

# Nelson and Siegel, no–arbitrage and risk premium

François Le Grand\*

**First draft:** 1st December 2005

**This draft:** 3rd January 2006

## PRELIMINARY DRAFT

### Abstract

In this article, we measure how no-arbitrage helps to improve the standard Nelson and Siegel model. I begin with deriving a Nelson and Siegel representation from a two factors term structure model. I estimate then the two representations with Kalman filter and maximum likelihood on US data from 1952 to 1998. The comparison reveals that a proper risk premium contributes to explain better several stylized facts (variance decomposition, rejection of the expectation hypothesis, out-of-sample forecast), without deteriorating usual Nelson and Siegel achievements (yield curve fitting and parameters parsimony).

We therefore propose an easy solution to integrate in a standard two–factors N&S model, a risk premium which satisfies properly main empirical features.

---

\*PSE (joint research unit CNRS–EHESS–ENPC–ENS), Paris and Barclays Capital, Paris. Email: [francois.le.grand@pse.ens.fr](mailto:francois.le.grand@pse.ens.fr) and [francois.legrand@barclayscapital.com](mailto:francois.legrand@barclayscapital.com). Postal Address: PSE–ENS, 48 boulevard Jourdan, 75014 Paris, France and Barclays Capital, 21 boulevard de la Madeleine, 75001 Paris, France.

# 1 Introduction

Nelson and Siegel models are popular amongst some financial experts, notably certain asset managers, central bankers or debt managers. As reported in a recent report from the Bank for International Settlements (Oct. 2005) [16] regarding zero-coupon yields curves, most of central banks use Nelson and Siegel approaches. To my mind, there are two reasons for this popularity. First, it does a pretty good job in fitting the yield curve. Second, it provides a simple framework as well as for the parameters estimation as for the economic interpretation. Nelson and Siegel gives indeed a simple analytical expression for the zero-coupon yield. It expresses the yield as the weighted sum of three terms: a level factor, a slope one and a curvature one. One of its main drawbacks is its lack of theoretical foundations. Whereas a wide part of interest rates models refers to no-arbitrage theory, Nelson and Siegel do not. As a consequence, it does not propose an explicit term structure risk premium modeling.

I have two main objectives in this article: (i) deriving properly a no-arbitrage Nelson and Siegel representation and (ii) investigating the empirical comparison with standard model through several aspects: yield curve fitting, risk premium modeling, variance decomposition and out-of-sample forecast.

Several reasons motivate imposing no-arbitrage in Nelson and Siegel models. First, from a financial point of view, no-arbitrage appears as a reasonable assumption for bonds markets, which are deep and liquid. In a pricing perspective, no-arbitrage is also a preliminary for the derivatives (vanilla caps and floors for the simplest ones) pricing. Standard Nelson and Siegel does not allow it, whereas the no-arbitrage one could. Moreover, as it is pointed out by Ang and Piazzesi [1], a proper risk premium definition - notably time varying - permits to reject the Expectation Hypothesis test and improves out-of-sample forecasts.

This article has been partly inspired with Diebold, Piazzesi and Rudebush work [10], who have the first been interested in a no-arbitrage Nelson and Siegel representation. I will follow them regarding the way of imposing no-arbitrage conditions. I look under which conditions an affine factor term structure model may admit a Nelson and Siegel representation. By construction the factor term structure model is arbitrage free, and thus the Nelson and Siegel expression too. Diebold, Piazzesi and Rudebush look at a simple case with only two independent factors. I complete their approach with first a complete derivation of the no-arbitrage Nelson and Siegel representation and second empirical comparisons with standard representation.

I estimate both models on Fama-Bliss dataset. It contains the US monthly zero-coupon prices from January 1952 to December 1998. I use the six following maturities: 3, 6, 12, 24, 60 and 120 months.

I have chosen to use Kalman filter techniques to estimate the likelihood of models. I then maximize this likelihood through optimization techniques: Berndt-Hall-Hausman algorithm and Levenberg-Marquardt algorithms. This technique provides an interesting trade-off between the cross sectional and the times series constraints. One has indeed to keep in mind that at each period, the model has to fit the yield curve, i.e. the six yields of the dataset, whose maturities vary between 3 months and 10 years. To my mind, my choice avoids notably to suppose that certain yields are priced without errors on the whole sample, which is for example the case in the Chen and Scott [5] estimation method. The choice of these perfectly measured yields may influence the para-

meters estimation and I do not know any simple criteria which can select them. On the contrary, with the Kalman filter, all yields are supposed to be priced with errors. The first explanation for them is simply the fact that zero-coupon yields are not directly the prices of market assets but are extracted from various bond prices.

After the estimation of both models, I compare them and I focus especially on the following aspects: share of term structure variance explained for different horizons, rejection of the expectation hypothesis and out-of-sample forecasts. Globally, the no-arbitrage Nelson and Siegel model performs significantly better than the standard one, although both are as tractable as each other. Let us look more at results into details. I can regroup my models comparisons in three fields: practical one, financial and finally economic ones.

From a practical point of view, the estimation of the no-arbitrage model is not significantly more complex than the standard Nelson and Siegel one. In fact, even if supplementary terms are necessary for imposing no-arbitrage constraints, the number of free parameters is in both models strictly identical. I address through these specifications the lack of parsimony which is one of the usual drawbacks of factor term structure. Moreover, one can use in both models the same estimation methods with very few modifications. Empirical costs (including computational and tractability ones) are in both models analogous. Finally, no-arbitrage becomes a supplementary channel to understand the functional forms chosen for Nelson and Siegel models by Diebold and Li [9], or its 4 and 5 factors extensions by Björk and Christensen [3].

From a financial point of view, the no-arbitrage model exhibits, as the standard version, a very good fit on yield curves. The bonds pricing errors are indeed very low and are, on average, about 10 basis points for a given maturity. Because yield curve fitting is a central feature of standard Nelson and Siegel models, it is crucial that imposing no-arbitrage constraints does not deter it. In the no-arbitrage Nelson and Siegel, the explicit definition of a risk neutral measure allows moreover to investigate the pricing of interest rates derivatives (e.g. caps, floors), which is not possible through the standard version.

From an economic point of view, the standard interpretation of factors in terms of level and slope is still robust with no-arbitrage. The explicit definition of the term structure risk premium allows us to investigate the market expectations for various maturities across time. For instance, whereas the portfolio optimization (and especially the duration choice) in a standard Nelson and Siegel rely on unknown term structure risk premium, the no-arbitrage version allows a proper job.

I perform several tests to assess the preceding affirmations. One of the standard tests for measuring the accuracy of the risk premium consists in testing the rejection of the expectation hypothesis. Under this hypothesis, the excess return of a zero-coupon yield of maturity  $n$  at time  $t$  is an unbiased predictor of yields variations between  $t$  and  $t+1$ . The regression of yield variation on corresponding excess returns should lead to a unity coefficient. In fact, empirics reveals that this coefficient is always negative and decreasing with maturity (at least on US data). In order to avoid small sample biases, I do not only perform this test on in ample modeled yields. I use Monte Carlo simulations to generate 1000 paths with both models. Each of them is a time series, containing as many yield curves as the original dataset. I therefore test if the model with its coefficients estimated through maximum likelihood is able to reject properly

the expectation hypothesis. As a result, I find that my no-arbitrage Nelson and Siegel model performs better than the standard one. The explicit time-varying risk premium improves therefore the standard Nelson and Siegel model.

I am also interested in the variance decomposition to compare both models performances in forecasting variances at different horizons. Whatever the horizon (and even for the unconditional variance, implying an infinite horizon), the no-arbitrage model performs better than the standard one. The share of residual variance, i.e. unexplained variance, is always smaller in my model. One can also gauge the respective impact of both factors in the variance forecast. The longer the horizon, the more important the level. On the other side, the slope factor accounts more for the short-term of the curve.

The last point I look at is the out-of-sample forecast performances of both models. I use the random walk as a benchmark. Duffee [12] reports indeed that factor term structure models do not easily beat the random walk. I compute one-month forecast for the last sixty months of the sample. For a forecast at date  $t$ , I estimate each model on data until  $t-1$ , which implies 60 calibrations. I compare then this forecast to the actual yield using two criteria: the mean absolute deviation and the root mean squared error. For all maturities, my no-arbitrage model outperforms the Nelson and Siegel one and the random walk. As reported by Ang and Piazzesi, the explicit definition of a time-varying risk premium is of great help in this achievement. My model also offers two other channels to understand this performance. First, it remains as parsimonious as the standard Nelson and Siegel one. This is one of the usual reproaches addressed to factor term structure models when they perform poorly in out-of-sample forecast. Second, The level factor of my model is constrained to be a random walk under the risk neutral measure. If the RW had beaten my model, it would have meant that adding a risk premium and a second factor would have been useless in improving the forecast performance.

To conclude with, I propose a model with no-arbitrage and a Nelson and Siegel representation. The usual advantages of Nelson and Siegel are not affected with no-arbitrage constraints. My model fits the yield curve as well as the standard Nelson and Siegel and is as tractable and as parsimonious. Moreover, it outperforms significantly in the rejection of the expectation hypothesis and the out-of-sample forecast. The first point is particularly important for portfolio management (for assets or debt) by understanding market expectations, whereas the second one is of major interest in monetary policy.

In the rest of the article, we note N&S the standard Nelson and Siegel representation and NA-N&S the no-arbitrage one.

## 2 A two factors approach

We begin with a two factors approach. Two reasons motivate this choice. As said in the introduction, the first one is to continue the work of Diebold, Piazzesi and Rudebush [10]. The second one is to give in a simple framework, the no-arbitrage Nelson and Siegel representation.

Let us precise the background, i.e. the factor term structure model and the Nelson and Siegel one.

## 2.1 Factor Term Structure model

We modify slightly Diebold, Piazzesi and Rudebush framework to make it compatible with Dai and Singleton classification [7]. If we use their notations, we consider a  $\mathbb{A}_0(2)$  model. It is moreover a little simplified because we do not take into account crossed terms in the factors dynamics.

We suppose that the short rate is the sum of two Ornstein Ühlenbeck processes, whose dynamics under the historical probability  $\mathbb{P}$  is:

$$r_t = x_t^{(1)} + x_t^{(2)} \quad (2.1)$$

$$dx_t^{(i)} = -\tilde{\kappa}_i x_t^{(i)} dt + \sigma_i d\widetilde{W}_t^{(i)} \quad i = 1, 2 \quad (2.2)$$

The price of a zero-coupon issued in time  $t$  and maturing  $\tau$  periods later is noted  $P^{(FTSM)}(t, \tau)$  and is equal to the expectation under the risk-neutral measure  $\mathbb{Q}$  of the stochastic discount factor  $\exp(-\int_t^{t+\tau} r_s ds)$ .

Let us note  $\mu_t$  the associated market risk premium. It means that  $W_t = \widetilde{W}_t - \mu_t$  is a Brownian motion under the risk neutral measure  $\mathbb{Q}$  (under regularity conditions regarding  $\Lambda_t$ ). According to the affine term structure literature, we suppose that this risk premium is affine:  $\mu_t = \mu_0 + \mu_1 X_t$ . For a simplification purpose (and in line with Diebold, Piazzesi and Rudebush paper), we suppose that  $\mu_1$  is diagonal. We do not allow crossed interactions for the risk premium.

Under the risk neutral measure, the dynamics of factors  $x_t^{(i)}$  is the following:

$$dx_t^{(i)} = (\mu_0^{(i)} - (\tilde{\kappa}_i - \mu_1^{(i)}) x_t^{(i)}) dt + \sigma_i dW_t^{(i)} \quad i = 1, 2$$

We define from now on  $\kappa_i = \tilde{\kappa}_i - \mu_1^{(i)}$ , which is the mean reverting speed under the risk neutral measure.

Using standard calculus, we can derive the price of a zero-coupon issued in time  $t$  and maturing  $\tau$  years later and noted  $P^{(FTSM)}(t, \tau)$ .

$$P^{(FTSM)}(t, \tau) = \exp(A(\tau) + B(\tau)X_t) \quad (2.3)$$

Since Duffie and Kan [13], we know that the coefficients  $A(\tau)$  and  $B(\tau)$  are defined through the two following ODE with boundary conditions  $A(0) = B_1(0) = B_2(0) = 0$ . After some calculation, we can show that the derivatives of both coefficients are equal to:

$$\begin{aligned} B'_i(\tau) &= -e^{-\kappa_i \tau} & B_i(\tau) &= \frac{e^{-\kappa_i \tau} - 1}{\kappa_i} \\ A'(\tau) &= \sum_{i=1}^2 \left[ \mu_0^{(i)} B_i(\tau) + \frac{\sigma_i^2}{2} B_i(\tau)^2 \right] \end{aligned} \quad (2.4)$$

As we will see it, we do not need the complete analytical expression of  $A$  to compute the no-arbitrage Nelson and Siegel representation.

## 2.2 Nelson and Siegel representation

The Nelson & Siegel model [17] provides a very simple expression for the zero-coupon yields. In line with the factor term structure model, we are only interested in a two-factors model. The expression of a zero coupon yield  $y^{(NS)}(t, \tau)$

at time  $t$  and maturing at date  $t + \tau$  is the following:

$$y^{(NS)}(t, \tau) = f_t^{(1)} + f_t^{(2)} \frac{1 - \exp(-\lambda\tau)}{\lambda\tau} \quad (2.5)$$

The zero-coupon yield is expressed as the weighted sum of two factors  $f^{(i)}$ . The first weight ( $\equiv 1$ ) is usually interpreted as the level of the yield curve. The first factor  $f^{(1)}$  affects all maturities (at a given time  $t$ ) in a uniform way. A modification in this factor translates identically the whole curve.

The second one ( $\equiv \frac{1 - \exp(-\lambda\tau)}{\lambda\tau}$ ) is interpreted as the slope. This weight decreases with maturity and its limits in 0 and  $\infty$  are respectively 1 and 0. The factor  $f^{(2)}$  affects thus more the short end of the curve than the long end.

### 3 Nelson and Siegel representation of a factor term structure model

In this section, we want to find a Nelson and Siegel representation of the factor term structure model which is described in the preceding section. We expose here a general method which can be applied to a more sophisticated framework (e.g. as we will do in the three factors case). We could however apply a more straightforward calculus in this simple context.

We want to find functional coefficients  $\{f_t^{(1)}, f_t^{(2)}, f_t^{(3)} \mid t \geq 0\}$  such that prices of the zero coupon with the factor term structure model and the Nelson and Siegel one are the same. Using the equations (2.3) and (2.5), the problem is summarized in the following equality (we suppose that  $\kappa_1$  and  $\kappa_2$  are not equal to 0) :

$$\tau y^{(NS)}(t, \tau) + A(\tau) + B(\tau)^\top X_t = 0 \quad (3.1)$$

Because we cannot explicitly compute the coefficient  $A(\tau)$ , we focus on the derivation of the equalities (3.1) relative to  $\tau$ , which is much more tractable. If we use equations (2.4) and (2.5), the derivation of (3.1) leads to the following equality:

$$\begin{aligned} 0 = & f_t^{(1)} - \left( \frac{\mu_0^{(1)}}{\kappa_1} + \frac{\mu_0^{(2)}}{\kappa_2} \right) + \frac{\sigma_1^2}{2\kappa_1^2} + \frac{\sigma_2^2}{2\kappa_2^2} \\ & + e^{-\lambda\tau} f_t^{(2)} + \sum_{i=1}^2 e^{-\kappa_i\tau} \left( \frac{\mu_0^{(i)}}{\kappa_i} - \frac{\sigma_i^2}{\kappa_i^2} - x_t^{(i)} \right) + \sum_{i=1}^2 e^{-2\kappa_i\tau} \frac{\sigma_i^2}{2\kappa_i^2} \end{aligned} \quad (3.2)$$

We now want to identify coefficients to get the no-arbitrage Nelson and Siegel representation. We distinguish two cases regarding the number of terms in the Nelson & Siegel representation. In the first case, we constraint our model to take into account only models with a constant term ( $\equiv 1$ ) and an exponential one ( $\equiv \frac{1 - \exp(-\lambda\tau)}{\lambda\tau}$ ). Unfortunately, those representations always lead to trivial term structures. Obviously we cannot consider them as satisfying term structure modeling. To cope with this issue, we must add terms to catch properly no-arbitrage conditions. As we will see it, our results will be in line with the five factors extension of standard N&S representation by Björk and Christensen [3].

Let us begin with a strict two factors N&S representation.

### 3.1 A strict two–factors Nelson & Siegel representation

We now want to identify factors  $f^{(i)}$  in the equation (3.2) without adding any term. We can firstly notice that if real coefficients  $(a_i)_{i=1\dots n}$  are different by pairs, the family  $(e^{a_i})_{i=1\dots n}$  is independent. The identification process is thus very easy. We need however to distinguish cases regarding the value of  $\kappa_i$  and  $\lambda$ . We suppose that  $\lambda$  is strictly positive.

- (i)  $\kappa_i, 2\kappa_i, \lambda$  different by pairs. The independence of exponential functions implies that factors  $x_t^{(i)}$  are equal to 0 and that the term structure is flat and equal to  $x_0^{(1)} + x_0^{(2)}$ .
- (ii)  $2\kappa_i = \lambda$  The process  $x_t^{(i)}$  is necessarily equal to 0. The yield curve is again constant.
- (iii)  $\kappa_i = 0$ . The equation (3.2) is not more valid but (2.4) is still. A straightforward calculus shows that we need again to have  $x_t^{(i)} \equiv 0$ . The term structure is again constant.
- (iv)  $\kappa_i = \lambda$ . In this first case, the process  $x_t^{(i)}$  is determinist ( $\sigma_i = 0$ ). More precisely, we get the following result:

$$\begin{aligned} x_t^{(i)} &= x_0^{(i)} + \mu_0^{(i)} t \\ f_t^{(1)} &= \frac{\mu_0^{(1)} + \mu_0^{(2)}}{\lambda} \\ f_t^{(2)} &= \sum_{i=1}^2 \left( x_t^{(i)} - \frac{\mu_0^{(i)}}{\lambda} \right) \end{aligned}$$

Even if the term structure is not more constant over time, this case remains undoubtedly uninteresting. The zero–coupon yield curve is still completely determinist and therefore useless for financial and economic purposes.

The restriction to a simple two factors representation always leads to poor term structures. These are in the best configuration determinist and even often constant. The factor term structure model is degenerated and is out of purpose. As a consequence, we cannot properly take into account no–arbitrage constraints in a simple two–factors N&S representation. We need to add terms, which will help to overcome preceding drawbacks. Those terms aim at avoiding zero identification. As we will see it, those terms have already been introduced by Björk and Christensen [3] in their 5 factors N&S extension.

### 3.2 An extended N&S representation

#### 3.2.1 Motivations

As we explained in the preceding paragraph, we wish we got a no–arbitrage N&S representation, which won't be simply determinist or even worse constant. This goal cannot be reached with a simple two factors N&S model, as described in equation (2.5). To cope with this issue, we can consider supplementary terms in our initial representation. Considering the equation (3.2) and the calculus

done in the preceding paragraph, we need terms in  $e^{-2\lambda\tau}$ ,  $\tau$  and  $\tau^2$  for the identification of the equality (3.2). This implies notably terms in  $\frac{1-e^{-2\lambda\tau}}{2\lambda\tau}$ ,  $\tau$  and  $\tau^2$  in the original Nelson & Siegel equation (2.5). These belong exactly to the terms chosen by Björk and Christensen [3] in their extension.

With this new terms, the extended Nelson & Siegel representation becomes now:

$$y_t^{(\tau)} = f_t^{(1)} + \frac{1 - e^{-\lambda\tau}}{\lambda\tau} f_t^{(2)} + \frac{1 - e^{-2\lambda\tau}}{2\lambda\tau} f_t^{(3)} + \tau f_t^{(4)} + \tau^2 f_t^{(5)} \quad (3.3)$$

Aruoba, Diebold and Rudebusch [11] explain that these factors do not improve the calibration on zero-coupon yield curves. Our estimations will not infrm this point. However, as we will see it, they will be of great help in the risk premium modeling and in the Expectation Hypothesis rejection. Those terms do not improve the yield fit (but do not deter it neither) but help the model to pass other important tests. For example, the risk premium modeling (and the rejection of the Expectation Hypothesis) is crucial to compare portfolio strategies. The use of a standard N&S seems to be of poor help in this purpose, whereas no-arbitrage one becomes meaningful.

### 3.2.2 Factors expressions

We investigate the computation of factors  $f^{(i)}$ . We need to modify the equation (3.2) to take into account our supplementary terms. For sake of precision, we consider in our expression the possibility that  $\kappa_i = 0$ . In this case, the equations (2.4) are not more valid, since we have  $B_i(\tau) = -\tau$ . We note  $\mathbb{1}_A$  the function equal to 1 if the event  $A$  is true and 0 else.

We can now generalize the equation (3.2) as follows:

$$\begin{bmatrix} f_t^{(1)} \\ f_t^{(2)} \\ f_t^{(3)} \\ f_t^{(4)} \\ f_t^{(5)} \end{bmatrix} \begin{bmatrix} 1 \\ e^{-\lambda\tau} \\ e^{-2\lambda\tau} \\ \tau \\ \tau^2 \end{bmatrix} = \sum_{i=1}^2 \begin{bmatrix} 1 \\ e^{-\kappa_i\tau} \\ e^{-2\kappa_i\tau} \\ \tau \\ \tau^2 \end{bmatrix} \cdot \begin{bmatrix} \left[ \frac{\mu_0^{(i)}}{\kappa_i} - \frac{\sigma_i^2}{2\kappa_i^2} \right] \mathbb{1}_{\kappa_i \neq 0} + x_t^{(i)} \mathbb{1}_{\kappa_i = 0} \\ - \left[ \frac{\mu_0^{(i)}}{\kappa_i} - \frac{\sigma_i^2}{\kappa_i^2} - x_t^{(i)} \right] \mathbb{1}_{\kappa_i \neq 0} \\ - \frac{\sigma_i^2}{2\kappa_i^2} \mathbb{1}_{\kappa_i \neq 0} \\ \frac{\mu_0^{(i)}}{2} \mathbb{1}_{\kappa_i = 0} \\ - \frac{\sigma_i^2}{6} \mathbb{1}_{\kappa_i = 0} \end{bmatrix} \quad (3.4)$$

This is the complete system of equations defining the no-arbitrage N&S representation. Because we assume that  $\lambda > 0$ , we only have three possibilities to conclude (out of which two are symmetric):  $\kappa_1 = \kappa_2 = \lambda$ ;  $\kappa_1 = 0$  and  $\kappa_2 = \lambda$  and vice-versa.

**A one-factor model** First, we can choose  $\kappa_1 = \kappa_2 = \lambda$ , which leads to the following Nelson and Siegel factors:

$$\begin{aligned} f_t^{(1)} &= \frac{\mu_0^{(1)} + \mu_0^{(2)}}{\lambda} - \frac{\sigma_1^2 + \sigma_2^2}{2\lambda^2} \\ f_t^{(2)} &= x_t^{(1)} + x_t^{(2)} + \frac{\sigma_1^2 + \sigma_2^2}{\lambda^2} - \frac{\mu_0^{(1)} + \mu_0^{(2)}}{\lambda} \\ f_t^{(3)} &= -\frac{\sigma_1^2 + \sigma_2^2}{2\lambda^2} \\ f_t^{(4)} &= 0 \end{aligned}$$

We only add a supplementary term in the N&S representation, whose weight is  $\frac{1-e^{-2\lambda\tau}}{2\lambda\tau}$ . Moreover, the factor ( $\equiv -\frac{\sigma_1^2 + \sigma_2^2}{2\lambda^2}$ ) is constant overtime. The two factors driving the term structure remain  $x_t^{(1)}$  and  $x_t^{(2)}$ . However, both of them affects only the slope weight. This model is equivalent to a one-factor model. We lose the dependency in one factor (the level is constant) and the richness of the yield patterns are affected. This model is simply a version of classical Vasicek model [18].

We can indeed set  $\mu^{(0)} = \mu_1^{(0)} + \mu_2^{(0)}$ ,  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ . Then  $x_t = x_t^{(1)} + x_t^{(2)}$  is an Ornstein–Ülhenbeck process<sup>1</sup> with volatility  $\sigma$ , mean  $\mu$  and mean reverting speed  $\lambda$ .

**A two-factors model** Second, as in Diebold, Piazzesi and Rudebusch, we choose that  $\kappa_1 = 0$  and  $\kappa_2 = \lambda$  (the vice versa choice would of course have been completely equivalent). In this case we get an extended five-factors extension of the standard N&S. None of weights is equal to 0.

This case is more satisfying than the preceding one because it allows to deal not only with a single parameter but with two independent parameters. The factor  $x^{(1)}$  drives the level and the factor  $x^{(2)}$  the slope. Consequently, we have been able to get a no-arbitrage N&S representation for a two-factors model. This representation is empirically comparable to the standard N&S version and we will do it in the next section.

To sum up in a couple of equations this new representation, we present the factors dynamics (under the historical probability) and the coefficients  $f^{(i)}$ :

$$\begin{aligned} dx_t^{(1)} &= -\mu_1^{(1)} x_t^{(1)} dt + \sigma_1 d\widetilde{W}_t^{(1)} & f_t^{(1)} &= x_t^{(1)} - \frac{\sigma_2^2}{2\lambda^2} + \frac{\mu_2^{(0)}}{\lambda} \\ dx_t^{(2)} &= -(\mu_2^{(1)} + \lambda) x_t^{(2)} dt + \sigma_2 d\widetilde{W}_t^{(2)} & f_t^{(2)} &= x_t^{(2)} + \frac{\sigma_2^2}{\lambda^2} - \frac{\mu_2^{(0)}}{\lambda} \\ & & f_t^{(3)} &= -\frac{\sigma_2^2}{2\lambda^2} \\ & & f_t^{(4)} &= \frac{\mu_1^{(0)}}{2} \\ & & f_t^{(5)} &= -\frac{\sigma_1^2}{6} \end{aligned} \quad (3.5)$$

<sup>1</sup>Whose associated BM is  $\frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} W_t^{(1)} + \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} W_t^{(2)}$ .

**Interpretation** We want to understand to what extent no-arbitrage modifies the N&S representation. For sake of simplicity let us write the expression of a zero-coupon yield  $y^{(NANS)}(t, \tau)$  of maturity  $\tau$  at time  $t$  according to both representations.

$$y^{(NANS)}(t, \tau) = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \frac{1-e^{-2\lambda\tau}}{2\lambda\tau} \\ \tau^2 \end{bmatrix}^\top \begin{bmatrix} -\frac{\sigma_2^2}{2\lambda^2} \\ \frac{\sigma_2^2}{\lambda^2} \\ -\frac{\sigma_2^2}{2\lambda^2} \\ -\frac{\sigma_1^2}{6} \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \tau \end{bmatrix}^\top \begin{bmatrix} \frac{\mu_2^{(0)}}{\lambda} \\ -\frac{\mu_2^{(0)}}{\lambda} \\ \frac{\mu_1^{(0)}}{2} \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} \quad (3.6)$$

$$y^{(NS)}(t, \tau) = \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} \quad (3.7)$$

Before any interpretation, one has to keep in mind that dynamics of  $x^{(i)}$  is different in each case. In the no-arbitrage model, the level factor is a random walk under the risk neutral measure. The slope one is a mean reverting process toward 0 and with mean reverting speed  $\lambda$  (always under the risk neutral measure), the same as for the weight terms in exponential. We can notice that in this representation, the convergence speed toward the long term value of the weight and the associated factor (considered under the risk neutral measure) are equal. The level weight (1) and its factor (RW) both have 0 mean reverting speed. For slope,  $\lambda$  is the common value. In the standard N&S model, both factors are also mean reverting processes (under the historical probability, because we do not take care of the risk neutral one.) toward 0 but without any constraint on the mean reverting speed.

The expression (3.6) is the sum of three different terms. The last one is analogous to the standard Nelson and Siegel one in (3.7) and refers to the level and slope of the yield curve. It is the only term depending on date  $t$  through the channel of processes  $x^{(i)}$ . The main difference with standard expression consists in the dynamics of processes under the risk neutral probability.

The other terms in (3.6) and (3.7) do not depend on time  $t$  and deform the yield curve in an uniform way, independently of time. In the standard model, this deformation is simply bivariate and consists in the sum of a constant deformation ( $\theta_1$ ) of the level and another constant one ( $\theta_2$ ) for the slope. Unfortunately, we can hardly go further in the interpretation. In the no-arbitrage representation, two terms deform uniformly the yield curve, independently of the time  $t$ . The first one is a Jensen term whereas the second one reflects the constant part of the risk premium and thus proportional to  $\mu^{(0)}$ . This zero-coupon yield expression offers consequently a straightforward economic interpretation.

In the estimation section, we will compare impacts of different terms on the yield curve deformation.

## 4 Empirical Estimations

We want to measure how the no–arbitrage conditions modify the modeling of the yield curve with Nelson and Siegel factors. We want to investigate this question regarding first the yield curve fitting, the modeling of the risk premium and the forecast performance of each model. We need estimating properly both models.

For the data, we use Fama and Bliss discount factors, i.e. monthly nominal zero coupon prices from June 30, 1952 to December 31, 1998 with maturities of 3, 6, 12, 24, 60 and 120 months.

For the estimation we do not follow Chen and Scott technique [5] but we use instead Kalman filter and maximum likelihood estimation. This was also the choice of Dai and Philippon[6]. Let us briefly call back the main reasons of our choice.

In the Chen and Scott technique, one needs to assume that certain yields are perfectly priced and do not suffer on the whole sample from any measurement error. Using these yields, you can compute the dynamics of your latent factors. Assuming then that other yields are priced with IID errors, you can compute the likelihood and, through maximization, estimate parameters. The main issue of this technique lies in the selection of perfectly priced yields. Up to my knowledge, there is no criteria to select them, except the feeling of the author.

That is why we have decided to use maximum likelihood with Kalman filter. We suppose that all yields are priced with IID errors. The selection issue disappears. This technique allows therefore a more natural trade–off between times–series properties of factors (i.e.  $x_t^{(i)}$ ) and the cross sectional fit of different maturities. We describe in details the estimation procedure for the standard representation. The technique is absolutely analogous in the no–arbitrage framework.

### 4.1 Standard N & S model

We describe our estimation technique for the standard N&S model.

Let us begin with some notations and definitions.  $\tau_1, \dots, \tau_N$  are the maturities available in our dataset, i.e. 3, 6, 12, 24, 60 and 120 months, meaning that  $N = 6$ .  $Y^{(NS)}(t)$  is the vector gathering zero–coupon yields of different maturities. We note  $\varepsilon_t$  the  $6 \times 1$  vector of yields measurement errors, whose covariance matrix  $\Sigma_\varepsilon^2 = \text{diag}(s_1^2, s_2^2, s_3^2, s_4^2, s_5^2, s_6^2)$  is supposed to be diagonal.  $X_t^{(NS)}$  is the vector of factors. We therefore suppose that shocks affecting various maturities are independent. One of the justifications behind this is that main causes of shocks are a priori independent. First, zero–coupon yields comes from interpolation of traded bond yields. Second, this errors can also be considered as a tool to overcome the preceding selection problem. This can be done with independent shocks.

For estimation purpose, we discretize the dynamics.  $\eta_t$  is the  $2 \times 1$  vector of shocks affecting both factors. We assume that the variance matrix is diagonal and equal to  $\text{diag}(\sigma_1^2, \sigma_2^2)$ . This in line with the assumption telling that both factors are independent.

The factors mean reverting matrix  $\Lambda = \text{diag}(1 - \kappa_1, 1 - \kappa_2)$  is diagonal. We note  $B$  the  $N \times 2$  matrix of factors weights and equal to  $\left[1 \quad \frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i}\right]_{i=1 \dots N}$ . Finally,  $A$  is the constant term deforming the yield curve:  $B [\theta_1 \quad \theta_2]$ .

The set of parameters we need to estimate in this model is the following one:  $\Theta^{(NS)} = \{\lambda, \kappa_1, \kappa_2, \theta_1, \theta_2, \sigma_1, \sigma_2, s_1, s_2, s_3, s_4, s_5, s_6\}$ .

With those notations, the state-space system expresses as follows:

$$Y_t^{(NS)} = A^{(NS)} + B^{(NS)} X_t^{(NS)} + \varepsilon_t^{(NS)} \quad (4.1)$$

$$X_t^{(NS)} = \Lambda^{(NS)} X_{t-1}^{(NS)} + \eta_t^{(NS)} \quad (4.2)$$

$$(4.3)$$

We use the Kalman filter techniques to compute the best forecast  $Y_{t+1|t}$  of  $Y_{t+1}$  at time  $t$  in function of factors  $X_t$ . We can therefore determine the distribution of  $Y_{t+1|t}$  knowing  $X_t$ . By recursion, we can then compute the total likelihood. You can find in appendix all the related technical stuff.

The estimation of the likelihood is performed through two algorithms: (i) Berndt–Hall–Hall–Hausman algorithm [2] and (ii) Levenberg–Marquardt algorithms [14] and [15]. Both are combined to modify traditional Newton–Raphson algorithms. The first one uses the equality, for the likelihood function, between the Hessian matrix and the opposite of the gradient covariance matrix. The equality is not valid for all values of parameters but only at the equilibrium value and we get then the Fisher information. This algorithm tells that this property remains true for values not too far from the equilibrium and that through a Newton–Raphson algorithm we converge toward the true value. The second algorithm modifies the step between each recursion of the numerical algorithm. Instead of simply subtracting the product of the Hessian inverse and the gradient, we multiply it with a constant so as to maximize the likelihood at this recursion step.

Precisions regarding the numerical techniques in the appendix ?

The table (TAB. 1) gathers estimation results for the parameters values and associated standard errors. Our estimation performs well and all coefficients are accurately estimated. Standard errors remain weak and all parameters are significant.

## 4.2 No–arbitrage N & S model

The estimation procedure is absolutely identical to the preceding one. In fact, we need very few modifications to take into account no–arbitrage constraints. Looking at zero–coupon yields expressions (3.5) and factors dynamics (3.6), it is rather intuitive to modify the standard N&S estimation equations and get the following state–space system:

$$Y_t^{(NANS)} = A^{(NANS)} + B^{(NANS)} X_t^{(NANS)} + \varepsilon_t^{(NANS)} \quad (4.4)$$

$$X_t^{(NANS)} = \Lambda^{(NANS)} X_{t-1}^{(NANS)} + \eta_t^{(NANS)} \quad (4.5)$$

$$(4.6)$$

As you can notice it, the structure of both estimations is identical and differences are thin. Computational costs as well as conceptual ones are in both cases

Parameters	Values $\times 10^{-2}$	Standard Errors $\times 10^{-5}$
$s_1$	0.439	0.123
$s_2$	0.204	0.067
$s_3$	0.048	0.051
$s_4$	0.106	0.027
$s_5$	0.043	0.058
$s_6$	0.183	0.052
$\lambda$	5.618	0.527
$\kappa_1$	1.025	3.833
$\kappa_2$	5.722	7.965
$\theta_1$	7.241	13.316
$\theta_2$	-1.276	4.197
$\sigma_1$	0.342	0.073
$\sigma_2$	0.564	0.084
<b>ML</b>	37273	0.015

Table 1: Estimation of the standard Nelson and Siegel model

of comparable magnitude. The set of parameters we need to estimate is now the following:  $\Theta^{(NANS)} = \{\lambda, \sigma_1, \sigma_2, \mu_1^{(0)}, \mu_2^{(0)}, \mu_1^{(1)}, \mu_2^{(1)}, s_1, s_2, s_3, s_4, s_5, s_6\}$ . We can notice that in no-arbitrage case the number of parameters to estimate is strictly equal to the one in standard N&S model. The no-arbitrage conditions allow us to decrease the number of 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.

As in the preceding case, we compute recursively the likelihood and its maximization leads to the estimations gathered in table (TAB. 2) . As in the standard model estimation, all the parameters are significant.

Parameters	Values $\times 10^{-2}$	Standard Errors $\times 10^{-5}$
$s_1$	0.439	0.137
$s_2$	0.210	0.063
$s_3$	0.008	0.797
$s_4$	0.128	0.034
$s_5$	0.103	0.045
$s_6$	0.137	0.056
$\lambda$	4.480	0.508
$\sigma_1$	0.235	0.056
$\sigma_2$	0.775	0.042
$\mu_1^{(0)}$	0.045	0.508
$\mu_2^{(0)}$	0.046	0.015
$\mu_1^{(1)}$	0.708	0.345
$\mu_2^{(1)}$	1.421	2.290
<b>ML</b>	37287	0.000

Table 2: Estimation of the no-arbitrage N&S model

We can notice that the maximum likelihood is greater in the no–arbitrage case than in the standard one, while the number of estimated parameters is the same.

### 4.3 Comparisons of both models

We want to compare both preceding models in terms of bond pricing and of risk premium modeling. The number of parameters is the same in both models and the likelihood is greater for the no–arbitrage representation. We expect therefore the no–arbitrage model to perform better than the standard one, especially what concerns the risk premium modeling and long maturities (notably because of Jensen terms and risk premium definition).

#### 4.3.1 Yield curve fitting

The following table (TAB. 3) gathers spreads between effective rates and real yields. The differences are expressed in basis points (bp = 0.01%). In parentheses, you can find standard errors coming from the estimation uncertainties.

<b>Spreads between model and reality (bp)</b>		
Maturity	N&S model	NA–N&S model
3	43.44 (0.0052)	43.68 (0.0098)
6	19.67 (0.0041)	20.91 (0.0042)
12	2.21 (0.0057)	0.01 (0.0048)
24	10.18 (0.0010)	12.56 (0.0031)
60	1.93 (0.0064)	8.60 (0.0057)
120	17.57 (0.0048)	11.19 (0.0073)
<b>Mean</b>	15.84	16.16

Table 3: Spreads between modeled and actual yields

Both models perform in a similar way and differences between both modelings are in terms of yields modeling not really significant. On average, these are about 0.3 bp.

One of the major advantage of standard Nelson and Siegel model is to fit particularly well the yield curve. We can notice that the no–arbitrage property does not deter this positive feature.

#### 4.3.2 Variance decomposition

We strengthen this affirmation by comparing the shares of yield variances explained by factors of each model at a given horizon. The table (TAB. 4) gathers results. For each model and for each maturity, we express which share of the total zero–coupon yield variance is explained by the first factor, by the second

one and which share remains unexplained. One can find in appendix the proper derivations for variances decompositions.

As we can see in the results array, the proportion of yield curve variance explained by the no-arbitrage model is always greater than the one explained by the standard representation. The maturity forecast does not influence this feature. We can also notice that the explanatory power of both models increases with forecast maturities in comparable proportions. The share of unexplained variances sinks from 8.08% for the NA-N&S model (resp. 10.96% for the standard N&S model) for a one month forecast down to 0.48% for a 10 years forecast (resp. 0.62%) and even 0.21% (resp. 0.59%) for a very long forecast, which is in fact the unconditional variance.

Let us look more into details. The explanatory power varies a lot regarding the considered yield. Models perform better for the long end of curve and especially the middle end, which means maturities between 1 year and 5 years. The average unexplained variance is 0.64% for the NA-N&S model (resp. 0.44% for the standard one), whereas for the remaining of the yield curve, which is the short end and the long end, this share is 3.63% (resp. 5.25%). The standard N&S model does slightly better than the no-arbitrage one for the middle end (especially for 2 years maturity) of the curve but the margin is very thin.

For both models, the 3 months zero-coupon yield is much less explained than the other ones. The unexplained variance is as high as 6.60% (9.01% for the standard N&S) for 3 months yields, whereas the average for other yields is not larger than 1.24% (resp. 1.61%). Explanatory power of both models is less for the short end (3 months and six months zero-coupon) of the yield curve, but the no-arbitrage model does better than the standard one. The 10 years zero-coupon yield is more interesting. For the standard N&S, it is the second worst maturity, just after the 3 months one. For the no-arbitrage model, the result is more moderated. For short forecasts (1 and 12 months in particular), the 10 years residual variance is greater than the 6 months one but for longer forecasts (typically longer than 5 years), the result is inverted. The no-arbitrage as well as Jensen terms help to improve the explanatory power for long forecast and long maturities.

Let us now focus on the role of both factors. The explanatory power of the first factor (interpreted as the level of the yield curve) increases with the forecast maturity from on average 14.8% (resp. 38.91%) on the yield curve for the no-arbitrage (resp. standard) N&S model for a one month forecast to 84.63% (resp. 78.13%) for unconditional variance. On the other side, the explanatory power of the second factor (i.e. the slope) decreases with the forecast maturities from 77.12% (resp. 50.42%) down to 15.16% (resp. 21.29%). The slope accounts more for explaining short term forecasts whether the level has a more important role on longer forecasts.

Globally, for a given forecast horizon, the explanatory power of the level increases with yield maturity whereas this of the slope decreases. For instance, let us consider the unconditional variance. For the no-arbitrage (resp. standard) N&S, the level explanation goes up from 77.55% (resp. 67.01%) for 3 months yields up to 94.68% (resp. 92.62%) for 10 years yields. On the other hand, the one of the slope sinks from 21.61% (resp. 30.74%) down to 5.22% (resp. 6.84%). On the whole, level is important to explain long-term forecast and to explain long-end of the yield curve. On the contrary, slope is important to explain short forecast and to explain short-end of the yield curve. We need however

Shares of yield variances						
Maturity (months)	NA-N&S			Standard N&S		
	First factor	Second factor	Unexp.	First factor	Second factor	Unexp.
<i>1 month forecast</i>						
3	6.81	69.40	23.79	19.40	48.59	32.01
6	8.81	84.12	7.06	27.27	63.00	9.74
12	10.62	89.38	0.00	33.34	66.00	0.66
24	12.56	83.73	3.71	38.59	57.67	3.74
60	20.14	75.99	3.87	55.69	43.42	0.88
120	29.85	60.12	10.04	59.17	23.86	16.97
<b>Mean</b>	14.80	77.12	8.08	38.91	50.42	10.67
<i>12 months forecast</i>						
3	14.11	81.72	4.18	36.47	57.93	5.60
6	15.40	83.56	1.05	40.02	58.65	1.33
12	17.28	82.72	0.00	44.29	55.62	0.08
24	20.76	78.72	0.52	51.10	48.44	0.46
60	31.63	67.85	0.52	66.84	33.06	0.10
120	46.00	52.69	1.31	77.97	19.94	2.08
<b>Mean</b>	24.20	74.54	1.26	52.78	45.61	1.61
<i>60 months forecast</i>						
3	37.20	60.43	2.37	59.01	38.19	2.79
6	39.42	60.00	0.58	62.22	37.14	0.64
12	42.69	57.31	0.00	66.13	33.83	0.04
24	48.34	51.40	0.26	72.00	27.80	0.20
60	62.30	37.48	0.22	83.20	16.76	0.04
120	75.34	24.20	0.46	89.89	9.37	0.74
<b>Mean</b>	50.88	48.47	0.65	72.08	27.18	0.74
<i>120 months forecast</i>						
3	51.98	46.21	1.81	65.02	32.60	2.38
6	54.32	45.25	0.43	68.00	31.46	0.54
12	57.64	42.36	0.00	71.59	28.38	0.03
24	63.10	36.72	0.19	76.84	22.99	0.17
60	75.12	24.73	0.14	86.47	13.50	0.03
120	84.81	14.91	0.28	91.99	7.43	0.59
<b>Mean</b>	64.50	35.03	0.48	76.65	22.73	0.62
$\infty$						
3	77.55	21.61	0.85	67.01	30.74	2.25
6	79.14	20.66	0.20	69.91	29.58	0.51
12	81.28	18.72	0.00	73.37	26.61	0.03
24	84.51	15.41	0.08	78.39	21.46	0.15
60	90.60	9.35	0.05	87.48	12.49	0.03
120	94.68	5.22	0.10	92.62	6.84	0.54
<b>Mean</b>	84.63	15.16	0.21	78.13	21.29	0.59

Table 4: Shares of variances

to be more cautious when explaining the role of the level, especially for short forecasts, because its explanatory power (for both models) is hump-shaped. For short forecasts, the slope explains more the middle-end of the curve than the short-end. These values correspond also to high unexplained variance shares, meaning that there is a lack in models for short term explanation of the short-end of the yield curve. Because it is a common feature to both models, we can assume that the reason of it is either the lack of a third factor (i.e. curvature) or of crossed terms in factors dynamics. My intuition is the lack of a third factor.

Let us now look into the modeling of the risk premium.

### 4.3.3 Comparison of risk premium modeling

After the comparison of the yield curve modeling, we want to show which specification leads to a proper risk premium. Therefore, we will focus on the Expectation Hypothesis test, as done for example in Campbell and Shiller [4] or Dai and Singleton [8].

We will run the following regressions for all  $n \in \{6, 12, 24, 60, 120\}$ :

$$y_{t+1}^{(n-3)} - y_t^{(n)} = a_n + b_n \frac{y_t^{(n)} - y_t^{(3)}}{n-3} + \varepsilon_t \quad (4.7)$$

Notice that  $n$  and  $t$  are expressed in months and that the minimum maturity we have in our dataset is 3 months.

We use the estimation procedure described in Dai and Singleton [8]. This investigation method for the Expectation Hypothesis tests if the model parameters (estimated through ML) themselves generate the correct pattern for the regression coefficient  $b_n$ . In fact, we do not want to test if the in-sample estimated yields exhibit this property. We therefore use Monte-Carlo simulations. For each model, we generate 1000 scenarios. Each of them is a path of zero-coupon yields simulated by the model. As in our sample (going from 1952, January to 1998, December), each path has 444 yields. Then we run our regression for each path and compute average and standard errors for each coefficient.

This coefficient must be in line with Campbell and Shiller [4] results. It should be always negative and should exhibit a downward sloping pattern with maturities.

As we can see in the table (TAB. 5), both models perform this test relatively well. Both exhibit a negative and downward sloping pattern. The implicit risk premium modeling in the standard Nelson and Siegel representation is therefore not too bad. However, the modeling of the no-arbitrage one is significantly better than the standard one, for all maturities and especially for long ones (3, 5 and 10 years). The difference between both models increases with maturity. The longer the maturity, the better the no-arbitrage N&S model relative to the standard one.

The explicit risk premium modeling is therefore better than the implicit one, even if we restrict to our particular case of a linear risk premium. The graph (FIG. 4.1) plots regression coefficients of (4.7) with their respective 95% confidence intervals (dashed lines). It illustrates to which extent no-arbitrage representation performs better than the standard one.

From a heuristic point of view, the actual regression coefficients almost lie inside the 95% confidence interval of the no-arbitrage model. On the other

Expectation Hypothesis Regressions			
Maturity	Data	Standard N&S	NA-N&S
6	-2.81	-0.57 (0.19)	-0.67 (0.25)
12	-4.97	-1.18 (0.50)	-2.20 (0.71)
24	-9.56	-5.22 (0.89)	-7.30 (0.86)
60	-22.83	-17.34 (1.51)	-22.24 (1.16)
120	-24.91	-18.16 (1.50)	-22.67 (1.04)

Table 5: Expectation Hypothesis Regressions

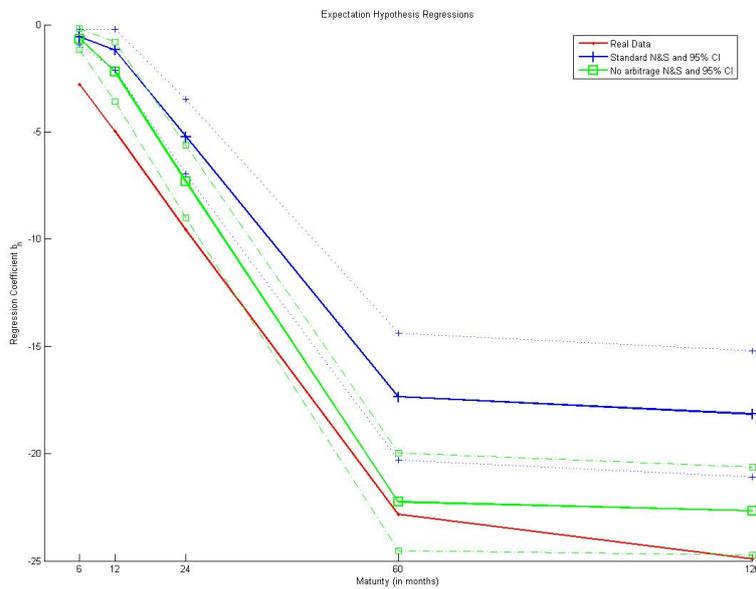


Figure 4.1: Regression coefficients (4.7) of  $V_s$  vs. maturities

side, confidence intervals of standard N&S and no–arbitrage one have a very thin intersection. If not a test, this illustrates how well the no–arbitrage model help to perform the rejection of the Expectation Hypothesis, especially for long maturities.

#### 4.3.4 Detailed comparisons of both representations

We want to understand what are the differences between both representations. We will begin with comparing the dynamics of both factors and then the impact of Jensen and constant risk premium terms as defined in expression (3.6).

For reading convenience, we sum up relevant expressions in both cases:

$$\begin{aligned}
y^{(NANS)}(t, \tau) = & \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \frac{1-e^{-2\lambda\tau}}{2\lambda\tau} \\ \tau^2 \end{bmatrix}^\top \cdot \begin{bmatrix} -\frac{\sigma_2^2}{2\lambda^2} \\ \frac{\sigma_2^2}{\lambda^2} \\ -\frac{\sigma_2^2}{2\lambda^2} \\ -\frac{\sigma_1^2}{6} \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} \\
& + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \tau \end{bmatrix}^\top \cdot \begin{bmatrix} \frac{\mu_2^{(0)}}{\lambda} \\ -\frac{\mu_2^{(0)}}{\lambda} \\ \frac{\mu_1^{(0)}}{2} \end{bmatrix} \tag{4.8}
\end{aligned}$$

$$X_t^{(NANS)} = \Lambda^{(NANS)} X_{t-1}^{(NANS)} + \eta_t^{(NANS)} \tag{4.9}$$

$$\begin{aligned}
y^{(NS)}(t, \tau) = & \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \end{bmatrix}^\top \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} \\
X_t^{(NS)} = & \Lambda^{(NS)} X_{t-1}^{(NS)} + \eta_t^{(NS)} \tag{4.10}
\end{aligned}$$

**Factors dynamics:** We compare factors dynamics in both representations, notably mean reverting speeds and standard errors. We begin with the mean reverting speeds, i.e. the diagonal terms of  $(I_2 - \Lambda^{(NS)})$  and  $(I_2 - \Lambda^{(NANS)})$ .

Mean reverting speeds $(\times 10^{-2})$		
	Standard N&S	No–arbitrage N&S
First factor	1.021	0.620
$\times 10^{-5}$	(5.627)	(2.831)
Second factor	5.633	5.481
$\times 10^{-5}$	(11.696)	(20.753)

In both cases, the first factor is the most persistent one but it is more persistent in the no–arbitrage case. On the other side, mean reverting speeds of the second factor are of comparable magnitude in each representation.

Let us now turn to the volatilities of factors shocks.

Factors Volatilities $(\times 10^{-3})$		
	Standard N&S	No-arbitrage N&S
First factor	3.417	2.338
$\times 10^{-5}$	(0.107)	(0.061)
Second factor	5.637	7.678
$\times 10^{-5}$	(0.123)	(0.222)

(4.12)

Factor volatilities are of the same order of magnitude in both cases. The first factor is less volatile in the no-arbitrage model whereas it is the opposite for the second factor.

If not equal, factors dynamics are comparable. As we can read in the table (TAB. 6), the financial interpretation of these factors in both representations is completely analogous. The correlation of first factors with the level factor (expressed as the long term zero-coupon rate, i.e. the 10Y in our case) is very high, even greater than 99%. Regarding second factors, the correlation with the level factor (expressed as the difference between the short rate, 3 months rate and the long rate, 10Y rate) is also very high in both cases, that is greater than 96%.

Correlations (%)		
	Standard N&S	No-arbitrage N&S
Level and first factor	99.45	99.18
Slope and second factor	96.46	96.04

Table 6: Correlations between models factors and yield curve moments

The no-arbitrage conditions does not deter significantly the economic interpretation of factors. The usual meanings of level and slope remain strongly.

**Constant terms in the zero-coupon yields** For a description purpose, it is useful to compare the constant terms in the standard N&S (i.e. terms depending on  $\theta_*$ ) and the no-arbitrage one (i.e. Jensen terms and terms in  $\mu_*^{(0)}$ ). Those terms are independent of the date of the term structure. They only depend on the maturity of the zero-coupon rate. The graph (FIG. 4.2) plots constant terms (normalized at 0 for maturity 1 month) for both models.

**Risk premium modeling and Jensen terms** We chose to model a linear risk premium depending on the factors  $X_t$ . We give its expression for reading convenience:  $\Lambda_t = \mu^{(0)} + \mu^{(1)}X_t$ . The two factors of the risk premium present patterns which are very similar to the level and slope of the term structure. The correlation between the level and the first factor of the risk premium is 99.21% and the correlation between slope and second factor is 95.75%. We interpret it as follows. Each component of the risk premium reflects a risk which is linked either to the level or to the slope. Our definition of the risk premium allows us to separate the risk lying of the level and the one on the slope. The level risk premium varies between 4.5 and 6.5 bp with a mean od 5.4 bp. The slope risk premium varies between -2 and 14 bp. Figures (FIG. 4.3) plot the evolution of these risk premia across time.

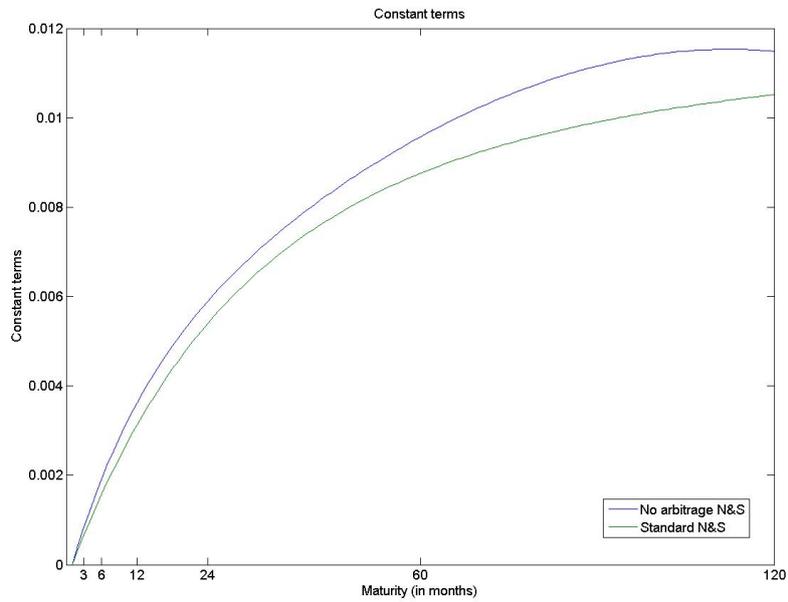


Figure 4.2: Constant terms in both models

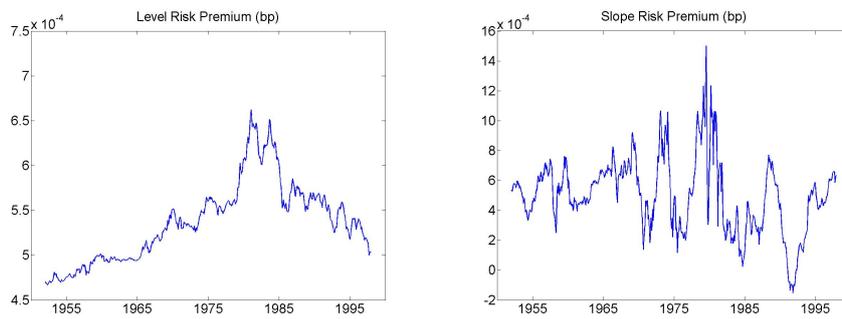


Figure 4.3: Risk premium components

## 5 Forecast performance

In this section, we want to investigate the out-of-sample performance of our models. We compute one-month forecast for the last 60 months. The procedure is the following. We estimate both models on a given sample and then compute the one month out-of-sample forecast for the 6 yields of our dataset (from 3 to 120 months). Then we compute the difference between the forecasts and the actual yields. Two criteria finally allow us to measure the performance of forecasts. Those criteria are the Root Mean Squared Error (RMSE) and the Mean Absolute Deviation (MAD). Those criteria are computed for the 6 maturities of our dataset. If we note  $\hat{y}(t, \tau)$  the model forecast of zero-coupon yield at time  $t$  and of maturity  $\tau$  noted  $y(t, \tau)$ , the expression of both criteria for a given maturity  $\tau$  is ( $T$  is the last date in our dataset):

$$MAD(\tau) = \frac{1}{60} \sum_{t=T-59}^T |\hat{y}(t, \tau) - y(t, \tau)| \quad (5.1)$$

$$RMSE(\tau) = \sqrt{\frac{1}{60} \sum_{t=T-59}^T (\hat{y}(t, \tau) - y(t, \tau))^2} \quad (5.2)$$

Looking at equations (5.1) and (5.2), we can understand that the smaller the criteria, the better the forecast performance of the model.

Duffee [12] explains that factor term structure models do not easily perform better than random walks in yields forecasting. The random walk performance will be used as a benchmark.

As we can see in table (TAB. 7), the no-arbitrage N&S model performs better than the random walk and than the standard N&S.

Forecast Performances ( $\times 10^{-3}$ )						
Maturity	MAD criteria			RMSE criteria		
	RW	NA-N&S	NS	RW	NA-N&S	NS
3	1.3532	<b>1.2091</b>	1.2102	1.8605	<b>1.5854</b>	1.6054
6	1.4265	<b>1.3016</b>	1.3149	1.9353	<b>1.6595</b>	1.6841
12	1.9860	<b>1.8664</b>	1.8815	2.5391	<b>2.3506</b>	2.3710
24	2.2433	<b>2.1459</b>	2.1619	2.8232	<b>2.6956</b>	2.7140
60	2.3105	<b>2.2646</b>	2.2703	2.8744	<b>2.8240</b>	2.8345
120	2.0973	<b>2.0772</b>	2.0812	2.6290	<b>2.6129</b>	2.6165

Table 7: Forecast performances of both models

Three reasons may explain this success in the out-of-sample forecast. First, as pointed out by Duffee, a time-varying risk premium, similar to ours, is of great help in forecasting performance. The explicit no-arbitrage constraints are useful in improving the forecast performance of the model. Second, our model offers a very interesting channel to understand how it helps to beat the random walk forecast. The random walk is in fact a polar case of our no-arbitrage N&S 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 a second factor would have been useless in improving the forecast performance.

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 relative to more parsimonious ones. Our no-arbitrage N&S model can indeed be also seen as a constrained factor term structure model.

It seems that all constraints we have incorporated in our model: no-arbitrage ones and Nelson & Siegel ones, help to improve the out-of-sample forecast performance.

## 6 Conclusion

This article provides a simple model which is close to standard N&S representation and which allows no-arbitrage cross-sectional restrictions. The main difference with an unconstrained factor term structure model is the factors risk-neutral dynamics. The mean reverting speeds must indeed verify several equalities to allow a N&S representation. The main difference with a standard N&S model are the Jensen terms (time series constant terms), whose expression is constrained with no-arbitrage.

In the appendix<sup>2</sup>, we derive more general conditions under which a factor term structure model admit a N&S representation. Without going deep in technical details, these conditions depend first on the factors volatility structure<sup>3</sup> and on the factors mean reverting speeds.

We have empirically estimated the no-arbitrage model and the standard N&S one using Kalman filter equations and maximum likelihood. It is before all important to notice that both models exactly have the same number of parameters and that the estimation technique is analogous. Both representations are therefore as parsimonious as each other. The comparison between both reveals that the no-arbitrage version performs better than the standard one, except for yield curve fitting, where results are comparable. The main successes of the no-arbitrage model are related to its proper definition for the risk premium, which means the rejection of the Expectation Hypothesis and performing out-of-sample forecast.

It would have been interesting to make this comparisons in a more general framework: general two factors model (with crossed factors) or even three factors model. One has however to keep in mind that analytical simplicity would strongly decrease.

Redo estimations for a general 2 factors model ? 3 factors model ?

---

<sup>2</sup>Not yet available.

<sup>3</sup>More precisely, our results depend strongly on the parameter used by Dai and Singleton in their classification, the rank of the 'stochastic' part of volatility.

## References

- [1] Andrew Ang and Monika Piazzesi. “A No–Arbitrage Vector Autoregression Of Term Structure Dynamics With Macroeconomic And Latent Variables”. *Journal of Monetary Politics*, 50: 745–787, 2003.
- [2] Ernst K. Berndt, Bronwyn H. Hall, Robert E. Hall, and Jerry A. Hausman. “Estimation and inference in nonlinear structural models”. *Annals of Economic and Social Measurement*, 3: 653–655, 1974.
- [3] Tomas Björk and Bent Jesper Christensen. “Interest Rate Dynamics and Consistent Forward Rate Curves”. *Mathematical Finance*, 9(4): 323–348, 1999.
- [4] John Y. Campbell and Robert J. Shiller. “Yield Spreads and Interest Rate Movements: A Bird’s Eye View”. *Review of Economic Studies*, 58: 495–514, 1991.
- [5] Ren-Raw Chen and Louis O. Scott. “Maximum Likelihood Estimation for a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates”. *Journal of Fixed Income*, 3: 14–31, December 1993.
- [6] Qiang Dai and Thomas Philippon. “Fiscal Policy and the Term Structure of Interest Rates”. *New York University Working Paper*, December 2004.
- [7] Qiang Dai and Kenneth J. Singleton. “Specification Analysis of Affine Term Structure Models”. *Journal of Finance*, 55(5):1943–1978, Oct. 2000.
- [8] Qiang Dai and Kenneth J. Singleton. “Expectation puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure”. *Journal of Financial Economics*, 63:415–441, 2002.
- [9] Francis X. Diebold and Calin Li. “Forecasting the Term Structure of Government Bond Yields”. *Working Paper, University of Pennsylvania*, Forthcoming, 2003.
- [10] Francis X. Diebold, Monika Piazzesi, and Glenn D. Rudebusch. “Modeling Bond Yields in Finance and Macroeconomics”. *American Economic Review (Papers and Proceedings)*, Session ID 71, 2005.
- [11] Francis X. Diebold, Glenn D. Rudebusch, and Boragan S. Aruoba. “The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach”. *Journal of Econometrics*, Forthcoming, 2005.
- [12] Gregory R. Duffee. “A Macro–Model Of The Term Structure, Monetary Policy, and the Economy”. *Journal of Finance*, 57: 405–443, 2002.
- [13] Darrell Duffie and Rui Kan. “A Yield Factor Model of Interest Rates”. *Mathematical Finance*, 6: 379–406, 1996.
- [14] K Levenberg. “A Method for the Solution of Certain Problems in Least Squares”. *Quarterly of Applied Mathematics*, 2(2): 164–168, July 1944.
- [15] D Marquardt. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”. *SIAM Journal of Applied Mathematics*, 11(2): 431–441, June 1963.

- [16] BIS Monetary and Economic Department. “Zero-coupon yield curves: technical documentation”. *BIS Papers No. 25*, October 2005.
- [17] Charles R. Nelson and Andrew F. Siegel. “Parsimonious Modeling of Yield Curves”. *The Journal of Business*, 60(4): 473–489, October 1987.
- [18] Oldrich Vasicek. “An equilibrium characterization of the term structure”. *Journal of Financial Economics*, 5(2): 177–188, November 1977.

# Appendix

## A Variance decomposition

In this section we derive explicit expressions for the variances decompositions. Suppose that we have the following state system (with obvious 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  $\Lambda$  lie in the unit circle<sup>4</sup>. We 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 effective  $Y_{t+h}$  is therefore:

$$Y_{t+h} - \hat{Y}_{t+h|t} = \sum_{k=0}^{h-1} B\Lambda^k \eta_{t+h-k} + \varepsilon_{t+h} \quad (\text{A.1})$$

Our both models shared some common features.  $\Lambda$  is  $2 \times 2$  diagonal matrix,  $B$  is identical,  $\eta_{t+h-k}$  and  $\varepsilon_{t+h}$  are independent and have diagonal covariance matrices. With obvious notations, the expression of the mean square error  $MSE_j$  for the component  $j$  is the following:

$$\begin{aligned} MSE^j(h) &= \mathbb{E} \left[ (Y_{t+h}^j - \hat{Y}_{t+h|t}^j)^2 \right] = \left[ \sum_{k=0}^{h-1} l_1^{2k} \right] \sigma_1^2 \\ &\quad + \left[ \sum_{k=0}^{h-1} l_2^{2k} \right] \sigma_2^2 \left[ \frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} \right]^2 \\ &\quad + s_j^2 \end{aligned}$$

We can then compute the variance share of yield  $j$  at horizon  $h$  explained by the first factor  $\omega_1^j(h)$ , by the second factor  $\omega_2^j(h)$  and the residual  $\omega_r^j(h)$ :

$$\begin{aligned} \omega_1^j(h) &= \frac{1 - l_1^{2h}}{1 - l_1^2} \frac{\sigma_1^2}{MSE^j(h)} \\ \omega_2^j(h) &= \frac{1 - l_2^{2h}}{1 - l_2^2} \frac{\sigma_2^2}{MSE^j(h)} \left[ \frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} \right]^2 \\ \omega_r^j(h) &= \frac{s_j^2}{MSE^j(h)} \end{aligned}$$

With  $h \rightarrow \infty$ , we get the unconditional variance decomposition.

---

<sup>4</sup>This assumption is true for our 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:

$$Y_t = A + B X_t + \varepsilon_t \quad (\text{B.1})$$

$$X_t = \Lambda X_t + \eta_t \quad (\text{B.2})$$

We suppose that the perturbations  $\varepsilon_t$  and  $\eta_t$  are independent of each other and of the initial state  $X_0$ . Their respective variance matrices are  $\Sigma_\varepsilon$  and  $\Sigma_\eta$ .

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

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

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

Regarding the initial conditions, they do not have any problem for  $X_{1|0}$ , which is the best forecast at date 0 of  $X_1$ . This is simply the unconditional mean of  $X$ , which is stationary under the historical probability  $\mathbb{P}$ .

$$X_{1|0} = \mathbb{E}^\mathbb{P} [X] \quad (\text{B.3})$$

Because the process  $X$  are stationary<sup>5</sup>, the initial mean squared error is defined as follows:

$$\text{vec}(P_{1|0}) = (I_{r^2} - \Lambda \otimes \Lambda)^{-1} \text{vec}(\Sigma_\eta) \quad (\text{B.4})$$

In the preceding equation,  $r$  is the number of components of  $X$  ( $= 2$ ),  $\otimes$  is the Kronecker product, and  $\text{vec}(M)$  is the vector representation of the matrix  $M$ <sup>6</sup>.

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

$$Y_t | \mathcal{I}_t \rightsquigarrow \mathcal{N} \left( A + B \hat{X}_{t+1|t}, B P_{t+1|t} B^\top + \Sigma_\varepsilon \right) \quad (\text{B.5})$$

We can compute easily the log likelihood  $\mathcal{L}$  corresponding to the observation  $\tilde{Y}_{t_k}$ .  $\Theta$  is the set of parameters we have to estimate.

$$\begin{aligned} \mathcal{L}_{\tilde{Y}_{t_k}}(\Theta) = & -\frac{1}{2} [\ln(2\pi) + \ln(\det(B P_{t+1|t} B^\top + \Sigma_\varepsilon))] \\ & + (\tilde{Y}_{t_k} - (A + B \hat{X}_{t+1|t}))^\top (B P_{t+1|t} B^\top + \Sigma_\varepsilon)^{-1} \times (\tilde{Y}_{t_k} - (A + B \hat{X}_{t+1|t})) \end{aligned} \quad (\text{B.6})$$

We can therefore compute the likelihood recursively and then maximize it relative to  $\Theta$ .

<sup>5</sup>At least under the historical probability and that what is here of interest.

<sup>6</sup>If  $M = (m_{ij})_{i=1\dots n, j=1\dots p}$  then  $\text{vec}(M) = [m_{11} \ m_{21} \ \dots \ m_{n1} \ m_{12} \ \dots \ m_{np}]^\top$