Measuring Term Premium. Robustness across Alternative Dynamic Term Structure Models*

(Very preliminary and incomplete)

Carlo Favero Iryna Kaminska

June 16, 2005

Abstract

This paper uses different term structure models for term premia estimation. In particular, we consider discrete Term Structure Models (TSM), which specify the driving stochostic process for the yield curve by Gaussian VAR. First, we refer to No-Arbitrage Affine TSM (see Duffie, Kan (96); Dai, Singleton (00)). Second, we estimate term premium by two-step procedure: we fit the yield curve by Nelson-Siegel model and then we add the assumption about the dynamics of the factors. Finally, the results are compared to the estimates implied by simple unrestricted VAR. We find that differences in term premia estimates among alternative specifications of Term Structure VAR Models are small, and thus the less computationally demanding method could be used to obtain the term premia.

1 Introduction

Over the last decade, different term structure models have been used for the analysis of the Term Premium, i.e. the difference between the yield to maturity of long term bond and the average of expected future short-term bond yields,

$$TP_{n,t} = y_t^n - \frac{1}{n}\sum_{j=0}^{n-1} E_t[i_{t+j}]. \tag{1}$$

The traditional expectations hypothesis, which states that the term premium is constant, is widely rejected by the empirical research: the term premium is time varying and, moreover, it appears to be important variable in finance and macroeconomic literature. Nevertheless, a suitable theory for the term premium is still required.

*_Favero_: IGIER, Universit'a Bocconi and CEPR, Via Salasco 5, 20136 Milano, Italy, carlo.favero@uni-bocconi.it; _Kaminska_: Universit'a Bocconi, Via Salasco 5, 20136 Milano, Italy, iryna.kaminska@uni-bocconi.it.

Despite the simplicity of the concept, there are severe challenges for the term premium estimation. First, the market expectations are not observable and there is neither a commonly accepted theoretical model nor an agreed method to proxy these expectations. Second, in order to estimate term premium for any maturity we have to work with the whole yield curve, which is also unobservable. Thus only when the term structure model could provide both, the framework for the shape of the yield curve and the proxy for the expectations, it could be considered as a flexible tool for the term premium analysis.

While it is not a problem to find a good yield curve fitting model, modeling the market expectations is a difficult task. The common approach is to use ex-post observed returns as a valid proxy for ex-ante expected returns. The approach has been questioned by Elton (1999), who provided ample evidence against the belief that information surprises tend to cancel out over time. Hence, realized returns cannot be considered as an appropriate proxy for expected returns. Campbell and Shiller (87) circumvent this problem and propose the VAR framework as an explicit model for the expected future rates. Given the path of VAR-projected future short rates, it is possible to construct yields to maturity consistent with the expectations theory and, as a residual, the term premium.

Even if it is reasonable to combine the yield curve fitting model with Campbell-Shiller VAR approach, VAR model is still too general framework to be a final solution to the problem. Two further questions remain open: which variables to include into the VAR?, and which restrictions to impose on the VAR coefficients?

Campbell and Shiller (87) use bivariate VAR model, according to which the only determinant of policy rates are long-term rates. However, the success of Taylor rules (Taylor,1993) points out an obvious potential mis-specification of the yields-only framework: the omission of macroeconomic variables to which the monetary policy maker reacts. Thus, focusing on the estimation of the expected future short term rates, it is natural to enrich the VAR with variables related to inflation and output. Clearly, including observable yields and macroeconomic variables into the unrestricted VAR will provide very large number of estimated coefficients, and hence it will be very imprecise model to work with.

An alternative solution is provided by Factor models, which assume certain relationship between yields of different maturities and therefore impose restrictions on the VAR coefficients.

Among the most popular dynamic term structure factor models are Affine Term Structure Models (ATSM) (see Duffie, Kan (96); Dai, Singleton (00)), which, in discrete time, imply VAR of yields with complex cross-equation restrictions due to the no-arbitrage assumption. Despite the high dimensionality and extreme non-linearity, many authors (see e.g. Ang et al (2004), Hördahl et al (2004)) use this type of models to estimate term premium (1).

In contrast, there are less computationally demanding ways to measure term premia by factor models. For example, it is possible to estimate term premium for any maturity using Nelson-Siegel (1987) parametric form for the yield curve and specifying additionally the VAR dynamics for the factors (as in Diebold and Li (2005), Carriero, Favero, Kaminska (2005)).

A natural question is: What is the impact of the alternative restrictions on the VAR-s for modeling the term premium? In this paper, we seek to answer this question by studying different discrete TSM, which specify the driving stochastic process for the yield curve by Gaussian VAR. First, we consider unrestricted VAR models. Second, to provide estimates of term premia, we estimate two types of ATSM: following the recent tendency, together with standard ATSM, we consider also joint macro-finance ATSM (as Ang, Bekaert (03), Rudebusch, Wu (04) etc.). Finally, we explore VARs implied by Nelson-Siegel dynamic factor model.

We find that differences in term premia estimates among alternative specifications of discrete Term Structure Models are small.

The paper is organized as follows. Section 2 defines the term premium from the perspective of the expectations hypothesis. Section 3 discusses VAR, dynamic Nelson-Siegel and No-Arbitrage Affine TSM approaches. Section 4 summarizes estimation details, while Section 5 compares the term premium estimates from different approaches. The last section concludes.

2 Expectations Hypothesis and Term Premium

The Expectations Hypothesis can be represented in several forms (Cox, Ingersoll, Ross (1985)). We work here with the Yield to Maturity Expectations Hypothesis in its logarithmic form[1].

Let y_t^n , i_t denote n-period yield and one-period interest rate respectively. Then logarithmic form of the Expectation Hypothesis states that

$$y_t^n = \frac{1}{n}\sum_{j=0}^{n-1} E_t[i_{t+j}] + TP_n, \tag{2}$$

where $E_t[i_{t+j}]$ denotes the market's expectations at time t of the one-period interest rate at time $t+j$. The term premium TP_n could be viewed as a sum of the risk and liquidity premium.

The traditional form of Expectations Theory (ET) assumes that a term premium is constant (zero in the case of Pure ET). Nevertheless the assumption of constant premium is merely a technical simplification of the theory. We follow Longstaff (1990) and Hamilton and Kim (2002) and from here and below consider a variable term premium[2]. Rearranging terms, we find an expression for the term premium:

$$TP_{n,t} = y_t^n - \frac{1}{n}\sum_{j=0}^{n-1} E[i_{t+j} \mid I_t] \tag{3}$$

[1] The log form is only the approximation of the EH and is not appropriate for the periods when the rates of returns take high values (like 1980-1983).

[2] With time varying term premium, the EH still holds if the term premium is restricted to be orthogonal to the spread.

The estimation of the term premium is difficult in practice as it involves expectations about the future path of the short-term interest rate, and alternative decompositions may differ substantially depending on how expectations are modelled.

3 VAR-based Models for Expectations

3.1 Unrestricted VAR Models

Our VAR-based approach is closely related to the paper by Campbell and Shiller (1987). By having an explicit model for the short-rate in VAR framework, they circumvent the use of ex-post realized returns as a proxy for ex-ante expected returns. By implementing the simulation based procedure one can explicitly measure deviations from the ET and, under the null that the proposed model delivers expected future policy rates not different from those expected by the market, interpret them as a measure of risk premium.

The bivariate CS approach has an implicit reaction function according to which the only determinant of policy rates are long-term rates. The success of Taylor rules (Taylor,1993) points out an obvious potential mis-specification of the yields-only framework: the omission of macroeconomic variables to which the monetary policy maker reacts. In general, standard approaches include in the VAR models the interest rates and inflation in levels, alternative specifications include the measures of the real activity as well (i.e. Kozicky, Tinsley (2001)). Ang, Piazzesi, Wei (2005) also derive expectations for future policy rates considering a vector of state variables that follows a Gaussian Vector Autoregression with one lag:

$$Y_t = \mu^U + \Phi^U Y_{t-1} + \Sigma \epsilon_t \tag{4}$$

In their case, the vector Y_t contains two factors from the yield curve, the 3-month rate, i_t^1, expressed at a quarterly frequency, to proxy for the level of the yield curve, and the 5-year term spread, $i_t^{20} - i_t^1$, to proxy for the slope of the yield curve, the last factor is the quarterly real GDP growth, $\Delta_4 y_t$. Expected risk-free rate are derived by simulating the VAR from (4) forward.

In this paper, working with unrestricted VAR, we consider two alternative measures of TP_t. The first one is obtained by applying unrestricted VAR(1) model to the vector Y_t, which contains the short yield, y_t^3, expressed at a monthly frequency, and the 5-year yield, y_t^{60} : $Y_t' = \left[y_t^3, y_t^{60}\right]$. The success of Taylor rules (Taylor,1993) points out an obvious potential mis-specification of the yields-only framework: the omission of macroeconomic variables to which the monetary policy maker reacts. We shall assess potential mis-specification effects by using an extended VAR, so that in the second case the vector Y_t contains the one quarter yield, y_t^3, the 5-year yield, y_t^{60}, and inflation π_t : $Y_t' = \left[y_t^3, y_t^{60}, \pi_t\right]$.

In each case, we simulate the estimated model (4) forward, to obtain projection for all relevant policy rates and to construct $\widehat{ET}_t$, which is the Expectations Theory consistent spreads, as follows:

$$\hat{ET}_t = \frac{1}{20}\sum_{j=0}^{19} E[y^3_{t+j} \mid \Omega_t] \tag{5}$$

where, $E[y^3_{t+j} \mid \Omega_t]$ are the VAR-based projections for the future changes in policy rates, hence Ω_t is the information set used by the econometrician to predict on the basis of the estimated unrestricted VAR model . The unrestricted VAR-based measure of term premium is then

$$TP^U_{3,60,t} = y^{60}_t - \frac{1}{20}\sum_{j=0}^{19} E[y^3_{t+j} \mid \Omega_t] \tag{6}$$

The information set Ω_t does not include any theoretical restriction. Alternatively, when working with term structure models, it would be natural to include information on the absence of arbitrage opportunities into the information set of the econometrician. We address how to impose the No-Arbitrage restrictions into the VAR model in the next section.

3.2 No-Arbitrage Term Structure Models

According to financial No-Arbitrage TSM, there are only few factors, X_t, relevant for pricing risk in the bond sector. If agents are risk neutral, then No-Arbitrage implies that any bond is priced by the following rule:

$$P_t = E_t(e^{-r_t} P_{t+1}), \tag{7}$$

with bond's price $P_t = P(X_t)$, $X_{t+1}|X_t \sim N(\mu_t, \Sigma_t\Sigma_t')$, $\mu_t = \mu(X_t)$, $\Sigma_t = \Sigma(X_t)$. If agents are not risk-neutral, then their behavior can be represented as that of risk-neutral with "distorted beliefs" about the distribution of $X_t :\ X_{t+1}|X_t \sim N(\mu^Q_t, \Sigma_t\Sigma_t')$, where $\mu^Q_t = \mu_t - \Sigma_t\Lambda_t$. Assuming that market uses these distorted beliefs to evaluate $P_t = E^Q_t(e^{-r_t} P_{t+1})$, the density to find these expectations would be

$$\begin{aligned}
f^Q_t(X_{t+1}) &= (2\pi)^{-\frac{N}{2}} \left|\Sigma_t\right|^{-1} \exp(-\frac{1}{2}(X_{t+1} - \mu^Q_{t+1})'(\Sigma_t\Sigma_t')^{-1}(X_{t+1} - \mu^Q_{t+1})) = \\
&= (2\pi)^{-\frac{N}{2}} \left|\Sigma_t\right|^{-1} \exp(\frac{(X_{t+1} - \mu_t + \Sigma_t\Lambda_t)'(\Sigma_t\Sigma_t')^{-1}(X_{t+1} - \mu_t + \Sigma_t\Lambda_t)}{-2}) \\
&= f_t(X_{t+1}) \exp(-\frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\Sigma_t^{-1}(X_{t+1} - \mu_t)) \\
&\equiv f_t(X_{t+1}) \exp(-\frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\Sigma_t^{-1}\varepsilon_{t+1})
\end{aligned}$$

The pricing kernel, M_{t+1}, is given by

$$\begin{aligned}
M_{t+1} &\equiv e^{-r_t} e^{-\frac{1}{2}\Lambda_t'\Lambda_t - \Lambda_t'\Sigma_t^{-1}\varepsilon_{t+1}} && (8) \\
P_t &= E^Q_t(e^{-r_t} P_{t+1}) = E_t(M_{t+1}P_{t+1}) && (9)
\end{aligned}$$

Estimating market prices of risk econometrician faces numerous challenges. Importantly, the presence of unobservable (latent) variables and the absence of the closed-form solution of the system of stochastic difference equations for bond prices do not allow one to use the maximum likelihood estimation. Instead, a closed-form solution for bond prices could be obtained imposing the affine structure into the model. Therefore, the majority of empirical studies adopt a specification which is affine.

We follow this direction and consider only Affine Term Structure Models. Canonical ATSM contains 3 basic equations:

1) Transition equation for the state vector relevant for pricing bonds (Gaussian VAR):

$$X_t = \mu + \Phi X_{t-1} + \Sigma\varepsilon_t \tag{10}$$

2) Definition of one period rate as a linear function of the state variables,

$$r_t = \delta_0 + \delta_1 X_t. \tag{11}$$

3) The price of risk, Λ_t, is associated with shocks ε_t and it is identified as an affine function of the state of economy (see Duffee (02)).

$$\Lambda_t = \lambda_0 + \lambda_1 X_t \tag{12}$$

Under these assumptions, as we show in the Appendix, the interest rate of any maturity is affine function of the state variables: $r_{t,n} = A_n + B_n' X_t,$ where A_n, B_n are the functions of the parameters $\{\lambda_0, \lambda_1, \delta_0, \delta_1, \mu, \Phi, \Sigma\}$:

$$\left\{ \begin{array}{c} A_{n+1} = A_n + B_n'\mu - B_n'\Sigma\lambda_o + \frac{1}{2}B_n'\Sigma\Sigma' B_n \\ B_{n+1}' = B_n'\Phi - \delta_1' - B_n'\Sigma\lambda_1 \end{array} \right\} \tag{13}$$

The yield on a zero-coupon bond of maturity n, is affine structure of the state:

$$y_t(n) = -\frac{1}{n}(A_n + B\prime_n X_t) \equiv a_n + b_n' X_t \tag{14}$$

3.2.1 No-Arbitrage VAR

In this subsection we are going to show how the No-Arbitrage assumption results in a set of restrictions on VAR from (4).

To estimate parameters and extract the factors in ATSM we use the approach due to Chen and Scott (1993). In this setting, N unobservable factors are measured by assuming that N bonds, $\hat{R}_t$, in the cross section are priced without error:

$$\hat{R}_t = \mathbf{A1} + \mathbf{B1} X_t. \tag{15}$$

Other interest rates are assumed to be priced with errors. Once all parameters of the model are estimated, the factors, X_t, could be extracted by inverting the pricing relationship of the model:

$$X_t = \mathbf{B1}^{-1}[\hat{R}_t - \mathbf{A1}] \tag{16}$$

Given the dynamics of the latent factors (10), the dynamics of the observed bond yields can be retrieved by combining equations, (15), (16) in the following way:

$$\begin{aligned}\hat{R}_t &= \mathbf{A1} + \mathbf{B1}\left(\mu + \Phi X_{t-1} + \Sigma\varepsilon_t\right) \\ &= \left(\mathbf{A1} + \mathbf{B1}\mu - \mathbf{B1}\Phi\mathbf{B1}^{-1}\mathbf{A1}\right) + \mathbf{B1}\Phi\mathbf{B1}^{-1}\hat{R}_{t-1} + \mathbf{B1}\Sigma\varepsilon_t \qquad (17)\end{aligned}$$

Note that the No-Arbitrage assumption implies the VAR for the observable variables with complex cross equation restrictions. We denote this VAR as

$$Y_{t-1} = \mu^{NA} + \Phi^{NA} Y_{t-1} + \Sigma^{NA}\epsilon_t, \qquad (18)$$

where Y_t stands for the observed yields (in our case $Y_t = [y_t^3, y_t^{60}]$),

$$\begin{aligned}\mu^{NA} &\equiv \mathbf{A1} + \mathbf{B1}\mu - \mathbf{B1}\Phi\mathbf{B1}^{-1}\mathbf{A1} \qquad (19) \\ \Phi^{NA} &\equiv \mathbf{B1}\Phi\mathbf{B1}^{-1} \qquad (20)\end{aligned}$$

The No-Arbitrage based measure of term premium is then

$$TP_{3,60,t}^{NA} = y_t^{60} - \frac{1}{20}\sum_{j=0}^{19} E[y_{t+j}^3 \mid \Omega_t^{NA}], \qquad (21)$$

where information set of the econometrician, Ω_t^{NA}, includes the theoretical assumption of No-Arbitrage.

The huge popularity of the No-Arbitrage ATSM in finance is due to the fact that implied affine functions of few unobservable (latent) factors could explain almost all movements of the yield curve (see Duffie, Kan (96); Dai, Singleton (00))[3]. Nevertheless, the pure ATSM has not gained the same popularity among economists since the model is not useful for Macroeconomic Policy. There is no any theory behind the NA-ATSM apart from the No-Arbitrage assumption and the economic nature of the latent factors is unknown. Observing that short term rate is a Policy Rate, macroeconomists have proposed a possible solution: to combine ATSM No-Arbitrage models with Macroeconomic models.

3.2.2 ATSM joint with Macroeconomic Models

Ways of incorporating the macroeconomic theory into the No-Arbitrage models could be divided into three groups. First, one can extract the unobservable factors from the pure yields model and then to look for their macroeconomic interpretation using Taylor rules, or other standard macroeconomic relationships for output and inflation (see Rudebusch, Wu (04)). Second, some authors claim that all factors relevant for the bond pricing are observable. The example of this approach is the paper of Ang, Piazzesi, Wei (04), who add to VAR-based macro-model only the No-Arbitrage assumption. The paper by

[3] just 2 factors could explain more than 99%

Ang and Bekaert (04) belongs to the third group, which is a mixture of the first two approaches. The authors assume that the state vector relevant for the bond pricing consists from both, latent and observable, factors. The approach is very popular and there is a bunch of papers devoted to it: Hordhal, Tristani, Vestin (04); Ang, Piazzesi (03), etc. Table 1 summarizes the recent work on the affine term structure models enriched by macroeconomic information.

Table 1. Different types of joint macro-finance models

Paper	State vector	F	Dynamics	Method	Yields exact	with error	Sample
Ang, Bekaert (04)	2 latent+π	Q	RS-VAR(1)	MLE	1q, 5y	1y, 3y	1952-2000
Ang, Piazzesi (03)	3 latent +"π"+"y"	M	VAR(12)	2-step LS	1m, 1y, 5y	1q, 3y	1952-2000
Ang et al (04)	r1+R10+growth	Q	VAR(1)	2-step LS	1q, 5y	-	1964-2001
Dai et sl (03)	3 latent	M	RS-VAR(1)	MLE	2q, 2y, 10y	5y	1970-1995
Hordhal et al (04)	1 latent +r+π+y	M	VAR(3)	MLE	1m, 3y	1q, 2q, 1y, 7y	1975-1998
Rudebusch, Wu (04)	2 latent	M	VAR(1)	MLE	1m, 5y	1q,1y, 3y	1988-2000

The empirical studies reported in Table 1 employ quite a variety of model specifications and data. Consequently, the results are difficult to compare directly. Nevertheless, there is a number of results which are robust. For instance, in the model of Rudebusch and Wu (2004), the level factor reflects market participants' views about the inflation target of the central bank. Diebold, Rudebusch, and Aruoba (2005) also find that the level factor is highly correlated with inflation, while the slope factor in their model is highly correlated with real activity. Inflation turns to be a priced risk factor (Buraschi, Jiltsov (04), Ang, Piazzesi(03) etc.).

To apply no-arbitrage macro-finance framework to the analysis of the term premium, we concentrate on the approach of Ang and Bekaert (04) and include inflation in ATSM as an observable factor. In this case the measure of the term premium is derived from the simulation of VAR given by equation (18), with $Y_t = [y_t^3, y_t^{60}, \pi_t]$

3.3 Nelson-Siegel Approach

In this section we use an alternative method to extract latent factors driving the yield curve. We estimate the yield curve at each point in time by the help of the simple term structure model proposed by Nelson and Siegel (1987). At each point of time, we construct financial factors by estimating (by non-linear least squares, on the cross-section of observed yields) the following Nelson-Siegel model :

$$y_t^k = L_t + SL_t \frac{1 - \exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}} \tag{22}$$

The parameter τ_1 is kept constant over time[4], as this restriction decreases the volatility of the parameters $X_t^{NS} = (L_t, SL_t)'$, making them more predictable in

[4] We restrict τ_1 at the value of 1.8.

time. As discussed in Diebold and Li (2002) the above interpolant is very flexible and capable of accommodating several stylized facts on the term structure and its dynamics. In particular, L_t, SL_t, which are estimated as parameters in a cross-section of yields, can be interpreted as latent factors. L_t has a loading that does not decay to zero in the limit, while the loadings on all the other parameters do so, therefore this parameter can be interpreted as the long-term factor, the level of the term-structure. The loading on SL_t is a function that starts at 1 and decays monotonically towards zero; it may be viewed a short-term factor, the slope of the term structure. In fact, $r_t^{rf} = L_t + SL_t$ is the limit when k goes to zero of the spot and the forward interpolant. We naturally interpret r_t^{rf} as the risk-free rate. Obviously, SL_t is the slope of the yield curve. The repeated estimation of loadings using a cross-section of yields at different maturities allows to construct a time-series for our factors.

Interestingly, Nelson-Siegel model of the term structure is consistent with the implications of the No-Arbitrage ATSM presented in previous subsection. As before, the yields are affine in state factors:

$$y_t^k = a^{NS,k} + b^{NS,k} X_t^{NS}, \tag{NS}$$

where the following restriction holds:

$$a^{NS,k} = 0 \tag{23}$$

$$b^{NS,k} = \left(1, \frac{1-\exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}}\right) \tag{24}$$

The often-quoted shortcoming of the Nelson-Siegel model is its static nature: the factors are extracted from the current yield curve and the information about the shapes of the past yield curves is omitted. To overcome this drawback, it is naturally to add the assumptions on the dynamics of the Nelson-Siegel factors. Following Diebold and Li (05) and Carriero-Favero-Kamisnka (05), we assume that the factors follow the Gaussian VAR process:

$$X_t^{NS} = \mu^{NS} + \Phi^{NS} X_{t-1}^{NS} + \Sigma^{NS} \epsilon_t. \tag{25}$$

Finally, given the factor dynamics (25) and linear relationship between yields and factors (**??**), it can be easily shown that the corresponding dynamics of the Nelson-Siegel yields is also described by VAR:

$$Y_{K,J,t}^{NS} = \mu_{K,J}^{NS} + \Phi_{K,J}^{NS} Y_{K,J,t-1}^{NS} + \Sigma_{K,J}^{NS} \epsilon_t, \tag{26}$$

$$\mu_{K,J}^{NS} \equiv B_{K,J}^{NS} \mu^{NS} \tag{27}$$

$$\Phi_{K,J}^{NS} \equiv B_{K,J}^{NS} \Phi^{NS} \left(B_{K,J}^{NS}\right)^{-1} \tag{28}$$

$$\Sigma_{K,J}^{NS} \equiv B_{K,J}^{NS} \Sigma^{NS} \tag{29}$$

$$B_{K,J}^{NS} \equiv \begin{bmatrix} 1 & \tau_1 \frac{1-\exp\left(-\frac{k}{\tau_1}\right)}{k} \\ 1 & \tau_1 \frac{1-\exp\left(-\frac{j}{\tau_1}\right)}{j} \end{bmatrix} \tag{30}$$

where $Y^{NS}_{K,J,t}$ denotes the vector of K- and J- maturity yields implied by Nelson-Siegel model (22). In fact, the model implies that model-implied yields follow VAR process, that is, Nelson-Siegel parametric restrictions are imposed on the VAR coefficients for yields of different maturities.

The Nelson-Siegel VAR-based measure of term premium $TP^{NS}_{60,t}$ is then

$$TP^{NS}_{3,60,t} = y^{NS,60}_t - \frac{1}{20}\sum_{j=0}^{19} E[y^{NS,3}_{t+j} \mid \Omega^{NS}_t]. \tag{31}$$

Finally, it is worth noting that the structure of the Nelson-Siegel model does not necessarily rule out the No-Arbitrage assumption. In fact, the basic relationships of the ATSM model (10), (11) and (14), have the same form as the relationships (25), (**??**) for the Nelson-Siegel model. Therefore, the Nelson-Siegel factors satisfy the assumption of the No-Arbitrage if the parameters $a^{NS,k}, b^{NS,k}$ are consistent with system of difference equation (13). For the details, see Diebold, Piazzesi and Rudebusch (05), who examine under which restrictions the No-Arbitrage restriction can be applied to the Nelson-Siegel term structure model.

4 Estimation

We consider monthly data on bond yields. In particular, we focus on the US zero-coupon bonds, assuming them to be default-risk-free. The data is available on the G. Duffee's home page.[5] and consists of time series of six yields,. $[y^3, y^6, y^{12}, y^{24}, y^{60}, y^{120}]$.

First, we estimate two-factor models: standard ATSM ($ATSM(2,0)$), dynamic Nelson-Siegel model[6], and unrestricted VAR of observed yields.

Second, to apply our framework to the analysis of the US term structure we consider a standard specification of the macroeconomic structure by including the annual US CPI inflation at time t, π_t, into the model. Again, we estimate a one-lag VAR model for three cases: ATSM with 2 latent factors and inflation as an observable factor($ATSM(2,1)$); VAR of Nelson-Siegel-implied yields and inflation; and unrestricted VAR of observed yields and inflation.

We limit the sample to 1988:1-1997:12 for all models under consideration. The choice of the sample was influenced by the question of the stability of the estimates. While the relationship between yields might remain stable over time, the relationship between interest rates and macroeconomic variables has changed over time. Thus, including observable macrofactors into ATSM, we have to limit samples to short intervals of plausible stability in MP regime (as Rudebusch, Wu (04)).[7]

[5] Duffee uses mixed data sourses. The data through February 1991 are from McCulloch's home page, After February 1991, the data are from Rob Bliss.

[6] The Nelson-Siegel term structure approximation is based on five yields from the Duffe's data set, $[y^3, y^6, y^{12}, y^{24}, y^{60}]$.

[7] The ATSM with Regime Switching (RS) could be an alternative solution. Dai, Singleton, Yang (03) and Ang and Bekaert (04) develop and empirically implement an arbitrage-free, dynamic term structure model (DTSM) with regime-shifts. However, the method is com-

For every VAR model under consideration, given the results of the estimation, the companion matrix is retrieved. For each point of our sample, the VARs are then projected for an horizon up to twenty observations to generate observation of the ET-consistent five year yield. The ET-consistent yields are then to be compared with observed yields, and the difference is interpreted as a term premium. The procedure is repeated for a total of 120 simulations of each model.

4.1 Estimation of ATSM

The most computationally demanding model is ATSM. We detail how to compute the likelihood function in Appendix.

The Chenn-Scott estimation approach requires to assume that, in two latent factors models, there are exactly two reference yields specified without errors. The likelihood function is the likelihood of the yields measured without error multiplied by the likelihood of the measurement errors. The well-known problem of the Chen-Scott approach is that different choices of the reference bonds imply different state variable realizations. In order to choose the reference yields, we estimate the model for all possible combinations of the pairs of reference yields and compare their fit and stability (i.e. the model should produce good fit for any out-of sample long-term yield, which is y^{120} for our case). The best performance of the two-factor $ATSM$ model is achieved when the reference yields are chosen to be $[y^3, y^{60}]$.

In order to limit the number of parameters $\{\lambda_0, \lambda_1, \delta_0, \delta_1, \mu, \Phi, \Sigma\}$ to be estimated, we de-mean the values of the variables for all models (the procedure should not distort the results, since we limit the sample to the stable interval). [8] De-meaning allows us to set $\lambda_0 = \delta_0 = \mu = 0$, and thus, to significantly decrease a number of parameters to be estimated. For the sake of the factors identification, we assume that Σ is diagonal, while Φ is lower trigonal. The parameters in δ_1 are normalized to be 1 in the case of the two-factor model, while in the case of the joint macro-finance model, the inflation factor loading is unconstrained, i.e. $\delta_1 = [1\ 1\ \delta_\pi]$.

We solve the nonlinear optimization problem of maximizing log-likelihood function by using the MATLAB 6.5. routine $fminsearch$ which represents a generalization of the Nelder-Mead simplex algorithm. Finally, we compute the standard errors for the estimated parameters using an approximation of the parameter covariance matrix based on the inverse of the Hessian matrix evaluated numerically.

putationalyy demanding and, in order to obtain closed form solution, a number of strong restrictions must be imposed on RS process.

[8] Imposing No-Arbitrage assumption into VAR by itself creates significant computational problems, which become really huge with additional macro-factors included. The trade-off is to limit certain parameters in $\{\lambda_0, \lambda_1, \delta_0, \delta_1, \mu, \Phi, \Sigma\}$, which is, of course, not the first best solution.

5 Empirical Results

5.1 Parameter Estimates for ATSM

We report the parameter estimates for both ATS models in Table 2.

Table 2 . Parameter estimates for ATSM.

	2-factor ATSM	3-factor ATSM
λ_1	$\begin{bmatrix} -1.0585 & 6.7534 \\ 1.8218 & -4.9356 \end{bmatrix}$	$\begin{bmatrix} 2.5627 & -18.1559 & -3.0091 \\ 0.0311 & -1.9476 & 0.7869 \\ -4.3771 & -0.0551 & 1.7464 \end{bmatrix}$
δ_1	-	1.1476
Φ	$\begin{bmatrix} 0.98507 & 0 \\ 0.027779 & 0.90766 \end{bmatrix}$	$\begin{bmatrix} 0.9279 & 0 & 0 \\ 0.0034305 & 0.98587 & 0 \\ -0.053238 & -1.2307 & 0.8839 \end{bmatrix}$
Σ	$\begin{bmatrix} 0.011451 & 0 \\ 0 & 0.010427 \end{bmatrix}$	$\begin{bmatrix} 0.025531 & 0 & 0 \\ 0 & 0.0014736 & 0 \\ 0 & 0 & 0.015463 \end{bmatrix}$

All factors appears to be higly persistent. Interestingly, our estimates confirm the welknown results by Clarida, Gali, Gertler (2000) of Fed "active" policy since 1980s, since coefficient on inflation in policy rule is larger than one, $\delta_\pi = 1.14$.

5.2 Two factor models and implied VARs of yields

As described above, we consider two modelling strategies for the latent factors: ATSM and Nelson-Siegel approaches. Additionally, we employ simple unrestricted VAR framework based on the observable yields, so that we end up with three different estimates of the companion matrix for $[y^3, y^{60}]$. The estimates of the companion matrices Φ^U, Φ^{NA}, Φ^{NS} from VARs (4), (18), (**??**) are provided in Table 3.

Figure 1: Term premia impied by alternative two-factor models.

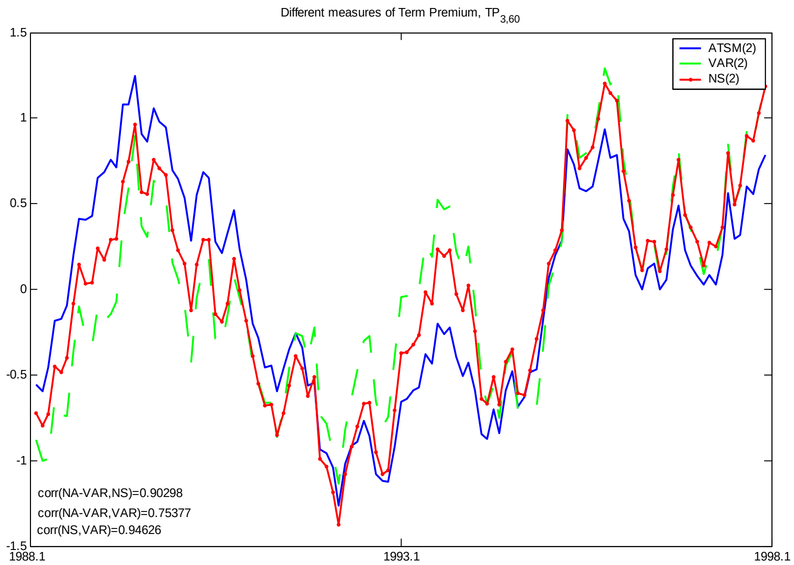


Table 3. Alternative estimates of companion matrix $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$, describing bivariate VAR(1) dynamics of the reference yields $[y^3, y^{60}]$

	Φ^U	Φ^{NA}	Φ^{NS}
ϕ_{11}	0.9555 (0.0236)	0.9564 (0.0056)	0.9596 (0.0256)
ϕ_{12}	0.0601 (0.0339)	0.0588 (0.0152)	0.0507 (0.0367)
ϕ_{21}	0.0061 (0.0322)	0.0238 (0.0066)	0.0088 (0.0357)
ϕ_{22}	0.9652 (0.0463)	0.9363 (0.0177)	0.9564 (0.051)

Note: Sample is 1988.01-1997.12. Standard errors are in parentheses.

An important implication of the estimation is that the corresponding coefficients from three alternative models do not differ significantly. The results clearly show that different VAR-based models will imply similar ET-consistent yields. As a conclusion, Figure 1 shows the striking similarity between the term premia obtained by alternative models.

5.3 Joint macro-finance models

In this section we report the results of estimations using alternative macro-finance specifications of VAR and show that our results are robust. We provide the relevant evidence in Table 4, where we report the results of estimating all our models with inflation included in VAR.

Table 4. Alternative estimates of companion matrix $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}$, describing bivariate VAR(1) dynamics of the reference vector $[y^3,\ y^{60},\ \pi]$

	Φ^U	Φ^{NA}	Φ^{NS}
ϕ_{11}	0.9696 (0.023)	0.88390 (0.0180)	0.9712 (0.0251)
ϕ_{12}	0.1044 (0.0352)	0.1595 (0.1591)	0.0879 (0.0377)
ϕ_{13}	-0.0903 (0.0322)	-0.02290 (0.0134)	-0.0821 (0.0284)
ϕ_{21}	0.0275 (0.0327)	0 –	0.0050 (0.0361)
ϕ_{22}	0.9533 (0.0500)	0.9852 (0.0088)	0.9443 (0.0542)
ϕ_{23}	0.0243 (0.0391)	-0.0283 (0.0007)	0.0265 (0.0409)
ϕ_{31}	0.0216 (0.0248)	0 –	0.0209 (0.0248)
ϕ_{32}	0.0444 (0.0379)	-0.0014 (0.2339)	0.0361 (0.0373)
ϕ_{33}	0.9281 (0.0296)	0.9286 (0.0197)	0.9360 (0.0281)

Note: Sample is 1988.01-1997.12. Standard errors are in parentheses.

Table 5 summarizes the key results of our analysis. The results show that our estimates of term premium are robust both to the choice of the model and to the inclusion of macroeconomic information into the model.

Table 5. Correlations across alternative term premium estimates

	ATSM(2)	ATSM(3)	NS(2)	NS(2)+π	VAR(2)	VAR(3)
ATSM(2)	1.000000	0.878535	0.902975	0.930844	0.753773	0.866314
ATSM(3)	0.878535	1.000000	0.931729	0.886514	0.823887	0.826537
NS(2)	0.902975	0.931729	1.000000	0.988757	0.946256	0.964834
NS(2)+π	0.930844	0.886514	0.988757	1.000000	0.929358	0.978774
VAR(2)	0.753773	0.823887	0.946256	0.929358	1.000000	0.969941
VAR(3)	0.866314	0.826537	0.964834	0.978774	0.969941	1.000000

5.4 Time Consistent Estimation of Term Premium

The VAR-based projections described in the previous sections have some limitations. For all considered models, the VAR is estimated only once on the full-sample and therefore VAR based projections are not based on the information available in real time to agents. Such procedure cannot simulate the investors' effort to use the model in 'real time' to forecast future monetary policy rates, as the information from the whole sample is used to estimate parameters while investors can use only historically available information to generate (up to n-period ahead) predictions of policy rates. In this paper at each point in time we estimate, using the historically available information, a model on and then we use it to project out-of-sample policy rates up to the nth-period ahead. Given the path of simulated future policy rates, we can construct yield to maturities consistent with the Expectations Theory and, as a residual, the term premium.

In this section we simulate the real time decision of agents who forecast policy rates by projecting forward a model to generate long-term yields consistent with the expectations theory. We propose measures for ER_t and TP_t.To construct such measures we estimate at each point in time, using the historically available information, the following model:

$$\begin{aligned} X_t &= \mu + \Phi(L) X_{t-1} + \Sigma \epsilon_t \\ X_t' &= \left[y_t^3, y_t^{60}, \pi_t \right] \end{aligned}$$

We then simulate the estimated model forward, to obtain projection for all the relevant policy rates and to construct ET, which is the ET-consistent long term yield, as follows:

$$\hat{ER}_t = \frac{1}{20} \sum_{j=1}^{19} E[y_{t+j}^3 \mid \Omega_t] \tag{32}$$

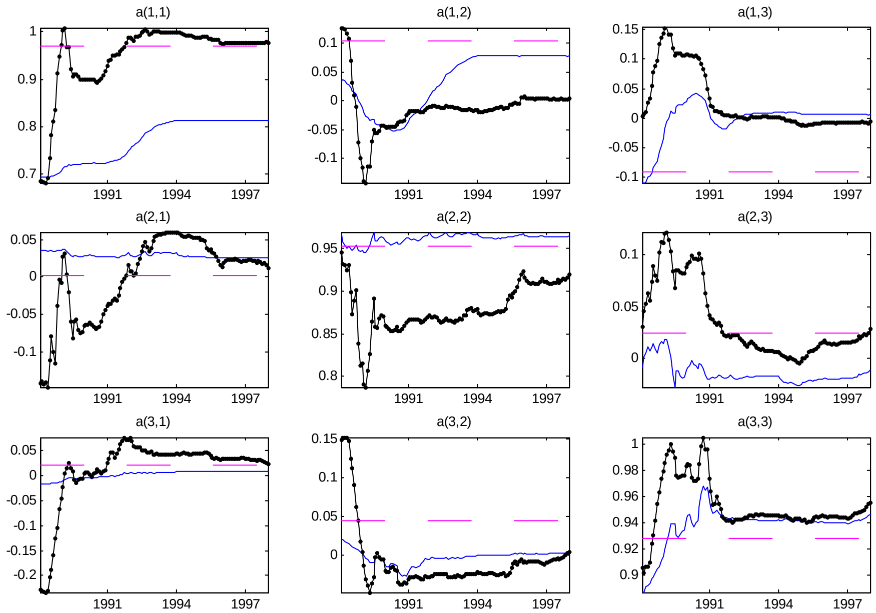
where, $E[y_{t+j}^3 \mid \Omega_t]$ are the VAR-based projections for the future changes in policy rates, hence Ω_t is the information set used by the econometrician to predict on the basis of the estimated VAR model .

Importantly, in implementing our procedure the econometrician uses the same information available to market participants in real-time. Future policy rates at time t are constructed using information available in real time for parameters estimation and forward projection of the model.

Unrestricted VAR and dynamic Nelson-Siegel procedures can easily accommodate the time consistent estimation of the term premium. The time varying parameters of the VAR can be obtained by estimating the model on one sample and then re-estimating it on the next.[9]

However, this is not the case for the ATSM, since the theoretical structure of the model assumes constant parameters.

[9] For the proof of the importance of the time varying coefficients see Figure ??.



Ang and Bekaert (04) propose an alternative solution for time varying parameters of the ATSM model, that is ATSM with Regime Switching (RS). The RS - approach to the ATSM is not new. Dai,Singleton, Yang (03) also develop and empirically implement an arbitrage-free, dynamic term structure model (DTSM) with regime-shifts. With RS, the ATSM become more computationally difficult, nevertheless both models provide considerable limitations. Dai, Singleton and Yang use only finance information in their model. Ang and Bekaert (04), in order to receive a solution in the closed form, assume that only mean and volatility of the variables change across regimes, while the mean-reversion of all variables is not regime-dependent. It is very restrictive assumption and is in odds with the results established by the authors in their previous works.[10] Moreover, Dai, Singleton, and Yang (03) strongly reject the assumption of RS with constant transition probability, which is the case of the model by Ang and Bekaert (04).

6 Conclusion

In this paper we employ alternative discrete term structure models in order to estimate the term premium. In particular, we consider the models, which

[10] Bekaert et al (01) and Ang and Bekaert (02) reported evidence on state-dependent mean-reversion in short rates. Evidence on state-dependent mean-reversion in inflation is reported in Evans and Lewis (95).

characterize the expectations of the future yields by VAR framework. The simple VAR of the observed yields is too limited framework since it does not allow to estimate the term premia for any particular maturity. On the other hand, the dynamic term structure models produce information about the whole yield curve and thus provide a flexible tool for term premium analysis.

Among the most popular dynamic term structure models are the No-Arbitrage Affine Term Structure Models (ATSM). Following the recent tendency, together with standard two-factor ATSM, we consider also joint macro-finance ATSM and enrich the model with macroeconomic information. All types of ATSM impose on the VAR complex cross-equation restrictions due to the no-arbitrage assumption. The model is high dimensional and extremely non-linear. This produces the maximum likelihood function with numerous local optima and implies a difficult optimization problem for ATSM estimation.

On the other hand, we estimate term premium for any maturity by less computationally demanding model (Nelson-Siegel approach) specifying additionally only dynamics of the state vector. We chose the input variables to be identical for the considered TSM, therefore the premia estimates are the functions of the same variables and could be compared directly. Another advantage of the Nelson-Siegel procedures is that it can easily accommodate the time consistent estimation of the term premium, while ATSM does not permit time variability of the parameters.

The main result of our study is that alternative approaches produces the strongly correlated term premia, and thus the less computationally demanding and more flexible method could be used in order to obtain the term premia.

References

[1] Ang, A. and G. Bekaert, 2002, Regime Switches in Interest Rates, Journal of Business and Economic Statistics 20, 2, 163-189.

[2] Ang, A. and G. Bekaert, 2004. The Term Structure of Real Rates and Expected Inflation. Working paper, Columbia University.

[3] Ang, A. and M. Piazzesi, 2003, A No-Arbitrage Vectorautoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, Journal of Monetary Economics 50, 4, 745-787.

[4] Ang, A., M. Piazzesi, and M. Wei ,2004, "What does the Yield Curve Tell us about GDP Growth?," forthcoming Journal of Econometrics.

[5] Berardi, A. and W. Torous, 2002, Does the Term Structure Forecast Consumption Growth?, Working paper, UCLA.

[6] Buraschi, A. and A. Jiltsov, 2005, "Inflation risk premia and the expectations hypothesis," forthcoming Journal of Financial Economics

[7] Campbell, John Y. and Robert J. Shiller, 1987, Cointegration and tests of present value models," Journal of Political Economy 95 (5), 1062—1088.

[8] Carriero, A., Favero C., Kaminska,I., 2005, Financial Factors, Macroeconomic Information and the Expectation Theory of the Term Structure of the Interest Rates, forthcoming Journal of Econometrics.

[9] Clarida, Richard, Jordi Gali, and Mark Gertler, 2000, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory" Quarterly Journal of Economics, 115 (1), 147-180

[10] Cox, J. C., J.E. Ingersoll, S.A. Ross, 1985, " A theory of the Term Structure of the Interest Rates", Econometrica (53), 1985, 385-407

[11] Dai, Q. and K. Singleton, 2000, Specification Analysis of Affine Term Structure Models, Journal of Finance 55, 1943-78.

[12] Dai, Q. and K. Singleton, 2002, Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure, Journal of Financial Economics 63, 415-41.

[13] Dai, Q., K. Singleton, and W. Yang, 2003, Regime Shifts in a Dynamic Term Structure Model of the U.S. Treasury Yields, Working paper, NYU.

[14] Diebold, F.X. and Li., C. ,2002, "Forecasting the Term Structure of Government Bond Yields," Manuscript, Department of Economics, University of Pennsylvania.

[15] Duffee, G.R., 2002, Term premia and interest rate forecasts in affine models, Journal of Finance, 57, 405-443.

[16] Duffie, D. and R. Kan, 1996, A Yield-Factor Model of Interest Rates, Mathematical Finance 6, 379-406.

[17] Elton, Edwin J., 1999, Expected return, realized return and asset pricing tests, Journal of Finance 54 (4), 1199—1220.

[18] Evans, M.D.D., 2003, Real risk, inflation risk and the term structure, Economic Journal, 113, 345-389.

[19] Hamilton, J.D., 1989, A New Approach To The Economic Analysis Of Nonstationary Time Series And The Business Cycle, Econometrica 57, 2, 357-384.

[20] Hamilton, James D. and Dong Heon Kim (2002), "A reexamination of the predictability of economic activity using the yield spread," Journal of Money, Credit, and Banking 34 (2), 340—360.

[21] Hördahl, P., O. Tristani and D. Vestin, 2004, A Joint Econometric Model of Macroeconomic and Term Structure Dynamics, Working Paper, European Central Bank.

[22] Kozicki, Sharon, P.A. Tinsley, 2001, "Term structure views of monetary policy under alternative models of agent expectations", Journal of Economic Dynamics & Control, 25, 149-184

[23] Longstaff, Francis A., 1990, Time Varying Term Premia and Traditional Hypothesis about the Term Structure, The Journal of Finance 45 (4), 1307-1314

[24] McCallum, 1994, "Monetary Policy and the Term Structure of Interest Rates", NBER Working Paper 4938

[25] Nelson and Siegel, 1987, Parsimonious modelling of yield curves, Journal of Business, 60, 473-89

[26] Rudebusch, G.D., and T. Wu, 2004, A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy, Working Paper, Federal Reserve Bank of San Francisco.

[27] Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-4. Discussion paper, Centre for Economic Policy Research(1051).

7 APPENDIX

7.1 Bond Prices

Since the pricing kernel, M_{t+1}, prices all bonds in the economy, for return of any asset

$$E_t(M_{t+1}(1+R_{t+1}))=1,$$

Then the above equation allows bond prices to be computed recursively:

$$P_t(n)=E_t\left\{M_{t+1}P_{t+1}(n-1)\right\} \tag{A1}$$

To keep matters simple, we assume that bond prices are exponential affine functions of X_t', $P_t(n)=\exp(A_n+B\prime_n X_t')$, so that the log prices of bonds with maturity n are given by:

$$p_t(n)=A_n+B\prime_n X_t' \tag{A2}$$

Under the assumption that M_{t+1} is conditionally lognormally distributed and $X_{t+1}$is normally distributed, we can take logs of the Pricing Kernel to obtain

$$p_t(n)=E_t\left\{m_{t+1}+p_{t+1}(n-1)\right\}+\frac{1}{2}Var_t\left\{m_{t+1}+p_{t+1}(n-1)\right\} \tag{A3}$$

The next step is to identify the recursive structure of the coefficients in the bond pricing equation. Putting together all assumptions made above, we get

$$\begin{aligned}
p_t(n+1) &= E_t\left\{m_{t+1}+p_{t+1}(n)\right\}+\frac{1}{2}Var_t\left\{m_{t+1}+p_{t+1}(n)\right\}= \\
&= E_t\left\{-R_{1,t}-\frac{\lambda_t'\lambda_t}{2}-\lambda_t'\varepsilon_{t+1}+A_n+B_n'X_{t+1}\right\}+\frac{1}{2}Var_t\left\{m_{t+1}+p_{t+1}(n)\right\}= \\
&= -R_{1,t}-\frac{\lambda_t'\lambda_t}{2}+A_n+E_t[B_n'(\mu+\Phi X_t+\Sigma\varepsilon_{t+1})]+\frac{1}{2}Var_t\left\{-\lambda_t'\varepsilon_{t+1}+B_n'\Sigma\varepsilon_{t+1}\right\}= \\
&= -\delta_0-\delta_1'X_t-\frac{\lambda_t'\lambda_t}{2}+A_n+B_n'(\mu+\Phi X_t)+\frac{1}{2}Var_t\left\{\left(B_n'\Sigma-\lambda_t'\right)\varepsilon_{t+1}\right\}= \\
&= -\delta_0-\delta_1'X_t-\frac{\lambda_t'\lambda_t}{2}+A_n+B_n'(\mu+\Phi X_t)+\frac{\left(-\lambda_t'+B_n'\Sigma\right)\left(-\lambda_t'+B_n'\Sigma\right)'}{2}= \\
&= \left(\delta_0+A_n+B_n'\mu+\frac{1}{2}B_n'\Sigma\Sigma'B_n-\lambda_0'B_n'\Sigma\right)+\left(-\delta_1'-B_n'\Sigma\lambda_1'+B_n'\Phi\right)X_t
\end{aligned}$$

We get A_{n+1}, B_{n+1} as a solution of the system of difference equations with initial condition $A_1=\delta_0,\ B_1=-\delta_1'$:

$$\left\{\begin{array}{c} A_{n+1}=A_n+B_n'\mu-B_n'\Sigma\lambda_o+\frac{1}{2}B_n'\Sigma\Sigma'B_n \\ B_{n+1}'=B_n'\Phi-\delta_1'-B_n'\Sigma\lambda_1 \end{array}\right\} \tag{A4}$$

The yield on a zero-coupon bond of maturity n, is affine structure of the state:

$$y_t(n)=-\frac{1}{n}(A_n+B\prime_n X_t)\equiv a_n+b_n'X_t \tag{A5}$$

7.2 Maximum Likelihood estimation

We solve the nonlinear optimization problem of maximizing L by using the MATLAB 6.5.routine $fminsearch$ which represents a generalization of the Nelder-Mead simplex algorithm.