points: (i) The yield curve fitting (ii) the variance decomposition (iii) the rejection of the expectation hypothesis and finally (iv) the out-of-sample forecast performances. We explain more in details the meaning and the relevance of each of these aspects in the next paragraphs.

Models	NS		NANS	
Parameters	**Values**$\times 10^{-2}$	**Std Err.**$\times 10^{-2}$	**Values**$\times 10^{-2}$	**Std Err.**$\times 10^{-2}$
$\lambda^{(\star)}$	7.433	0.172	8.577	0.202
$\sigma_1^{(\star)}$	0.322	0.013	0.288	0.007
$\sigma_2^{(\star)}$	0.876	0.041	0.982	0.045
$\sigma_3^{(\star)}$	0.705	0.037	0.806	0.041
$\theta_1^{(\star)}$	7.852	1.276	0.064	0.001
$\theta_2^{(\star)}$	-1.706	0.739	0.052	0.034
$\theta_3^{(\star)}$	0.274	0.221	0.122	0.028
$\kappa_{11}^{(\star)}$	0.673	0.745	0.210	0.216
$\kappa_{12}^{(\star)}$	-2.528	0.601	-3.117	0.866
$\kappa_{13}^{(\star)}$	-4.728	1.060	-5.608	1.519
$\kappa_{21}^{(\star)}$	1.747	2.055	-0.477	0.714
$\kappa_{22}^{(\star)}$	5.206	1.590	-3.589	1.544
$\kappa_{23}^{(\star)}$	0.533	2.839	-9.364	2.792
$\kappa_{31}^{(\star)}$	1.766	1.672	0.851	0.599
$\kappa_{32}^{(\star)}$	2.412	1.320	3.599	1.337
$\kappa_{33}^{(\star)}$	13.301	2.294	7.234	2.336
Log Lik	28034.23	11.64	28280.98	44.16

Table 1: Estimation of both models

4.1 Likelihood Comparison

The log-likelihood comparison implies that the no-arbitrage model should be preferred to the standard one. Even if both models are not nested, it is however possible to compare their accuracy. Because they have strictly the same number of parameters, we can simply compare their log-likelihood. The standard model selection criteria, as the Akaike's information criterion (Aic) or the Schwarz's Bayesian information criterion (Bic), will indeed provide the same result. As we computed in the last line of the table (Tab. 1), it appears that the log-likelihood of the no-arbitrage model, equal to 28281, is significantly larger than the NS one, which reaches 28034. This first criterion pleads therefore in favor of no-arbitrage constraints.

4.2 Factor signification

We compute in this section the correlation between modeled factors and empirical level, slope and curvature of the yield curve. All correlations are high and the financial interpretation of the three factors is analogous in both models, which confirms the use of level, slope and curvature as factor labels.

No-arbitrage constraints make correlations larger and especially the one between the third factor and the curvature, which goes up from 88.6% to 92.1%.

Correlations (%)		
	Standard NS	**No-arbitrage NS**
Level and first factor	98.58	98.74
Slope and second factor	95.44	96.05
Curvature and third factor	88.62	92.07

Table 2: Correlations between models factors and yield curve moments

4.3 Yield curve fitting

To measure the yield curve fitting, we compare the mean square errors between modeled yields and actual ones for the 16 maturities of the dataset. The NANS model fits better the yield curve than the standard one for almost all maturities. No-arbitrage restrictions improve the cross-sectional behavior.

The table (Tab. 3) gathers the mean square errors between modeled yields and real ones. The differences are expressed in basis points. The bold number is the lowest spread between both. The average spread for the no-arbitrage model is 8.3 basis points whereas it is almost 9.8 bp for the standard one. The gain is important as well at the short end of the curve (maturities below 2 years) as the long end (maturities greater than 5 years).

Spreads between model and reality (bp)		
Maturity	NS	NANS
1	37.60	**27.28**
3	17.32	**13.08**
6	4.93	**3.89**
9	**5.75**	6.68
12	7.42	**7.37**
15	8.08	**5.76**
18	7.06	**4.94**
24	5.50	**5.22**
36	5.64	**4.25**
48	6.73	**6.64**
60	6.69	**5.89**
72	6.38	**6.14**
84	7.26	**7.08**
96	7.96	**6.55**
108	10.12	**9.73**
120	12.23	**11.73**
Mean	9.79	**8.26**

Table 3: Spreads between modeled and actual yields

4.4 Variance decomposition

We decompose the variance of the modeled forecast error in four terms. The three first ones depend on the three factors and the last one is considered as unexplained. The smaller it is, the better the explanatory power of the model is. For almost all maturities and all forecast horizons, the unexplained variance in the NANS model is lower than in the standard one. The difference between both models rises in importance for long maturities and long forecast horizons.

At a given horizon h, the mean square error between the actual yield $y_{t+h}^{(j)}$ of maturity j and its best forecast $\hat{y}_{t+h|t}^{(j)}$ at date t is equal to the sum of four terms. The first one depends only on the volatility[2] σ_1 of the first factor $x_t^{(1)}$. The two following terms have analogous interpretation and depend only respectively on σ_2 and σ_3, which are the volatilities of the two other factors. The last term is only function of the measurement standard error s_j for the yield of maturity j: This is the unexplained part of the total variance. We derive properly in appendix the computation of variance decomposition.

We compute the shares for 5 maturities: 1, 12, 36, 60 and 120 months and 5 horizons: 1, 12, 60, 120 months and ∞. The variance at the ∞ horizon is nothing else, but the unconditional variance. The numbers in the table are percentages of the total variance for the given maturity and the given horizon. For example, the number 49.82 in the ninth row and second column means that for the NANS model, 49.82% of the yield with a maturity of 36 months at a 12 month forecast is explained by the second factor.

The proportion of yield curve variance explained by the no-arbitrage model is almost always greater than the one explained by the standard representation. The only exception is the one month forecast of the 10Y zero coupon. As for the in sample fit, the explanatory power of the no-arbitrage model is better for the short end and the long end of the yield curve. The relative gain increases with the forecast horizon. The longer the horizon, the better the no-arbitrage model relative to the standard NS one. No-arbitrage constraints and Jensen terms help to improve the explanatory power of factors and this is especially true for long horizons.

The explanatory power of the three factors is also comparable for both models. For all horizons and both models, the first factor share increases with maturity (e.g. from 8.47% to 52.61% at 12 month horizon for the no-arbitrage model), whereas the second factor role is globally decreasing. The third one can have the second role in the no-arbitrage model, whereas in the standard one it always plays the third role.

[2] We skip the superscripts $^{(\text{NS})}$ and $^{(\text{NANS})}$ as soon as equalities hold for both models.

	Shares of yield variances							
	NANS				**Standard NS**			
Maturity (months)	First factor	Second factor	Third factor	Unexp.	First factor	Second factor	Third factor	Unexp.
	1 month horizon							
1	4.89	51.66	33.33	10.10	6.72	46.35	27.84	19.06
12	15.02	67.41	15.54	2.01	19.37	62.89	15.62	2.10
36	46.02	50.64	0.77	2.54	50.55	45.40	1.15	2.89
60	65.95	28.35	0.01	5.67	69.18	25.20	0.04	5.57
120	66.65	12.24	0.00	21.09	72.56	6.76	0.00	20.67
Mean	39.71	42.06	9.93	8.28	43.68	37.32	8.93	10.06
	12 months horizon							
1	8.47	59.63	30.35	1.53	9.49	60.34	26.87	3.29
12	16.97	63.32	19.50	0.19	20.15	63.43	16.16	0.23
36	34.10	49.82	15.90	0.16	42.37	47.67	9.70	0.24
60	44.01	37.78	17.87	0.32	55.42	33.95	10.19	0.43
120	52.61	25.09	20.52	1.76	66.45	19.85	11.89	1.79
Mean	31.23	47.13	20.83	0.79	38.78	45.05	14.96	1.19
	60 months horizon							
1	15.47	55.43	28.43	0.65	11.23	63.68	22.95	2.12
12	21.38	53.06	25.49	0.05	20.94	62.23	16.70	0.11
36	27.89	44.56	27.51	0.03	34.60	48.40	16.90	0.08
60	30.54	40.06	29.34	0.05	40.66	40.11	19.09	0.12
120	32.71	36.00	31.03	0.25	45.55	32.42	21.54	0.48
Mean	25.60	45.82	28.36	0.21	30.59	49.37	19.43	0.58
	120 months horizon							
1	18.51	51.80	29.20	0.47	11.40	63.68	22.81	2.08
12	23.15	49.27	27.53	0.03	21.07	62.01	16.79	0.11
36	27.38	43.43	29.15	0.01	33.95	48.73	17.23	0.07
60	28.92	40.72	30.32	0.02	39.37	41.16	19.34	0.11
120	30.10	38.41	31.33	0.14	43.61	34.33	21.62	0.43
Mean	25.61	44.73	29.51	0.13	29.88	49.98	19.56	0.56
	∞ *horizon*							
1	21.37	48.28	30.03	0.30	11.41	63.68	22.81	2.08
12	24.58	46.23	29.15	0.02	21.08	62.00	16.80	0.11
36	27.09	42.66	30.22	0.01	33.91	48.76	17.25	0.07
60	27.91	41.17	30.89	0.01	39.27	41.25	19.35	0.11
120	28.53	39.94	31.44	0.07	43.46	34.48	21.61	0.42
Mean	25.90	43.65	30.35	0.08	29.83	50.03	19.56	0.56
Mean	29.62	44.68	23.80	1.90	34.55	46.36	16.49	2.60

Table 4: Shares of variances

4.5 Comparison of the accuracy in risk premium modeling

Because it is not possible to measure directly the accuracy of term structure risk premia, we use two regressions to infer it. We do not want to test, if the in sample estimated yields exhibit this property (i.e. generate the proper regression result) but if the model parameters, estimated through maximum likelihood, are able to imply the predicted results, which is a stronger result. We use the procedure described in Dai and Singleton (2002), which is the following. For each model, we generate 500 scenarii through Monte-Carlo simulations. Each of them is a path of zero coupon yields, which counts 353 yields as the original dataset (going from Aug. 1971 to Dec. 2000). We then run the two regressions for each path and compute the mean and standard deviations of coefficients.

4.5.1 Description of the two regressions

The first regression consists in measuring to what extent both models reject the expectation hypothesis. This hypothesis assumes that the zero coupon rate at date t of maturity τ is the average of expected future short rates between t and $t+\tau$. If the zero coupon yield is the sum of expected future short rates, it is equivalent (from date t point of view) to hold (i) a τ period zero coupon during one year and then sell it as a security of maturity $\tau-1$ and (ii) to hold a 1 period zero coupon until maturity. The expected holding return of a zero coupon does not depend on maturity and the expected excess holding return (relative to the one month zero coupon) is null. Under the expectation hypothesis, the so-called term structure risk premium (including Jensen terms) is constant through time. Testing the expectation hypothesis consists in running the following regression for all maturities $n \in \{3, 6, 12, \ldots, 156\}$:

$$y_{t+1}^{(n-1)} - y_t^{(n)} = a_n + b_n \frac{1}{n-1} \left(y_t^{(n)} - y_t^{(1)} \right) + \varepsilon_t \tag{18}$$

Under the expectation hypothesis, one should get $a_n = 0$ and $b_n = 1$. As shown in Campbell and Shiller (1991), it does not hold empirically: b_n is negative and decreases with the maturity n.

The second regression measures to what extent the dynamics of the model is properly defined under the risk-neutral measure. The expectation hypothesis should hold as soon as one accounts for risk. The expected holding returns, when risk is taken into account, should be the same whatever the maturity.

The regression is the following ($B(t,n)$ is the price at time t of a zero coupon of maturity n):

$$y_{t+1}^{(n-1)} - y_t^{(n)} + \frac{1}{n-1} e_t^{(n)} = \alpha_n + \beta_n \frac{1}{n-1} \left(y_t^{(n)} - y_t^{(1)} \right) + \varepsilon_t \tag{19}$$

$$e_t^{(n)} = \mathbb{E}_t \left[\ln \left(\frac{B(t+1, n-1)}{B(t,n)} \right) - y_t^{(1)} \right] = \text{expected excess holding period return}$$

If the model exhibits a proper risk neutral dynamics, the coefficient β_n should be equal to 1.

4.5.2 Regression results for both models

We run the preceding regressions on Monte Carlo simulations of the yield curves. It results that the NANS risk premium is more accurate than the standard one, even if risk neutral dynamics are both correctly defined.

The first regression. The graph (Fig. 1) plots results. The true value of coefficients b_n, given by the data, always lies in the confidence interval of the no-arbitrage model. For long maturities (greater than 4 years), the fit is almost perfect between actual regression coefficients and the mean of modeled ones. This is not the case with the NS model: Data values lie outside the confidence interval except for some maturities. The no-arbitrage model rejects correctly the expectation hypothesis, whereas the standard model fails to do so. It does not seem to exhibit the right pattern: The graph is too “flat” compare to two others. We compute the slope for the three cases and their 95% confidence intervals:

Slope of expectation hypothesis coefficients(%)			
	Data	**NS**	**NANS**
Average	-2.63%	-0.90%	-3.70%
95% interval	$[-2.15\% \; ; -3.10\%]$	$[-0.40\% \; ; -1.40\%]$	$[-3.00\% \; ; -4.30\%]$

Table 5: Average slopes of b_n

As the table (Tab. 5) shows, we cannot reject the fact that the slope of the NANS model is equal to the true value. On the opposite, the standard NS model does not exhibit the true value. This is another clue in favor of a proper modeling of the risk premium with no-arbitrage.

The second regression. Both models perform relatively well the regression: The theoretical value of 1 lies for all maturities in the confidence intervals of both models. However, average coefficients of the no-arbitrage model are always closer to 1, even if it is not statically significant. The graph (Fig. 2) plots regression coefficients of (19) with their respective 95% confidence intervals (dashed lines).

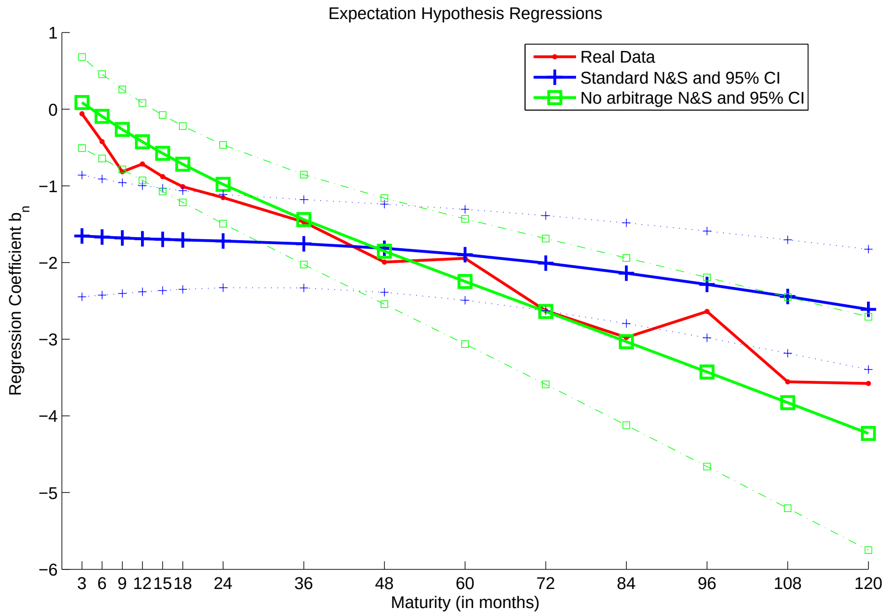

Figure 1: Regression coefficients (18) of Vs. maturities

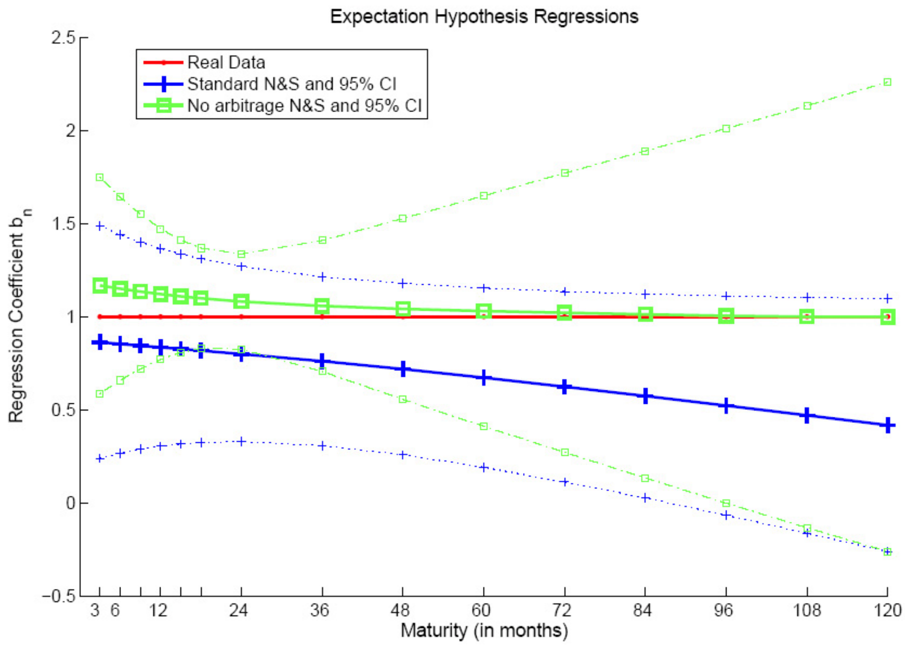

Figure 2: Regression coefficients (19) of Vs. maturities

4.6 Out-of-sample forecast performances

In this section, we compare the out-of-sample forecast performances of both models for two horizons: 1 month and 1 year. The random walk is used as a benchmark. The NANS always performs better than the standard model and than the random walk. Whatever the maturity and the forecast horizon, no-arbitrage constraints improve the out-of-sample behavior. The improvement with no-arbitrage constraints is for most maturities significant at the 10% level for the one month horizon and at the 5% level at the 1 year horizon. It means that the information contained in today's no-arbitrage NS term structure – and particularly in the risk premium – helps to better anticipate tomorrow's yields.

To estimate the forecasts at date $t+1$ and $t+12$, we estimate both models until date t. We compute then the one month and one year out-of-sample forecast $\hat{Y}_{t+1|t}$ and $\hat{Y}_{t+12|t}$ for the 16 yields of the dataset (1, 3, 6, 9, 12..., 120 months). We finally estimate these forecasts for the 60 last dates of the dataset, for the NANS model, the NS one and the random walk. Using these estimations, two criteria help to assess the accuracy of out-of-sample forecast performances: The Root Mean Squared Error (RMSE) and the Mean Absolute Deviation (MAD). If one notes $\hat{y}^{(\tau)}_{t|t-j}$ the j months ($j = 1$ or 12 months) forecast of yield $y^{(\tau)}_t$ of maturity τ at time t, the expression of both criteria is (T is the last date in the dataset):

$$MAD^{(j)}(\tau) = \frac{1}{60}\sum_{t=T-59}^{T} \left|\hat{y}^{(\tau)}_{t|t-j} - y^{(\tau)}_t\right| \tag{20}$$

$$RMSE^{(j)}(\tau) = \sqrt{\frac{1}{60}\sum_{t=T-59}^{T} \left(\hat{y}^{(\tau)}_{t|t-j} - y^{(\tau)}_t\right)^2} \tag{21}$$

Looking at equations (20) and (21), it is straightforward that the smaller the criterion is, the better the forecast performance of the model. We use Diebold and Mariano (1995) and (2000) test to measure whether the NANS forecasts are significantly more accurate than both other ones. For each maturity and each horizon, we compare NANS to the two other forecasts (RW or NS). In tables (Tab. 6) and (Tab. 7), one star means that the result holds at the 10% level and two mean that it holds at the 5% one. For both forecasts, NANS perform significantly better than the RW especially for both ends of the curve and for the long horizon (i.e. one year).

Three reasons may explain the success of NANS in the out-of-sample forecast. First, as pointed out by Duffee (2002), a time-varying risk premium, similar to ours, is of great help in forecasting, because it makes the information in today's term structure more relevant for predicting tomorrow's yields. Second, our model offers an interesting channel to understand how it helps to beat the random walk forecast.

1 month Forecast Performances ($\times 10^{-3}$)						
	MAD criteria			**RMSE criteria**		
Maturity	RW	NANS	NS	RW	NANS	NS
1	2.5487	**2.4408****	2.5458	3.6375	**3.0990****	3.5475
3	1.7217	**1.6734**	1.7302	2.4596	**2.3620**	2.3720
6	1.9335	**1.8703***	1.9218	2.5233	**2.4593***	2.4693
9	2.0653	**1.9996***	2.0532	2.7028	**2.6654**	2.6803
12	2.3523	**2.2938**	2.3429	2.9997	**2.9626**	2.9920
15	2.3860	**2.3266**	2.3800	2.9937	**2.9854**	3.0026
18	2.4508	**2.3962**	2.4516	3.0282	**3.0199**	3.0485
24	2.5568	**2.4936***	2.5552	3.1741	**3.1653**	3.1977
36	2.6116	**2.5315***	2.5859	3.1958	**3.1872**	3.2189
48	2.6225	**2.5602***	2.6085	3.2571	**3.2281**	3.2653
60	2.5999	**2.5343***	2.5801	3.2044	**3.1755**	3.2084
72	2.4717	**2.3995****	2.4405	3.0969	**3.0494***	3.0979
84	2.4187	**2.3414****	2.3739	3.0636	**3.0160***	3.0614
96	2.2827	**2.2249****	2.2635	2.9322	**2.8864***	2.9198
108	2.2633	**2.2023****	2.2433	2.9401	**2.8941***	2.9311
120	2.2830	**2.2156****	2.2464	3.0032	**2.9562***	2.9864
Mean	2.3480	**2.2815**	2.3327	3.0133	**2.9445**	2.9999

Table 6: Forecast performances of both models

12 months Forecast Performances ($\times 10^{-3}$)						
	MAD criteria			RMSE criteria		
Maturity	RW	NANS	NS	RW	NANS	NS
1	10.0970	**9.4646****	10.5225	12.8757	**11.9131****	13.2608
3	9.9668	**9.3626***	10.0844	12.9128	**12.1191***	12.9538
6	9.9633	**9.4821***	10.7225	12.8465	**12.2805***	13.6449
9	10.1248	**9.6144***	11.0781	12.9267	**12.4266***	14.0849
12	10.1598	**9.7236***	11.2087	12.9839	**12.5256***	13.7495
15	10.1043	**9.7701***	11.0899	12.8509	**12.4629***	14.1529
18	10.0815	**9.7242***	11.0216	12.7405	**12.3509***	14.0977
24	10.0852	**9.6041***	10.5948	12.4836	**12.0219***	13.6561
36	9.6209	**9.0444****	10.3803	11.7932	**11.1931****	12.9283
48	9.2699	**8.5731****	10.3801	11.2154	**10.4809****	12.587
60	9.1660	**8.3847****	10.2137	10.9690	**10.1018****	12.3202
72	8.7552	**7.8942****	10.6504	10.4415	**9.5314****	12.0781
84	8.7129	**7.7580****	10.9269	10.2765	**9.2708****	11.7688
96	8.3753	**7.3888****	11.3287	9.9143	**8.8874****	11.6258
108	8.2699	**7.2672****	11.3779	9.8485	**8.8166****	11.3635
120	8.2562	**7.2599****	11.0228	9.7736	**8.7099****	11.1515
Mean	9.4381	**8.7698**	10.7877	11.6783	**10.9433**	12.8390

Table 7: Forecast performances of both models

The random walk is a polar case of our no-arbitrage NS representation. It is the first factor of our model in the risk neutral world. If the RW had beaten our model, it would have meant that adding a risk premium and two other factors with no-arbitrage restrictions would have been useless in improving the forecast performance. The supplementary factors and associated constraints (no-arbitrage ones) are therefore likely to help in forecasting. Finally, our no-arbitrage model remains parsimonious, which avoids one of the traditional drawbacks of affine term structure models. The large number of parameters in those kinds of models is sometimes evoked to explain their failure in forecasting relative to more parsimonious ones.

4.7 Risk premia and trading the yield curve

In this section, we investigate the practical implications of preceding results in terms of (i) risk premium modeling and (ii) portfolio management. First risk premia implied with both models differ quite widely. NS risk premium exceeds NANS one with 30 bp (15% in relative terms). Second, portfolio returns with our model are larger than with NS. The average realized return on the last 60 months differs from 2 points in absolute terms. In relative terms, the NANS implied return is almost 3 times greater.

4.7.1 Term structure risk premium

We begin with comparing the risk premium modeling because it is a key factor in portfolio management. The difference between risk premia modeled by NS and NANS has an order of magnitude of 30 bp, which means more than 15% in relative terms. The relative cost of different maturities suffers from important differences, which is likely to impact strongly portfolio or debt management. This difference in risk premium modeling is a consequence of the differences in the expectation hypothesis regressions and in the out-of-sample forecasts. Preceding results plead therefore in favor of a more accurate risk premium for the NANS model relatively to the standard NS one, even if we need further investigation.

We include Jensen terms in the risk premium $RP_t^{(n)}$ and we define it as the differences between the zero coupon yield of maturity n at time t and the average of expected future short rates (equal in our model to the one month yield $y_{t+k}^{(1)}$):

$$RP_t^{(n)} = y_t^{(n)} - \frac{1}{n} \sum_{k=1}^{n-1} \mathbb{E}_t r_{t+k} \tag{22}$$

To compute the right hand side of (22), we come back to the yield definition and to the state-space systems (equations (14), (15), (16) and (17)), which have the following generic form for both models:

$$Y_t = A + B\,X_t + \varepsilon_t$$

$$X_t = \Lambda\,X_{t-1} + \eta_t$$

The best forecast $\hat{Y}_{t+k|t}$ at horizon k of Y_t is equal to $A + B\,\Lambda^k\,X_t$. Since $\mathbb{E}_t r_{t+k} = \hat{y}^{(1)}_{t+k|t}$ is the first component of vector $\hat{Y}_{t+k|t}$, this allows computing the right side of the equation (22).

Mean and Standard deviations of risk premia (%)				
	Standard NS		No-arbitrage NS	
5 Years: Mean (Standard deviation)	1.791	(1.131)	1.459	(1.076)
10 Years: Mean (Standard deviation)	2.268	(1.288)	1.729	(1.202)

Table 8: Mean and Standard deviations of risk premia

The average size of the NANS (NS) risk premium for the 5 year bond is 1.46% (1.79%) and 1.73% (2.27%) for the 10 year one. Time variations of the 10 Y premium are plotted on (Fig. 3). The graph (Fig. 4) shows the evolution of the difference between NS risk premium and NANS one. Even if not reported, patterns for the 5Y risk premium are similar. It reflects the overvaluation of the risk premium with standard NS relative to our model. Even if preceding comparisons between both models seem to show that the risk premium of our model is more accurate, we investigate it in portfolio management.

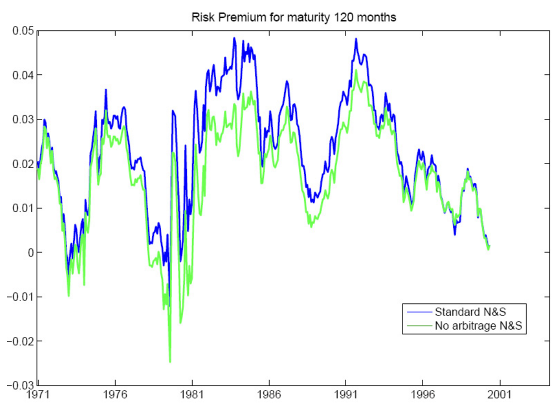

Figure 3: 10 year risk premium

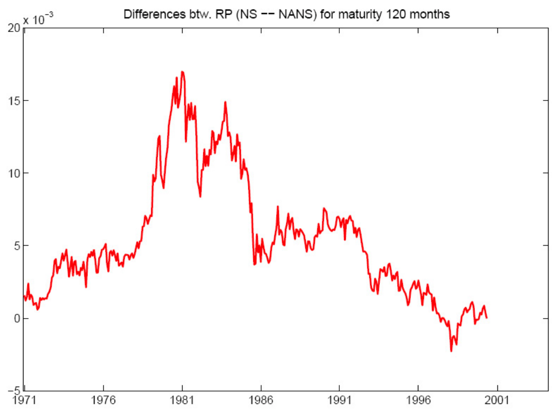

Figure 4: Difference btw. 10 year RP

4.7.2 Portfolio management

We study the consequences of preceding results regarding risk premia on portfolio management. In a simple mean-variance portfolio framework, the NANS model implies an average one month return that is 200 basis points greater than the one of the NS model. The result remains true for the ex-post welfare. This insures that a larger return on average is not the direct consequence of a riskier strategy.

We consider the following portfolio strategy. Each month, we invest a fraction ω_i of our wealth (normalized to 1) in a zero coupon bond of maturity τ_i, which lies between 1 month and 10 years. The investment follows a simple mean-variance criterion, which consists in maximizing the one month expected return of the portfolio minus a share of its expected variance. If we note R_i ($i = 1, \ldots 16$) the return of each zero coupon, R_π the return of the portfolio and γ the (constant absolute) risk aversion of the investor, the investment strategy sums up to:

$$\max_{\omega} \mathbb{E}R_\pi - \frac{\gamma}{2}\mathbb{V}R_\pi$$
$$\text{s.t.} \begin{cases} R_\pi(\omega) &= \sum_{i=1}^{16} \omega_i R_i \\ \sum_{i=1}^{16} \omega_i &= 1 \end{cases}$$

As for the out-of-sample forecast, we compute the optimal portfolio shares and the realized returns for the last 60 months of the dataset. Instead of the welfare, we compute more simply the realized average return minus $\gamma/2$ times the realized variance, which has the same meaning.

Portfolio Performances			
	CS	**NANS**	**NS**
		$\gamma = 1$	
Return (Y/Y %)	0.49%	**6.19%**	1.90%
Welfare (%)	0.49%	**4.67%**	1.79%
Times as 1st rank	12	**35**	13
		$\gamma = 2$	
Return (Y/Y %)	0.49%	**3.00%**	1.25%
Welfare (%)	0.48%	**2.37%**	1.17%
Times as 1st rank	12	**34**	14
		$\gamma = 3$	
Return (Y/Y %)	0.49%	**2.52%**	0.87%
Welfare (%)	0.48%	**2.01%**	0.83%
Times as 1st rank	11	**35**	14
		$\gamma = 4$	
Return (Y/Y %)	0.49%	**2.02%**	0.77%
Welfare (%)	0.47%	**1.62%**	0.71%
Times as 1st rank	10	**35**	15
		$\gamma = 5$	
Return (Y/Y %)	0.49%	**1.50%**	0.57%
Welfare (%)	0.47%	**1.22%**	0.56%
Times as 1st rank	12	**34**	14
		Average	
Return (Y/Y %)	0.49%	**3.05%**	1.07%
Welfare ($\times 10^{-3}$)	0.48%	**2.38%**	1.01%
Times as 1st rank	11.4	**34.6**	14

Table 9: Portfolio performances of both models

We compare then the results for our model (NANS), for the standard Nelson and Siegel (NS) and for the constant strategy (CS), which consists in investing an equal weight 1/16 of wealth in each bond. NANS model outperforms both others for the mean realized return and for the ex-post welfare: The larger return of NANS is not impaired by a larger variance.

For values of risk aversion varying between 1 and 5, the spread between returns of NS and NANS vary from 429 bp ($\gamma = 1$) to 93 bp ($\gamma = 5$) with a mean equal to 198 bp. The 'welfare' improvement varies from 0.66 points to 2.88 points.

Moreover, our model is the best in 58% of the cases. For almost 35 periods amongst the 60 considered, our model beats the two others. The constant strategy is the best in 19% of cases and the NS only in 23%. The table (Tab. 9) gathers results. The return is the average realized yearly return on a yearly basis. Times as 1st rank is the number of periods (amongst the 60), where a given model is better than the two others.

5 Conclusion

This article provides a simple model, close to the standard NS representation, with no-arbitrage cross-sectional restrictions. The main difference with an unconstrained factor term structure model is the factor risk-neutral dynamics. The main differences with a standard NS model are the Jensen and the constant risk premium ones (time series constant terms), whose expression is constrained with no-arbitrage.

The explicit risk premium defined in our model is on average 15% lower than the one with the standard NS, which leads to more efficient portfolio allocations. For the basic strategy we have considered, the average return is roughly 2 points larger than the one with Nelson and Siegel. This improvement is the consequence of a better replication of several stylized facts: The yield curve fitting, the variance explanation, the rejection of the expectation hypothesis and the out-of-sample forecast. It is noteworthy that this behavior is not a consequence of a richer framework because both models have the same number of parameters and are as parsimonious as each other. The explicit definition of a risk premium, even if linear, is efficient in capturing the information in the current curve for the prediction of tomorrow's one. No-arbitrage restrictions constrain the term structure modeling in a relevant way.

References

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974): "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3, 653–665.

BIS Monetary and Economic Department (2005): "Zero-Coupon Yield Curves: Technical Documentation," *BIS Papers No. 25*.

Björk, T., and B. J. Christensen (1999): "Interest Rate Dynamics and Consistent Forward Rate Curves," *Mathematical Finance*, 9(4), 323–348.

Campbell, J. Y., and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *The Review of Economic Studies*, 58, 495–514.

Chen, R.-R., and L. O. Scott (1993): "Maximum Likelihood Estimation for a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, 3, 14–31.

Dai, Q., and T. Philippon (2004): "Fiscal Policy and the Term Structure of Interest Rates," *New York University Working Paper*.

Dai, Q., and K. J. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55(5), 1943–1978.

——— (2002): "Expectation puzzles, Time–varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63(3), 415–441.

Diebold, F. X., and C. Li (2006): "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130(2), 337–364.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

——— (2000): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 20(1), 134–144.

Diebold, F. X., M. Piazzesi, and G. D. Rudebusch (2005): "Modeling Bond Yields in Finance and Macroeconomics," *American Economic Review (Papers and Proceedings)*, 95(2), 415–420.

Duffee, G. R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57(1), 405–443.

Duffie, D., and R. Kan (1996): "A Yield Factor Model of Interest Rates," *Mathematical Finance*, 6(4), 379–406.

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

HARRISON, M. J., AND D. M. KREPS (1979): "Martingales and Arbitrage in Multi-Period Securities Markets," *Journal of Economic Theory*, 20(3), 381–408.

HARRISON, M. J., AND S. R. PLISKA (1981): "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications*, 11(3), 215–260.

JUILLARD, M. (2006): "Dynare Manual," http://www.cepremap.cnrs.fr/dynare/download/manual/Dynare_UserGuide_WebBeta.pdf.

KREPS, D. M. (1981): "Arbitrage and Equilibrium in Economies with Infinitely many Commodities," *Journal of Mathematical Economics*, 8(1), 15–35.

LEVENBERG, K. (1944): "A Method for the Solution of Certain Problems in Least Squares," *Quarterly of Applied Mathematics*, 2(2), 164–168.

LITTERMAN, R., AND J. SCHEINKMAN (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1(1), 54–61.

MARQUARDT, D. (1963): "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal of Applied Mathematics*, 11(2), 431–441.

MCCULLOCH, H. J., AND H.-C. KWON (1993): "US Term Structure Data, 1947–1991," *Ohio State University Working Paper*, 93–6.

NELSON, C. R., AND A. F. SIEGEL (1987): "Parsimonious Modeling of Yield Curves," *The Journal of Business*, 60(4), 473–489.

SHILLER, R. J., AND H. J. MCCULLOCH (1990): "The Term Structure of Interest Rates," in *Handbook of Monetary Economics*, vol. 1, chap. 13, pp. 627–722. Elsevier.

Appendix

A Variance decomposition

In this section we derive explicit expressions for the variances decompositions. Suppose that one has the following state system (with standard notations):

$$\begin{aligned} Y_t &= A + BX_t + \varepsilon_t \\ X_t &= \Lambda X_{t-1} + \eta_t \end{aligned}$$

We suppose that all eigenvalues of Λ lie in the unit circle[3]. One can write the following $MA(\infty)$ representation for Y_t:

$$Y_t = A + \sum_{k=0}^{\infty} B\Lambda^k \eta_{t-k} + \varepsilon_t$$

The error between the forecast $\hat{Y}_{t+h|t}$ at date t of Y at horizon $h > 0$ and actual Y_{t+h} is therefore:

$$Y_{t+h} - \hat{Y}_{t+h|t} = \sum_{k=0}^{h-1} B\Lambda^k \eta_{t+h-k} + \varepsilon_{t+h} \tag{23}$$

Because the covariance matrix of η is diagonal and because η is independent of ε (this is true for our both models), the expression of the mean square error $MSE^j(h)$ for the component j at horizon h is the following (We note E_{ii} the $3\times$ 3 matrix with a 1 in position (i,i) and 0 elsewhere):

$$\begin{aligned} MSE^j(h) &= \mathbb{E}\left[(Y^j_{t+h} - \hat{Y}^j_{t+h|t})^2\right] \\ &= \sum_{i=0}^{3} \underbrace{\left\{\sum_{k=0}^{h-1} \left[B\,\Lambda^k\, E_{ii}\,\Lambda^{k\,\top}\, B^\top\right]\right\}_j}_{=K^j_i} \sigma_j^2 \\ &+ s_j^2 \end{aligned}$$

One can then compute the variance share of yield j at horizon h explained by the ith ($i = 1,2,3$) factor $\omega^j_i(h)$ and the residual $\omega^j_r(h)$:

$$\begin{aligned} \omega^j_i(h) &= \frac{K^j_i}{MSE^j(h)}\,\sigma_i^2 \qquad i = 1,2,3 \\ \omega^j_r(h) &= \frac{1}{MSE^j(h)}\,s_j^2 \end{aligned}$$

With $h \to \infty$, one gets the unconditional variance decomposition.

[3]This assumption is true for both models, the standard one and the no–arbitrage one.

B Kalman filtering and ML estimation

This section refers to the ML estimation using Kalman filter estimations. Suppose that the state–space system has the following expression:

$$Y_t = A + B\,X_t + \varepsilon_t \tag{24}$$

$$X_t = \Lambda\,X_t + \eta_t \tag{25}$$

We suppose that the perturbations ε_t and η_t are independent of each other and of the initial state X_0. Their respective variance matrices are Σ_ε and Σ_η.

The standard Kalman filter update equations provide an expression for the forecasts $\hat{X}_{t+1|t}$ of the state process X_t and associated mean squared errors $P_{t+1|t}$:

$$\hat{X}_{t+1|t} = \Lambda\hat{X}_{t|t-1} + \Lambda P_{t|t-1}B^\top\left(BP_{t|t-1}B^\top + \Sigma_\varepsilon\right)^{-1}(Y_t - A - B\hat{X}_{t|t-1})$$

$$P_{t+1|t} = \Lambda\left[P_{t|t-1} - P_{t|t-1}B^\top\left(BP_{t|t-1}B^\top + \Sigma_\varepsilon\right)^{-1}BP_{t|t-1}\right]\Lambda' + \Sigma_\eta$$

Regarding the initial conditions, $X_{1|0}$ is the best forecast at date 0 of X_1. Because of the stationarity of X under the historical probability $\mathbb{P}$, this is simply the unconditional mean of X.

$$X_{1|0} = \mathbb{E}^{\mathbb{P}}[X] \tag{26}$$

Again because the process X is stationary, the initial mean squared error is defined as:

$$vec(P_{1|0}) = \left(I_{r^2} - \Lambda\otimes\Lambda\right)^{-1} vec(\Sigma_\eta) \tag{27}$$

In the preceding equation, r is the number of components of X $(= 3)$, $\otimes$ is the Kronecker product, and $vec(M)$ is the vector representation of the matrix M[4].

One can now remark that the distribution of Y_t knowing $\mathcal{I}_t$ (representing the information available at time t) is the following:

$$Y_t\big|\mathcal{I}_t \rightsquigarrow \mathcal{N}\left(A + B\hat{X}_{t+1|t},\; BP_{t+1|t}B^\top + \Sigma_\varepsilon\right) \tag{28}$$

One can then compute the log likelihood $\mathcal{L}$ corresponding to the observation $\tilde{Y}_{t_k}$. Θ is the set of parameters one needs to estimate.

$$\begin{aligned}\mathcal{L}_{\tilde{Y}_{t_k}}(\Theta) = -\frac{1}{2}\big[&\ln(2\pi) + \ln\big(\det(BP_{t+1|t}B^\top + \Sigma_\varepsilon)\big) \\ &+(\tilde{Y}_{t_k} - (A\mu + B\hat{X}_{t+1|t}))^\top(BP_{t+1|t}B^\top + \Sigma_\varepsilon)^{-1}\times(\tilde{Y}_{t_k} - (A + B\hat{X}_{t+1|t}))\big]\end{aligned} \tag{29}$$

We compute likelihood recursively and then we maximize it relative to Θ.

[4] If $M = (m_{ij})_{i=1\ldots n, j=1\ldots p}$ then $vec(M) = [m_{11}\; m_{21}\ldots m_{n1} m_{12}\ldots m_{np}]^\top$