Term Structure Forecasting: No-arbitrage Restrictions Versus Large Information set*

Carlo A. Favero\textsuperscript{a}  Linlin Niu\textsuperscript{b}  Luca Sala\textsuperscript{c}\textsuperscript{†}

\textsuperscript{a,c}IGIER, Bocconi University
\textsuperscript{b}WISE, Xiamen University

May 2009

Abstract

This paper addresses the issue of forecasting the term structure. We provide a unified state-space modelling framework that encompasses different existing discrete-time yield curve models. Within such framework we analyze the impact of two modelling choices, namely the imposition of no-arbitrage restrictions and the size of the information set used to extract factors, on the forecasting performance. Using US yield curve data, we find that both no-arbitrage and large info help in forecasting but no model uniformly dominates the other. No-arbitrage models are more useful at shorter horizon for shorter maturities. Large information sets are more useful at longer horizons and longer maturities. We also find evidence for a significant feedback from yield curve models to macroeconomic variables that could be exploited for macroeconomic forecasting.

Keywords: Yield curve, term structure of interest rates, forecasting, large data set, factor models

JEL Classification: C33, C53, E43, E44

\textsuperscript{*}We thank Andrew Ang, Michel van der Wel, Marco Taboga and seminar participants at Bocconi University, Norges Bank for useful comments, WISE Information Center for computational support.

\textsuperscript{†}a,c: IGIER, Università Bocconi, Via Guglielmo Röntgen, 20136 Milan, Italy. E-mail: carlo.favero@unibocconi.it, and luca.sala@unibocconi.it.  
\textsuperscript{b}: WISE, Economics Building A301, Xiamen University, 361005, Xiamen, China. E-mail: llniu@xmu.edu.cn.
1 Introduction

Yields of maturities longer than one period are risk-adjusted averages of expected future short-rates. Short term rates are monetary policy instruments, controlled by central banks. Modelling and forecasting the term structure requires modelling and forecasting risk as perceived by the market and modelling and forecasting future monetary policy rates. This paper provides a unified framework encompassing the many different approaches employed in the literature to model the term structure and provides empirical evidence on their forecasting performance for U.S. data.

The amount of information relevant to modelling and forecasting expected monetary policy and risk is potentially enormous, therefore the choice of a "parsimonious" specification capable of capturing all relevant information is the most crucial step in the modelling strategy. Following this intuition, all the relevant information on pricing bonds at any given point in time is often summarized by a small number of factors. As a consequence, the task of forecasting the term structure is simplified to that of forecasting a small number of factors. Different modelling strategies are defined by the restrictions used to shrink the information available in a few factors (Litterman and Scheinkman, 1991). The traditional finance literature limits the information set to a number of observable yields and uses two alternative methods: extraction of latent factor via cross-sectional interpolation methods and extraction of latent factors by exploiting no-arbitrage restrictions.

Among the cross-sectional interpolation methods, the Nelson and Siegel (1987) approach is the most popular. The Nelson and Siegel three factor model explains most variances of yields at different maturities with a very good in-sample fit. Diebold and Li (2006) have successfully considered the out-of-sample forecasting performance of this model by assuming that the three factors follow AR(1) or VAR(1) processes.

Among no-arbitrage models, the common approach is to assume a linear model for the latent factors and to restrict the factor loadings so as to rule out arbitrage strategies on bonds of different maturities. No-arbitrage restrictions serve not only for reducing the dimension of the parameter space, but also contribute to the theoretical consistency of the model. Dai and Singleton (2000) and Piazzesi (2003) have surveyed the specification issues of affine term structure models in continuous time and discrete time, respectively. Duffee (2002) showed the usefulness of essentially affine term structure models ($A_0(3)^1$) in forecasting.

\(^1\)Essentially affine term structure model, denoted by $A_0(n)$, refers to the modelling
These two approaches have been recently merged in an affine arbitrage-free Nelson-Siegel (AFNS) model, see Christensen, Diebold and Rudebusch (2007) or Le Grand (2007), where the traditional Nelson and Siegel structure, by construction not compatible with no-arbitrage, is slightly modified to rule out arbitrage opportunities.

Models mentioned above are traditionally based only on the information contained in the term structure. Financial markets are clearly not insulated from the rest of the economy. The feedback from the state of the economy to the short term interest rate is explicitly considered in the monetary policy reaction function introduced by Taylor (1993) and by now widely adopted to explain the behaviour of central banks.

Several papers indicate that macroeconomic variables have strong effects on future movements of the yield curve (among others, Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2006) and Rudebusch and Wu (2008)). In particular, Ang and Piazzesi (2003) use an \( A_0(3) \) model, and show that a mixed model (with three latent financial factors plus output and inflation) performs better than a yields-only model in terms of one step ahead forecast at monthly frequency.

One question that naturally arises in this context is how to efficiently summarize the large amount of macroeconomic information available. Factor models suited to deal with large cross-sections have therefore become increasingly popular in the profession. As shown in Stock and Watson (2002) and Forni, Hallin, Lippi, and Reichlin (2005), by decomposing large panels of time series in common and idiosyncratic components, information can be used efficiently, dimensionality greatly reduced and forecasting efficiency improved. Giannone, Reichlin and Sala (2004) show that a two dynamic factor model produces forecasting accuracy of the federal funds rate similar to that of the market.

We will set up an encompassing framework in which we will assess the relative importance of no-arbitrage restrictions versus large information set in forecasting the yield curve. Our framework allows to accommodate the case in which factors are extracted from large information sets.

We choose to evaluate alternative term structure models on the basis of their out-of-sample forecasting performance for different yields. In this way, we will have a uniform ground to compare models with very different features, settings and number of parameters.

We are not the first to make this choice and to provide evidence on the forecasting performance of alternative term structure models: recently feature that in the affine ("A") model, there are "n" state variables, but none of the states drives the conditional variance of the state innovation, hence denoted by subscript "0".
Moench (2008) proposed a no-arbitrage factor-augmented VAR (FAVAR, see Bernanke, Boivin and Eliasz, 2005), in which financial factors are augmented with macroeconomic factors, and compared the forecasting performance of this models with a number of alternatives to show that a no-arbitrage FAVAR delivers almost uniformly better forecasts at horizons from 6 to 12 months ahead. Our exercise differs from that of Moench in many aspects. First, we propose a more general framework to evaluate systematically a larger set of models; second, we base our forecasting comparison exercise on a rolling window estimation with fixed size, in which parameters are re-estimated at each stage, while Moench’s forecast considered a recursive estimation strategy and expanded the estimation window as new observations are included in the sample; third, none of the models we compare is our proposed model; fourth, our empirical results are different in that our evidence is not overwhelmingly in favor of the FAVAR model; fifth, we provide a framework to understand the impact of estimation errors on bond-yield prediction with the no-arbitrage framework. This will help us to understand why this type of models performs unsatisfactorily in the prediction of the long-end of the yield curve.

Concluding, we will also discuss the reverse issue of forecasting macroeconomic variables with term structure models.

The paper is organized as follows. In Section 2 we propose a unified state-space framework to evaluate the effects of incorporating factor information and/or no-arbitrage restrictions on the forecasting performance of empirical models of the yield curve. In Section 3 we describe our data of the US yield curve and factor extraction from a panel of 162 macroeconomic time-series. In section 4 we apply our framework to the data, specify model settings, and evaluate the forecasting performance of various models. The models will differ on how yield factors are extracted and on how additional macroeconomic information is incorporated. Section 5 is devoted to the discussion of our empirical results. Section 6 tackles the issue of forecasting the macroeconomic variables with term-structure models and Section 7 concludes.

2 The general state-space representation

We study the dynamics of the term structure in the state-space model presented in equations (1) and (2). $y_{t,t+n}$ is the yield-to-maturity at time $t$ of a zero-coupon bond maturing at time $t + n$. Yields with different maturities are collected in a vector $y_t = [y_{t,t+1}, y_{t,t+2}, \ldots, y_{t,t+N}]'$. Equation (1) is the measurement equation, in which different yields $y_{t,t+n}$ are assumed to be de-
terminated by a set of state variables, collected in the vector $X_t$. Equation (2) is the state equation in which the states $X_t$ are assumed to follow a VAR(1) process.

$$y_{t,t+n} = \frac{-1}{n} (A_n + B_n' X_t) + \varepsilon_{t,t+n} \quad \varepsilon_{t,t+n} \sim \text{i.i.d.} \mathcal{N}(0, \sigma_n^2)$$  

$$X_t = \mu + \Phi X_{t-1} + v_t \quad v_t \sim \text{i.i.d.} \mathcal{N}(0, \Omega)$$  

The variables in $X_t$ can be either endogenous (that is, some of the elements of $y_t$ are also included in $X_t$) or exogenous, observable or latent.

The system composed of (1) and (2) is very general and can accommodate different specifications. Equation (1) illustrates how the yield curve is fitted. This can be done either by pure interpolation methods or by using models that impose no-arbitrage restrictions. When no-arbitrage restrictions are imposed, the entries of matrices $A_n$ and $B_n$ are constrained by cross-equation restrictions derived from economic theory. In equation (2) different specifications of the information set can be adopted according to the choice of what variables to include in $X_t$. As we shall see below, some models will include only factors extracted from yield curve data, while others will include a combination of factors from the yield curve and factors from macroeconomic data. Additional specifications are possible, as different choices are available to measure the relevant factors.

We take the forecasting performance as the metric to evaluate alternative models. We shall classify models along two dimensions: one will be based on the nature of the restrictions imposed on the measurement equation (1) - interpolation (or reduced form) versus no-arbitrage restrictions; the second will be based on the information set used to model the state dynamics - small versus large information set.

### 2.1 Interpolation versus no-arbitrage models

The number of models of the term structure available in the literature is vast. We select a limited number of models, each of which is to be considered as a representative of a class of models.

1. Diebold-Li model. In Diebold-Li (2006), three factors, extracted à la Nelson and Siegel (NS henceforth), are assumed to follow an unrestricted VAR. In this case we have:

   $$B_n' = \left[ -n, -\left(\frac{1 - e^{-\lambda n}}{\lambda}\right), -\left(\frac{1 - e^{-\lambda n}}{\lambda} - ne^{-\lambda n}\right)\right] \quad \text{and} \quad A_n = 0$$
We denote the three yield factors as $NS_t = [NS_{1,t}, NS_{2,t}, NS_{3,t}]'$ and define $X_t = NS_t$. Equation (1) then takes the form:

$$y_{t,t+n} = NS_{1,t} + NS_{2,t} \left( \frac{1 - e^{-\lambda n}}{\lambda n} \right) + NS_{3,t} \left( \frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n} \right) + \varepsilon_{t,t+n} \quad (3)$$

The dynamics of $NS_t$ is assumed to follow an unrestricted VAR(1):

$$NS_t = \mu + \Phi NS_{t-1} + v_t \quad (4)$$

$NS_{1,t}$, $NS_{2,t}$, and $NS_{3,t}$ are estimated as parameters in a cross-section of yields, letting $n$, the maturity date, vary. In the time series dimension, $NS_{1,t}$, $NS_{2,t}$, and $NS_{3,t}$ have an immediate interpretation as latent factors. The loading on $NS_{1,t}$ is the only element in $B_0$ that does not decay to zero as $n$ tends to infinity; $NS_{1,t}$ can therefore be interpreted as the long-term factor, the level of the term structure. The loading on $NS_{2,t}$ is a monotone function that starts at 1 and decays to zero; $NS_{2,t}$ can be viewed as a short-term factor, the slope of the term structure. $NS_{3,t}$ is a medium term factor: its loading starts at zero, increases and then decays to zero, with the speed of decay determined by the parameter $\lambda$. This factor is usually interpreted as the curvature of the yield curve.

This model will be considered as the benchmark in the class of unrestricted models.

2. No-arbitrage affine models, in which long yields are risk-adjusted expectations of average future short-rates and the coefficients of the state-space model are restricted so as to rule out arbitrage opportunities (see Appendix 1 for details). In this case, we follow the general discrete-time framework popularized after Ang and Piazzesi (2003). Defining the market price for risk associated with the state variables $X_t$ as $\Lambda_t = \lambda_0 + \lambda_1 X_t$ and given the measurement equation of the short rate, $y_{t,t+1} = -(A_1 + B'_{1} X_t) + \varepsilon_{t,t+1}$, it is possible to show that no-arbitrage imposes the following structure on the coefficients of the measurement equation (for $n \geq 1$)²:

$$A_{n+1} = A_n + B'_n (\mu - \Omega \lambda_0) + \frac{1}{2} B'_n \Omega B_n + A_1$$

$$B'_{n+1} = B'_n (\Phi - \Omega \lambda_1) + B'_1$$

²Differently from Ang and Piazzesi (2003) where they define $\Sigma \Sigma' = \Omega$ in the difference equations above, and identify the matrix $\Sigma$: we are only interested in $\Omega$. We do not aim at identifying $\Sigma$, which is not necessary to forecast. This implies that our $\lambda_0$ and $\lambda_1$ have different meaning and scale to theirs, because the pricing kernel we specify is:

$$m_{t+1} = \exp (-r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t - \Lambda'_t v_{t+1})$$

where $v_{t+1} \sim N(0, \Omega)$, while they assume $m_{t+1} = \exp (-r_t - \frac{1}{2} \Lambda'_t \Lambda_t - \Lambda'_t \varepsilon_{t+1})$, where $\varepsilon_{t+1} \sim N(0, I)$. 

6
The restrictions imply that once the coefficients on the short rate equation \((A_1, B'_1)\) are fixed, all the other coefficients for longer maturity yields are determined by the following equations:

\[
B_{n+1} = \left[ \sum_{i=0}^{n} (\Phi' - \lambda_i'\Omega)^i \right] B_1
\]

\[
A_{n+1} = (n + 1)A_1 + \sum_{i=0}^{n} B^{(i)}, \text{ where } B^{(i)} = B'_i(\mu - \Omega\lambda_0) + \frac{1}{2}B'_i\Omega B_i.
\]

In this setup the state vector is assumed to be of dimension 3. Following Chen and Scott (1993), the states are extracted by inverting the measurement equation, assuming that exactly 3 yields are observed without error (see details of this method and the corresponding likelihood function in Appendix 2). The Chen-Scott factors are denoted by \(CS_t = [CS_{1,t} \ CS_{2,t} \ CS_{3,t}]'\). We define \(X_t = CS_t\).

3. We also estimate an affine arbitrage-free Nelson-Siegel (AFNS) model as in Christensen, Diebold and Rudebusch (2007), similar to Le Grand (2007). This model imposes the Nelson-Siegel structure on the canonical representation of affine models, so that

\[
B_n' = \left[ -n, -\left( \frac{1 - e^{-\lambda n}}{\lambda} \right), -\left( \frac{1 - e^{-\lambda n}}{\lambda} - ne^{-\lambda n} \right) \right].
\]

Under the risk-neutral measure, it can be shown that the autoregressive coefficient matrix of the VAR(1) for the states is:

\[
\Phi^Q = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 - \lambda & \lambda \\
0 & 0 & 1 - \lambda
\end{pmatrix}.
\]

In order to exclude arbitrage opportunities, the measurement equation for the yields has to be adjusted by a constant term. Hence, different from the original Nelson-Siegel model, \(A_n \neq 0\).

Although the no-arbitrage Nelson-Siegel (NANS) model is appealing as a unifying framework that links the traditional Nelson-Siegel model to the affine no-arbitrage term structure model, it is significantly restrictive, as it does not allow more or less than three state variables and it does not allow the inclusion of different combinations of state variables, such as observable macro variables. For this reason, we decided to consider these models only in Section 5.3 below, where we address robustness issues.
2.2 State information set: small versus large

A second taxonomy considers small versus large information sets. We will add to the factors extracted from the yields different additional variables and we will compare their effectiveness in forecasting yields.

We define an information set as small if it contains only the yield factors and/or a small set of observed macroeconomic variables. These macro variables will be a measure of inflation and a measure of real activity. The huge literature on Taylor rules has shown that a nominal and a real variable are both important in driving the dynamics of nominal interest rates. We select the annual CPI inflation rate ($\pi_t$) as the measure of inflation, and the annual growth rate of the Index of Industrial Production ($IP_{g_t}$) as the measure of real activity.\(^3\)

1. In the class of small information set models we consider the following specifications.

1a) Unrestricted case. In addition to the $NS_t$ factors, we add $[\pi_t \quad IP_{g_t}]$ as observable factors. The state dynamics is described by a five variable VAR.

1b) Restricted case. In addition to the $CS_t$ factors, we consider $[\pi_t \quad IP_{g_t}]$ as observable factors; as in case 1a), the state dynamics is described by a five variable VAR.

2. In the class of large information set models, we extract common factors from a large panel of macroeconomic variables ($N = 162$). We estimate factors by static principal components, as in Stock and Watson (2002) and we call them $mf_t = [mf_{1,t} \quad mf_{2,t} \ldots mf_{k,t}]'$. We evaluate the forecasting performance of "large $N$" macroeconomic factors in the following specifications:

2a) Unrestricted case. The macro factors are added to the $NS$ factors: $X_t = [NS_t' \quad mf_t']'$.

2b) Restricted case. The macro factors are added to the $CS$ factors: $X_t = [CS_t' \quad mf_t']'$.

2c) The macro factors are used as explanatory variables in a "generalized" Taylor rule (see Bernanke and Boivin, 2003): $X_t = [y_{t,1} \quad mf_t']'$, both in the unrestricted and restricted models.

2d) The states are the macro factors: $X_t = mf_t$

We will employ two to four macro factors in our analysis.

\(^3\)We have also considered the unemployment rate as an alternative indicator of real activity. However this variable is outperformed by the index of industrial production for forecasting purposes.
Before discussing the empirical results some remarks are in order. The state space representation is so general and flexible that it can accommodate very different specifications not mentioned within our limited context. The advantage of using such an encompassing framework will be immediate when we report results along the two dimensions discussed.

There, we will clearly see what are the elements driving the different results and what is the role of each of them. For example, by going from reduced form to no-arbitrage model, given the information set and a similar number of parameters, we will see the role of no-arbitrage restrictions on the forecasting performance; by changing the risk price specification, given the information set, we will be able to study how sensitive the results are to the form of risk price; by adding different set of macro factors, while holding constant the number of yield factors, we will see how the macro factors affect the forecast performance.

3 Data and macroeconomic factors

Our basic data set consists of Bliss data set of zero-coupon equivalent US yields (1974:2-2003:9) at the following 11 maturities: 1-month, 3-month, 6-month, 9-month, 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year. The dynamics of the yield curve is shown in Figure 1.

We extract macroeconomic factors from a panel of 162 US macro monthly time series for the sample 1974:2-2003:9. The data set is the same as used in Giannone, Reichlin and Sala (2004). We have excluded nine interest rates from the original 171 series.

<table>
<thead>
<tr>
<th># of series</th>
<th>Categories</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>IP indices</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>39</td>
<td>Labor market</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>17</td>
<td>Sales, consumption spending</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>12</td>
<td>Inventory and orders</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>22</td>
<td>Financial markets, money and loans</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>25</td>
<td>Price indices</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>3</td>
<td>Import &amp; export</td>
<td>ln (X_t - \ln X_{t-12})</td>
</tr>
<tr>
<td>23</td>
<td>Capacity utilisation and inventory indices, etc.</td>
<td>(X_t)</td>
</tr>
</tbody>
</table>

The common factors have been extracted from the macro panel as follows. First, we transform the data to obtain stationarity. We take annual log-difference for the series that contain trends (production indices, price indices...
including asset prices, money stock, etc.) while series stationary by their nature (capacity utilization, sentiment indicators, etc.) are considered in levels.

Second, we estimate the factors by principal components (Stock and Watson (2002)). We rank the factors according to their explanatory power and consider up to the fourth in our analysis.

As reported in Table 1, the first four factors explain up to 68% of the total variance in the panel. In Figures 2 and 3 the two macro variables and the first four macro factors extracted are plotted. As can be seen, the dynamics of the first two factors are closely related to IP growth rate and CPI inflation rate respectively. The first factor highly correlates with output growth. The R-squares of a regression of various industrial production indices on the first factor are higher than 0.9 as shown in Table 1. The second factor closely follows inflation: it explains around 80% of variations in annual growth rates of PPI crude materials, CPI housing and CPI services as shown in Table 1. The third and fourth factor are more related to financial variables and to the effective exchange rate.

4 Estimation specification and forecast evaluation

4.1 Estimation specification

In all models considered, we assume the state dynamics to follow a VAR(1).

In the unrestricted models, we do not make restrictions on the parameters, hence a two-step OLS suffices to estimate the state and measurement equations.

In the restricted models, we make specifications according to how the state vector is composed of yield factors and macro factors.

In the restricted model with only latent yield factors, we assume that the factors have zero mean \( \mu = 0 \), and that the VAR coefficient matrix \( \Phi \) is lower triangular, with \( \Omega = I \). This is the most general identified representation for the class of essentially \( A_0(3) \) models (Dai and Singleton (2000)). In addition, in the short rate equation, \( y_{t,t+1} = -A_1 - B_1 X_t + \varepsilon_{t,t+1} \), we restrict \(-A_1 = \)

\[ \begin{array}{c}
4 \text{This is different from Ng and Ludvigson (2006). They construct a composite factor by combining several common factors according to their in sample significance in explaining the bond risk premia. We have tried to rank the factors according to their contribution to R-squares of yields, but we did not find clear evidence suggesting that such a strategy improves out-of-sample forecasting of yields.}
\end{array} \]
\( \bar{r} + B_1 \bar{X}, \) where \( \bar{r} \) is the historical mean of the short rate. In the restricted models with latent factors, we use the Chen-Scott method (see Appendix 2 for details) and estimate the model with maximum likelihood.

In the restricted model with large information set, when the state vector is assumed to be composed of \( CS_t \) yield factors and observable macro variables or factors, we assume the specification proposed by Pericoli and Taboga (2008), where no restrictions are imposed on the VAR coefficient matrix \( \Phi \), while the following general conditions need to be met:

1) the covariance matrix \( \Omega \) is block diagonal with the block corresponding to the unobservable yield factor being identity, and the block corresponding to the observable factors being unrestricted, i.e.

\[
\Omega = \begin{bmatrix}
I & 0 \\
0 & \Omega^o
\end{bmatrix}
\]

2) the loadings on the factors in the short rate equation are positive, \(-A_1 \geq 0\);
3) \( X^u_0 = 0 \).

Early specifications of this model often impose zero restrictions on the VAR coefficient matrix \( \Phi \) (in Ang and Piazzesi (2003), \( \Phi \) is block diagonal). These assumptions impose strong restrictions on the interaction between yield factors and macro factors, which if true would enhance the forecasting performance of the model. We have tried various specifications in our forecasting exercises. The result is that the general specification, despite its heavier paramerization, does not have an inferior forecasting performance\(^5\). We hold to the general specification below.

In these mixed factor models in which state vector contains latent yield factors, we also use the Chen-Scott method to estimate the model with maximum likelihood.

In the restricted model with only observable states, we follow a two step procedure. We first estimate the VAR for the states, then given \( (\hat{\mu}, \hat{\Phi}, \hat{\Omega}) \), we estimate the prices of risk, \( \lambda_0 \) and \( \lambda_1 \).

For all the above no-arbitrage models, we estimate three specifications for risk prices.

- Constant prices of risk: \( \lambda_0 \neq 0 \) and \( \lambda_1 = 0 \).

\(^5\)Additional empirical evidence is available upon request.
• Time-varying prices of risk, $\lambda_1 \neq 0$, $\lambda_1$ diagonal. This assumption, together with a diagonal $\Omega$, implies that prices of risk are independent and has been employed in Ang, Bekaert and Wei (2007).

• The third one is most general, with $\lambda_1$ being a full $K \times K$ matrix, $K$ denoting the number of state variables. As discussed in Dai and Singleton (2002), this specification is the combination of nonzero factor correlations through the matrix, $\Phi$, and state-dependent market prices of risk, $\lambda_1$.

In the restricted model with Nelson-Siegel factors, we estimate a correlated-factor AFNS model as in Christensen, Diebold and Rudebusch (2007), but for parsimony we choose a diagonal variance-covariance matrix $\Omega$, similar to Le Grand (2007). In this case, the constant term in the measurement equation of yield takes the following form

$$A_n = \sigma_{11}^2 \frac{n^3}{6} + \sigma_{22}^2 \left[ \frac{n}{2\lambda^2} - \frac{1 - e^{-\lambda n}}{\lambda^3} + \frac{1 - e^{-2\lambda n}}{4\lambda^3} \right] + \sigma_{33}^2 \left[ \frac{n}{2\lambda^2} + \frac{ne^{-\lambda n}}{\lambda^2} - \frac{n^2e^{-2\lambda n}}{4\lambda} - \frac{3ne^{-2\lambda n}}{4\lambda^2} - \frac{2(1 - e^{-\lambda n})}{\lambda^3} + \frac{5(1 - e^{-2\lambda n})}{8\lambda^3} \right]$$

4.2 Forecast procedure

In this section, we discuss how forecasts are computed. We obtain multiple step ahead forecasts for the states by iterating the one-step model forward:

$$\hat{X}_{t+h|t} = \sum_{i=0}^{h} \Phi^i \hat{\mu} + \hat{\Phi}^h X_t$$

Forecasts based on different specifications are computed as follows:

4.2.1 Unrestricted models

1. Diebold and Li (2006). We obtain the Nelson and Siegel factors from equation (3). We fix $\lambda$, the parameter governing the speed of decay in the exponential function, at 0.077, as calibrated in Diebold, Rudebusch and Aruoba.

---

6 The alternative would have been to obtain direct forecasts from projecting $h$-step ahead: $\hat{X}_{t+h|t} = \hat{\mu}_h + \hat{A}_h X_t$. We find that except the Taylor Rule VAR with two macro variables or two macro factors, iterated forecast does always better than direct forecast. Further, the restrictions on the no-arbitrage model are such that only iterated forecast can be computed.
Figure 4 shows the NS factors as well as their empirical correspondents. The first NS factor closely represents the 10 year yield. The second NS factor correlates well with the spread between long and short yields: $(10\text{year} - 1\text{month})$. The third NS factor is close to: $(2 \times 2\text{year} - (10\text{year} + 3\text{month}))$, a measure of curvature of the yield curve.

After having extracted the factors and estimated the unrestricted VAR(1), we obtain forecasts by iterated projections:

$$\hat{N}S_{t+h|t} = \sum_{i=0}^{h} \hat{\Phi}^i \hat{\mu} + \hat{\Phi}^h \hat{N}S_t$$

and by using the NS parameterization:

$$\hat{y}_{t+h|t} = \hat{N}S_{1,t+h|t} + \hat{N}S_{2,t+h|t} \left( \frac{1 - e^{-\lambda n}}{\lambda n} \right) + \hat{N}S_{3,t+h|t} \left( \frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n} \right)$$

2. Diebold-Li financial factors plus macro variables/factors. The Nelson-Siegel factors are extracted as before. The state vector becomes: $\hat{X}_t = [\hat{N}S_t \ z_t]'$, where $z_t$ contains the macro information and is modeled as a VAR(1). In this case, we disregard the specific form of the NS interpolants and project directly the yields on the states:

$$\hat{y}_{t+h|t} = \hat{a} + \hat{b} \hat{X}_{t+h|t}$$

$$\hat{X}_{t+h|t} = \sum_{i=0}^{h} \hat{\Phi}^i \hat{\mu} + \hat{\Phi}^h \hat{X}_t$$

We find that this leads to better forecast than fixing the loadings to those implied by equation (3). A possible interpretation of this is that the good forecasting performance of the Diebold-Li model largely comes from the NS factor extraction in the first step and not from the restrictions imposed by the specific functional form of the loadings.

3. Interest rate rule-type VAR in which the state equation is unrestricted. In this setting, the yields are directly projected onto the states. Both the measurement and the state equations are estimated with OLS as in (5).

---

7The factors extracted are insensitive to the choice of $\lambda$. A robustness check, estimating $\lambda$ period by period in the forecast procedure, is made in Section 5.3.
4.2.2 No-arbitrage models

The forecast in no-arbitrage models is obtained by using equation (5) in which the parameters are subject to the no-arbitrage restrictions.

4.3 Forecast comparison

We set the sample size at 180 periods. We implement rolling estimation by moving the sample forward one observation at a time and re-estimating the model at every step, starting from the sample period 1978:1-1992:12. We consider 4 forecasting horizons (denoted by $h$): 1 month, 6 months, 12 months, and 24 months. For the 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods; for the 6 month ahead forecast, we end up with a total of 124 forecasts, and so on, up to the 24 month ahead forecast, for which we end up with 106 forecasts.

We choose two measures of forecasting performance. One is the ratio of the root mean squared forecast error (RMSFE) of each model to the RMSFE of a random walk forecast. We show the comparison of forecasting results from different models in Table 2. The table shows better forecasts with respect to the random walk with bold characters for the range of $[0.9, 1)$, with added shading background for the range of $[0.8, 0.9)$, and with added underline for ratios smaller than 0.8.

RMSFE ratio tells the relative accuracy of each model for each maturity-horizon forecast compared to random walk. Forecast errors originate mainly from two sources: errors in forecasting yield levels, and errors in forecasting changes in yields. The RMSFE does not distinguish between these two types of errors. When models are subject to structural change and/or parameter instability in the mean, even if they can forecast changes in yields relatively well, the forecasted levels might deviate from the realised data persistently so that the RMSFE will be high. To take care of this problem, we complement the information in the RMSFE with a Bayesian probability indicator which rewards models that better predict yield changes and controls for biases in the forecasts.

We construct this indicator by taking observed yields as the dependent variable in a regression. We use as regressors two forecasts at a time: one from one of the $\bar{k}$ models discussed above and one from the random walk model. We assess the posterior probability of being included in the regression for each of the two forecasts. We then calculate the ratio of the posterior probabilities of the selected model forecast to be included in the general
regression to the random walk forecast. We repeat the above procedure for all the models discussed above. Let us go into the details of the procedure (a more general discussion can be found in Koop, 2003).

Define \( y = (y_1,...,y_t,...,y_T)' \) the realised yield at a specific maturity, where \( T \) is the total number of forecast.

We similarly define \( \hat{y}_k = (\hat{y}_{k,1},...,\hat{y}_{k,t},...,\hat{y}_{k,T})' \) as the vector collecting the forecasts from model \( k \).

We consider 2 forecasts at a time, the random walk forecast, \( \hat{y}_{RW} \), and the forecast from model \( k \), \( \hat{y}_k \) and define \( X_k = [\hat{y}_{RW} \hat{y}_k] \) as a \( T \times 2 \) matrix which contains the 2 forecasts of \( y \) at a specific forecast horizon \( h \).

Given \( X_k \), we can define three nested models \( M_r \) \((r = 1, 2, 3)\) where \( M_3 \) is the most general one:

\[
M_1: \quad y = \alpha + \hat{y}_{RW}\beta_1 + \varepsilon \\
M_2: \quad y = \alpha + \hat{y}_k\beta_2 + \varepsilon \\
M_3: \quad y = \alpha + \hat{y}_{RW}\beta_1 + \hat{y}_k\beta_2 + \varepsilon
\]

\( \alpha \) is the intercept and \( \varepsilon \) is a \( T \times 1 \) vector of errors, which is assumed to be a priori distributed as \( N(0_T, h^{-1}I_T) \). We assume all nested models have the same prior probability \( p(M_r) = 1/3 \).

This also implies there is equal probability of each regressor being included in the general model. For each model with \( \hat{k} \) regressors, we impose a relatively non-informative Normal-Gamma conjugate prior on the coefficients:

\[
\gamma \sim N(\gamma, h^{-1}V) \\
h \sim \Gamma(s^{-2}, v)
\]

where \( \gamma = (\alpha, \beta) \), by assigning:

\[
\alpha = 0, \beta_i = 1/\hat{k} \\
V \text{ diagonal, } V(1,1) = 25 \text{ and } V(j,j) = 1 \ (j > 1); \text{ and} \\
s^{-2} = 0.5 \text{ and } v = 4
\]

\(^8\)We do not explicitely write the index \( h \) to save on notation.

\(^9\)There are in principle \( 2^2 \) possible subsets of \( X_k \). We exclude the empty set and focus on those subsets which contains at least one yield model forecast. For our comparison, whether or not to include the empty set will not affect the results. Excluding it speeds up the computations.
These priors are chosen as such so that the implied prior variance of $\alpha$ is 100, and the implied prior variance of $\beta_i$ is 4. The prior means of $\alpha$ and $\beta_i$ imply that in each nested model the coefficients of different forecasts sum up to one and there is no persistent bias.

For each of the three nested models we calculate the posterior probability:

$$p(M_r|y) = \frac{p(y|M_r)p(M_r)}{\sum_{r=1}^{3} p(y|M_r)p(M_r)}$$

By integrating out the model posteriors, we can obtain the posterior probability that regressor $\hat{y}_k$ is contained in the forecast regression model:

$$p(I_k = 1) = \sum_{r=1}^{3} I_{k,r}p(M_r|y).$$

Our measure, the Bayesian Model Averaging Indicator (BMAI), will be the ratio:

$$BMAI_{k,RW} = \frac{p(I_k = 1)}{p(I_{RW} = 1)}$$

which can be written as:

$$BMAI_{k,RW} = \frac{p(M_2|y) + p(M_3|y)}{p(M_1|y) + p(M_3|y)}$$

If the ratio is bigger than 1, it means that regressor $\hat{y}_k$ has a higher probability than $\hat{y}_{RW}$ to be included in the forecasting regression. In Table 3 we report results. We highlight the ratio with a shaded area when it is bigger than 2, and add underline when it is bigger than 5\textsuperscript{10}.

When $\alpha$ is set to 0 and $\beta_i$ is assumed to sum to 1, the mean of the priors of the model in equation (8) is similar to a standard forecast comparison regression, as in Stock and Watson (1999):

$$y = \lambda \hat{y}_1 + (1 - \lambda) \hat{y}_2 + \varepsilon. \quad (9)$$

The regression in equation (9) is too restrictive if the forecasts are biased, which may be the case for some yield models. In such occasion, the indicator might penalize the forecast with higher systematic bias, attributing little weight to it. The fact that a model delivers biased forecasts does not mean that it does not have forecasting power. By allowing for $\alpha$ different from zero, we take into account the biases in the forecasts and identify models with positive bias but high correlation between forecasts and actual yields.

\textsuperscript{10}We have checked the robustness of these results to a range of different parameterizations for the relevant prior distributions.
5 Empirical Results

We present our empirical results by discussing first our main evidence on yield curve forecasting. We identify the nonlinear impact of estimation/specification errors on forecast errors across the yield curve under no-arbitrage restrictions, which explains unsatisfactory performance of no-arbitrage models under certain specifications. We then use the models with Nelson-Siegel factors to analyse some specific issues that could be driving forecasting performance.

5.1 Yield curve forecast: no-arbitrage and/or large information set?

We report our main results on the forecasting performance in Table 2 for the RMSFE ratio and in Table 3 for the BMAI, respectively. In each table, we select 16 representative models and put them into 6 rows and 3 columns according to their characteristics. We report each model’s forecasting performance in one sub-table along yield maturities (1, 24, and 120 months) and forecast horizons (1, 6, 12 and 24 months). Above each sub-table, we indicate the state vector.

The models are compiled along two dimensions. Horizontally, we compare reduced form versus no-arbitrage models. Among no-arbitrage models we present the results of two specifications: constant risk price and time-varying risk price with $\lambda_1$ a diagonal matrix. We didn’t list the results of time-varying risk price with full $\lambda_1$, because the results of this specification are the worst performing among the three. Interested readers can refer to our appendix file\(^{11}\) for related results. Vertically, we compare small information set models with large information set ones. For the class of small information set, we present three yield factor models and three yield factor models augmented with macro variables. For the class of large information set, we present a model with three yield factors and macro factors, a model with one yield factor (the short rate) and macro factors, and lastly a model with macro factors only. We can therefore compare how macro factors fare compared with macro variables, and what is the relative role of yield curve versus macroeconomic information.

Among these models, some are existing in the literature, others are similar to existing models with minor differences in their specifications. If we denote the model in row $m$ and column $n$ by $(m, n)$, then model (1,1) is the Diebold-Li model with VAR(1) states; models (1,2) and (1,3) are $A_0(3)$ models with constant risk price and time-varying risk price respectively; model (2,1) is

\(^{11}\)We provide the appendix file upon request.
similar to the yield-macro model of Diebold, Rudebusch and Aruoba (2006) but with a different set of macro variables; models (2,2) and (2,3) are similar to Ang and Piazzesi (2003), with inflation and IP growth as explicit macro variables, allowing for the interaction between the dynamics of yield and macro variables. Models in the third and fourth rows have three yield factors plus macro factors extracted from large macro panel. Models relative to entries (4,2) and (4,3) are not reported; those mixed-factor models have three latent factors intertwined with too many parameters, estimation uncertainty is very high, and results less reliable. Models (5,2) and (5,3) are Moench (2008) type models; model (5,1) is a reduced form version of it. We put in the sixth and last row models with macro factors only, to complement the information in the first row, in which models with only yield factors is presented. This gives us an interesting comparison on the information content in yield curve prediction.

Let us now proceed to examine our results through several routes. We will start from the Diebold-Li model and move along the two dimensions previously discussed.

5.1.1 Reduced form versus no-arbitrage restrictions.

Moving from left to right in the first row of Table 2, we find that in forecasting the short rate all three models have similar performance and they all beat the random walk model at all horizons. Compared with the Diebold-Li model, the no-arbitrage models with constant risk price has better forecast for medium and long term yields at all forecast horizons.

The forecasting performance of the no-arbitrage model with time-varying risk prices deteriorates, the more so, the longer the maturity: the no-arbitrage restrictions help to reduce the RMSFE as long as the risk price is constant.

When we look at the second row of Table 2 where macro variables are added to the three yield factors, the same pattern remains.

Does this findings dismiss the usefulness of time-varying risk price in forecasting yields? Not necessarily. We will go back to this issue when discussing the forecasting performance as measured by the BMAI criterion below.

5.1.2 Small information set versus large information set.

Moving downwards the first column of Table 2, we observe that the model in entry (2,1), in which the inflation rate and the IP growth rate are added to the three Nelson-Siegel factors, does not improve upon Diebold-Li in terms of RMSFE.
When two latent macro factors are introduced (as in model (3,1)), there is
not much improvement with respect to Diebold-Li. If compared with model
(2,1), there is a clear tendency towards improvement for yields with longer
maturities at longer forecast horizons. The two factors extracted from the
large panel of macro variables seem to better capture the real and the nominal
dimension of the economy than the two observable macro variables. When
three Nelson-Siegel factors are augmented with four macro factors, in spite
of the larger number of parameters, forecast at 6 to 24 month horizons of all
yields are improved compared with the model (2,1), and forecast for short
and medium term yields at 12 to 24 month horizons are better than both
Diebold-Li model and the random walk.

Moving to model (5,1) and (6,1), the advantage in forecasting short and
medium term yields at 12 to 24 month horizons remains, while the RMSFE
ratios increase substantially at short horizons across the yield curve as the
number of yield factors decreases. Kim (2007) discusses a similar finding
when yield factors are replaced with observable macroeconomic variables.

We detect a clear pattern: macro information tends to improve yield
forecasts at longer horizons; the macro factors extracted from the large macro
panel have robust forecasting power at 12 to 24 months ahead, and this is
ture even in a VAR without yield curve factors. While previous research such
as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006) find
evidence for the importance of macroeconomic information in forecasting
the yield curve, our results are even clearer, thanks to the design of our
experiments, in which the modification of one element at a time allows to
identify the specific role of additional information.

5.1.3 No-arbitrage restrictions versus large information set.

Moving from the Diebold-Li model to either direction on Table 2, both the
no-arbitrage restrictions and large information set have some value added in
forecasting the yield curve. The benefits work through different mechanisms
and have different impacts. Can we explore both at the same time? The
block of "no-arbitrage restricted" and "large N" models in Tables 2 and 3
gives us some hints.

Model (3,2) is a no-arbitrage restricted model with constant risk price,
with 3 latent yield factors and 2 macro factors. The restrictions reduce the
RMSFE ratios for medium and long term yields compared to model (3,1)
which has the same information set. For medium and long term yields at
12 and 24 month ahead forecast horizons, the large information set improves
upon the no-arbitrage models (1,2) and (2,2) with small information set. This model seems to enjoy the benefits of both imposing no-arbitrage restrictions and expanding its information set.

As we introduce time-varying risk price in model (3,3), although the short rate forecast in term of RMSFE ratios tends to improve for all horizons, the forecast of medium and long term yields deteriorates.

However, going to model (4,2) or (4,3) is too difficult to maintain the improvement given the technical challenge and model uncertainty.

Moench (2008) takes a different approach to combine no-arbitrage with macro factors; he substitutes latent factors extracted from the yield curve with latent macro factors and keeps the short rate as the single yield factor. The VAR for the states can then be interpreted as a "generalised Taylor rule".

Models (5,2) and (5,3) are simplified versions of Moench models, with less parameters in the risk price equations and less lags in the VAR. Due to the two step procedure and to the fact that the 1-month rate $y_{t,t+1}$ is a variable in the VAR, the forecast of $y_{t,t+1}$ is the same at all horizons in both no-arbitrage models and the same as in the unrestricted model (5,1). The pattern of the forecast performance of these models is similar, i.e. the long end of the yield curve is poorly predicted, but at longer forecast horizons the RMSFE ratio decreases, probably due to the presence of the macro factors.

Comparing the three models in the fifth row, the no-arbitrage models fare worse than the unrestricted one; among the no-arbitrage models, time-varying risk price tends to improve the RMSFE compared to the constant risk price model. The model in Moench (2008) did a better job at the long end of the yield curve, probably because of more lags in the VAR, different time-varying risk specification or different period chosen. In our framework on the contrary, the no-arbitrage models cannot improve upon the reduced form model.

Models (6,2) and (6,3) with pure macro factors under no-arbitrage restrictions strengthen the above findings with further evidence.

Let us now turn to discuss the evidence on the BMAI indices in Table 3. Models are collected in the same order as of Table 2.

In the first row of Table 3, the BMAI ratios are mostly around 1 for all three models compared with random walk. However, for the 10 year yield at 24 month forecast horizon, the ratios are above 5 for both no-arbitrage models; the time-varying risk price model does even better than the constant risk price model at 2 year horizon. In the second row of Table 3, the constant
risk price no-arbitrage model does not outperform neither the random walk nor the reduced form model, but the time-varying risk price model does better than the random walk at 1 to 2 year horizons, and better than the reduced form model at these horizons except for one entry. The evidence of a better performance of time-varying risk price models is still present when we move down to the third row when macro factors are added to the state vector.

Moving downwards the first column of Table 3, when the three Nelson-Siegel factors are augmented with the inflation rate and the IP growth rate, there is improvement in the BMAI ratios in the 24 month ahead forecasts, especially for medium and long term yields, with respect to the Diebold-Li model. When two macro factors are added to the model (3,1), there is improvement both in the 12 and 24 month ahead forecasts. This again indicates that the two factors extracted from the large macro panel seem to better capture the real and nominal dimensions of the economy than the observable macro variables (as in Giannone, Reichlin and Sala, 2004). When four macro factors are augmented with three Nelson-Siegel factors, the forecasts at 12 to 24 month horizons of short and medium yields improve when compared to the model (2,1), although this improvement does not extend to long term yields. Moving to models (5,1) and (6,1) where the yield factors are excluded from the state vector but four macro factors are used, the improvement in forecasting short and medium term yields at 12 to 24 month horizons remains. In this case, the noisy forecast that had a high RMSFE ratios are not punished by the BMAI ratio at short horizons.

In order to visualise some of the above findings, we plot the forecast of yields from models in the first and last rows of Table 2 in Figure 5. Figure 5-1 to Figure 5-3 display 1-month ahead, 12-month ahead and 24-month ahead yield forecast respectively. On each figure, the left column shows forecast from models with 3 yield factors, together with random walk forecast and the realised yields. The right column shows forecast from models with 4 macro factors, together with realised yields. The upper panels are for 1-month yield, y(1), middle panels for 2-year yield, y(24), and lower panels for 10-year yield, y(120). Some observations arise which confirm some of the findings above when using the RMSFE ratios and BMAIs.

Figure 5-1 shows that for 1-month ahead forecast, models with macro factors only (those on the right column) tend to mispredict the yield levels persistently when compared to the three yield factor models (those on the left column), the more so, the longer the maturity. When the forecast horizon increases, the persistent bias of models with macro factors become less severe, while performance of yield factor models deteriorates, as can be seen in
Figures 5-2 and 5-3. At 12-24 month ahead, the prediction of $y(1)$ and $y(24)$ from models with macro factors traces the realised yields even better than yield factor models. The unrestricted model (blue solid line) tends to perform better than models with no-arbitrage (red line with circle for constant risk price and green line with cross for time-varying risk price).

On each figure, there are often persistent differences between forecasts with constant risk price and those with time-varying risk price. In addition, the longer the maturity of yields, the bigger is the difference in forecasts. On the left column (that shows models with three yield factors), forecast errors from the constant risk price are smaller than those from the time-varying risk price models.

Overall, our results can be summarized as follows:

(a) no-arbitrage restrictions with constant risk price in a three yield factor model generate low RMSFE;
(b) macroeconomic factors are useful at predicting the yield curve at longer horizons;
(c) among no-arbitrage models, the RMSFE is systematically high for the longer maturities and time-varying risk price is useful in capturing the future movement of the yield curve;
(d) in models in which the state vector is composed of observable macro factors with no or little information on the yield curve, yields of longer maturities are poorly predicted; the forecasts worsen when no-arbitrage restrictions are imposed.

Why is it difficult to gain in forecasting by using the theoretically consistent and tractable restrictions of no-arbitrage under time-varying risk price in a three yield factor model? What has hindered the value of combining no-arbitrage with observable macro factors? What makes the prediction errors higher at longer maturities? In what follows, we will shed some lights on these issues by analysing the sensitivity of yield curve coefficients in the no-arbitrage model.

5.2 No-arbitrage restrictions and the nonlinear and non-equal prediction errors across the yield curve.

No-arbitrage restrictions lead to parsimonious and consistent modelling of bond yields. However, the nonlinearities involved in the coefficients of the yield equations imply that bond yield predictions can be very sensitive to any estimation or specification error. In general, in these models the impact of
estimation or specification error on yield curve estimation and prediction is not evenly distributed across the yield curve. This is due to the nonlinearity in the bond pricing formula which links the fundamental parameters to the loadings on the yield curve factors.

We use a one factor affine term structure model to illustrate this point.

\[
y_{t,t+n} = a_n + b_n x_t + \varepsilon_{t,t+n}, \quad \varepsilon_t \sim i.i.d. N(0, \sigma^2 I) \tag{10}
\]

\[
x_t = \mu + \phi x_{t-1} + v_t, \quad v_t \sim i.i.d. N(0, \omega^2) \tag{11}
\]

with parameters \( \theta = (\mu, \phi, \omega^2, a_1, b_1, \lambda_0, \lambda_1) \). We can derive:

\[
b_{n+1} = \frac{1}{(n+1)} \left[ \sum_{i=0}^{n} (\phi - \lambda_1 \omega^2)^i \right] b_1
\]

\[
a_{n+1} = a_1 + \frac{1}{(n+1)} \sum_{i=1}^{n} \left[ i b_i (\mu - \omega^2 \lambda_0) - \frac{1}{2} i^2 b_i^2 \omega^2 \right]
\]

If we suppose the true parameter vector is \( \hat{\theta} \), any estimation error \( \Delta \hat{\theta} \) will generally have different effects on the yield curve coefficients because of the nonlinear structure of the above difference equations. Let us evaluate the impact around \( \hat{\theta} \), and define \( \Delta_b \) and \( \Delta_a \) as the marginal effects on \( b_{n+1} \) and \( a_{n+1} \) of errors in \( b_1 \) and \( \mu \):

\[
\Delta_b \equiv \frac{\partial b_{n+1}}{\partial b_1} = \frac{1}{(n+1)} \sum_{i=0}^{n} (\hat{\phi} - \hat{\lambda}_1 \hat{\omega}^2)^i \\
\Delta_a \equiv \frac{\partial a_{n+1}}{\partial \mu} = \frac{1}{(n+1)} \sum_{i=1}^{n} ib_i
\]

\( \Delta_b \) and \( \Delta_a \) can be regarded as two kernels, because the impact of estimation errors from other parameters are mostly proportional to them:

\[
\frac{\partial b_{n+1}}{\partial \phi} = \hat{b}_1 \Delta_b \\
\frac{\partial b_{n+1}}{\partial \lambda_1} = \hat{b}_1 \hat{\omega}^2 \Delta_b \\
\frac{\partial b_{n+1}}{\partial \omega^2} = -\hat{b}_1 \hat{\lambda}_1 \Delta_b \\
\frac{\partial a_{n+1}}{\partial \lambda_0} = \hat{\omega}^2 \Delta_a \\
\frac{\partial a_{n+1}}{\partial \omega^2} = -\lambda_0 \Delta_a - \frac{\hat{b}_1 \hat{\lambda}_1}{n+1} \Delta_b \sum_{i=1}^{n} \left[ i (\hat{\mu} - \hat{\omega}^2 \hat{\lambda}_0) - i^2 b_i \hat{\omega}^2 \right] - \frac{1}{2(n+1)} \sum_{i=1}^{n} i^2 b_i^2
\]
Assuming the risk adjustment term $\lambda_1 \omega^2$ is fairly small compared to the autoregressive coefficient $\phi$, then $\Delta_b$ is either hump shaped or upward sloping when $\phi$ is close to 1. If $\hat{b}_{n+1}$ is increasing in $n$, $\Delta_a$ is also an increasing function of $n$. The shape of $\Delta_a$ will depend on the value of $\phi$. When $\hat{\phi}$ gets closer to 1 (which is often the case in monthly data), the longer the maturity, the greater the error. A graphical illustration is given in Figure 6.

Although the effects of estimation errors on different parameters might work on different directions, they can hardly cancel out completely. The total effect will be non-linear on the yield curve coefficients across different maturities, and the medium-to-long maturities will be more affected. When there is more than one factor in the state vector, the situation will become more complicated and involved.

When the estimated parameters are used to compute out-of-sample forecast, any error in $\hat{X}_{t+h}$ will be multiplied by the coefficient error in $b_{n+1}$, with the medium-to-long term yields more sensitive. When the forecast horizon $h$ increases, this will lead to bigger errors cumulated in $\hat{X}_{t+h}$ and also the prediction errors in medium-to-long term yields will become more severe.

Let us go back to the questions that motivated our analysis.

Among the parameters in no-arbitrage models, the risk price parameters $\lambda_0$ and $\lambda_1$ are subject to the researcher’s specification assumptions, which are very likely to be misspecified or have identification problems. The dimensionality curse together with the non-linear effects of estimation errors in risk parameters will potentially make the forecast from no-arbitrage models less accurate than a reduced form model, in particular for yields with longer maturities. In fact, this is what we see from our results in models with three yield factors.

In no-arbitrage models in which most of the information is extracted from macro factors and in which less information is extracted from the yield curve, the lack of high-frequency fluctuations in the states may lead to bad forecasting performance, especially at short forecasting horizons. Kim (2007) discusses these effects in detail. In light of our analysis, the induced mis-specification and/or estimation errors can be transmitted to the long end of the yield curve and negatively affect the out-of-sample forecasts. A comparison between unrestricted and no-arbitrage models in the 5th and 6th row of Tables 2 and 3 makes this point clear. Because of the two step estimation of no-arbitrage models, the VAR coefficients and state variable forecast are the same across the three models in each row, and the only difference comes from differences in the estimated coefficients of the yield equations. Compared with the unrestricted models, the no-arbitrage models have higher prediction errors measured by RMSFE for longer term yields, because the estimation/specification errors in the coefficients for longer term yields are
higher due to the nonlinear effect we discussed above.

5.3 Further issues on model specification and forecast procedure

In this Section, we discuss several issues related to forecasting with the Nelson-Siegel model. Results are reported in tables 4 for the RMSFE ratios and 5 for the BMAIs. We denote the model in row \( m \) and column \( n \) by model \( (m,n) \).

1) State dynamics: VAR(1) versus AR(1). As discussed above, we have chosen the Diebold-Li model with VAR(1) states as a benchmark for the class of unrestricted models (these results are reported in the (1,1) entry of Tables 4 and 5). Diebold-Li (2006) find that having the states follow 3 independent AR(1) further improves the forecasting performance. We compare the AR(1) specification with the VAR(1) and find that the AR(1) underperforms the VAR(1) forecasts for short rate at short horizons, and that there might be some gains at longer horizons for long rates. We provide these results in the rows labeled "AR" of Tables 4 and 5.

2). Iterated versus dynamic forecast. Diebold-Li (2006) obtain \( h \)-steps ahead forecasts by directly projecting \( X_{t+h} \) on \( X_t \):

\[
\hat{X}_{t+h|t} = \hat{\mu} + \hat{\Phi}X_t,
\]

In our benchmark results we have used iterated projections. Marcellino, Stock and Watson (2006) find that dynamic forecast is more robust under model misspecification, while iterated forecast is more efficient otherwise. We compared the two forecasting methods and we find mixed results: the dynamic forecast does not outperform the iterated forecast. Models (1,1) and (1,2) show that the iterated forecast is even slightly better.

3). Calibrated \( \lambda \) versus estimated \( \lambda \). So far, we have used two step OLS estimation to extract the Nelson-Siegel factors with a fixed \( \lambda \), calibrated at the value used by Diebold, Rudebusch and Aruoba (2006). An obvious alternative would be to estimate \( \lambda \). In the third column of Tables 4 and 5, we show the forecast results when we estimate \( \lambda \) with maximum likelihood. Compared with the first column, the forecast performance is similar in BMAI

\[\text{We also considered models in which the Nelson-Siegel factors are augmented with macro factors. There, the AR(1) has a bad forecasting performance. Appendix file for the results are available upon request.}\]
ratios, with long term yield forecast slightly better but short rate forecast
less accurate in terms of RMSFE ratios. We conclude that the estimation of
\( \lambda \) does not matter much for our general results and our conclusions.

4). \( A_0(3) \) versus no-arbitrage Nelson-Siegel. In the introduction we briefly
discussed a modification of the Nelson and Siegel approach that rules out
arbitrage opportunities (see Christensen, Diebold and Rudebusch (2007),
and Le Grand (2007)). We estimate this model and present the results in the
fourth column of Tables 4 and 5. Compared with the Diebold-Li model, the
forecasting performance of the no-arbitrage Nelson-Siegel model is similar,
with a slight improvement towards the longer term yield at long horizons.

In conclusion, our results seem to be robust to the 4 issues analyzed in
this Section.

6 Forecasting macroeconomic variables with
yield curve information: what can we learn
from this framework?

Our framework not only allows to investigate the issue of yield curve forecast,
but provides a laboratory to study whether models of the yield curve are
useful to forecasting macroeconomic variables. Within this framework, we
can examine the following issues: 1) whether the information on the term
structure can contribute to the forecast of macro variables compared to a
simple time series model, say an AR(1); 2) how sensitive are the results to
specification of risk prices in no-arbitrage models; 3) the relative performance
of no-arbitrage models and reduced form models; 4) the relative performance
of Nelson-Siegel yield factors models to models with one observable yield
factor (1-month rate).

Tables 6 and 7 report RMSFE ratios and BMAIs for inflation and IP
growth respectively. Column one contains results from the unrestricted mod-
els with three Nelson-Siegel yield factors. Column two shows forecasts from
unrestricted models with the short rate as the single yield factor. Column
three contains results from no-arbitrage models with 3 latent yield factors
under three specifications of risk prices. Each sub-table presents results from
a model with specified information set \( X \).

1) The RMSFE ratios in Table 6 show that when compared with a fore-
cast from a simple time series AR(1), term structure information does not
contribute to the predictability of inflation, but does increase the forecasta-
bility of real activity such as the IP growth. This is largely consistent with
Ang, Bekaert and Wei (2007) for inflation forecast, and the findings of Ang, Piazzesi and Wei (2006) for GDP forecast. The BMAIs in Table 7 show that the inflation forecast from no-arbitrage models 24 month ahead might contain more useful information than the AR(1) model.

2) Within the no-arbitrage models, the results are very sensitive to the risk price settings. We compare three risk price specifications. Comparing results in the third column, a parsimonious modelling is more favorable in terms of lower RMSFE ratios and higher BMAIs, especially for inflation, and the more so when forecast horizon increases. Fully time varying risk prices are in general worse than the other two cases, except for the forecast of IP growth as in model (3,3) where IP growth alone is augmented to the three yield factors. Although time-varying risk price is likely to capture patterns of time-varying risk premia of yields, it is more important for yield dynamics than for macro variables. On the other hand, the underlying parameter uncertainty increases when \( \lambda_1 \) becomes more flexible and the gain for macro variable forecast, if exists, will also quickly diminish.

3) With the first and third columns we can compare the performance of reduced form models with Nelson-Siegel factors with no-arbitrage latent yield factors models. The results show that the no-arbitrage models have better forecasting performance for inflation, especially for medium to long horizons, as long as the risk price setting is simple. For IP growth, the reduced form models do a better job in RMSFE ratios while the BMAIs are not conclusive.

4) The second column only has one yield factor - the short rate (1-month yield), in the unrestricted models. Compared with the first column, although the 3 yield factor models has significantly more parameters than 1 yield factor model which might affect their forecast accuracy, the results indicate that the information contained in slope and curvature contributes to the forecast performance of IP growth - the measure of real activity. The evidence from BMAI ratios is mixed.

7 Conclusions

We propose a general state-space modelling framework to accommodate a vast number of different yield curve models in the macro finance literature. Within this framework, we systematically examine the relative importance of no-arbitrage restrictions versus large information set in forecasting the yield curve. The way we conduct our experiment and comparison helps us to reveal a number of interesting aspects on yield curve forecasting with these modelling tools.

This paper takes a first step in identifying the nonlinear impact of es-
estimation/specification errors on forecast errors across the yield curve under no-arbitrage restrictions. The forecast of the long term yields are particularly sensitive to estimation/specification errors due to the nonlinearities implicit in the no-arbitrage restrictions. Although models with simple (constant prices of risk) no-arbitrage restrictions help to improve the overall forecast performance of the yield curve, more complex model specifications (with time-varying risk prices or with exogenous variables in the state vector) are associated with higher prediction errors, especially at longer maturities.

For these reasons, a combination of large macro information and no-arbitrage restrictions does not yield better forecasts if compared to an unrestricted model. Our paper, however, finds evidence of important effects of macroeconomic variables on the future yield curve, in line with the findings of Ang and Piazzesi (2003), Diebold, Rudebusch and Aruoba (2006). We find the predictive power of macro information is high at relatively long forecast horizons (1 to 2 years) and for yields with short-to-medium maturities. This effect remains even in models with only macro factors and no yield factors.

Our framework is useful in dealing with the reverse question, namely forecasting macro variables with yield curve information. We find that the yield curve is more useful in forecasting real activity than inflation; we also find that, in addition to the level factor of yield, the slope and curvature factors contribute to the forecast of macro variables.

In future work, we plan to extend our framework taking into account stochastic volatility in the state dynamics, in order to explore its role in modelling and forecasting the yield curve.

8 References


9 Appendix

9.1 Appendix 1. No-Arbitrage Restrictions on Bond Pricing Parameters

1. State variable dynamics.
   Transition equation for $X_t$ follows VAR(1):
   \[ X_t = \mu + \Phi X_{t-1} + v_t, \]
   $v_t$ is i.i.d. $N(0, \Omega)$.

2. Short rate equation.
   \[ r_t = \delta_0 + \delta_1' X_t \]
   $\delta_0$: a scalar.
   $\delta_1$: $K \times 1$ vector.

3. Time-varying prices of risk (associated with the sources of uncertainty $v_t$).
   \[ \Lambda_t = \lambda_0 + \lambda_1 X_t \]
   $\Lambda_t$: $K \times 1$ vector.
   $\lambda_0$: $K \times 1$ vector.
   $\lambda_1$: $K \times K$ matrix.
   If investors are risk-neutral, $\lambda_0 = 0$ and $\lambda_1 = 0$, hence $\Lambda_t = 0$, no risk adjustment. If $\lambda_0 \neq 0$ and $\lambda_1 = 0$, then price of risk is constant.

   No arbitrage opportunity between bonds with different maturities implies that there is a discount factor $m$ linking the price of yield of maturity $n$ this month with the yield of maturity $n - 1$ next month.
   \[ P_t^{(n)} = E_t \left[ m_{t+1} P_{t+1}^{(n-1)} \right] \]
   The stochastic discount factor is related to the short rate and risk perceived by the market,
   \[ m_{t+1} = \exp \left( -r_t - \frac{1}{2} \Lambda_t' \Omega \Lambda_t - \Lambda_t' v_{t+1} \right) \]
No-arbitrage recursive relation can be derived from the above equations as:

\[
P_t^{(n)} = E_t \left[ P_{t+1}^{(n)} \right] = E_t \left[ m_{t+1} m_{t+2} P_{t+2}^{(n-2)} \right]
\]

\[
= E_t \left[ m_{t+1} m_{t+2} \ldots m_{t+n} P^{(0)}_{t+n} \right] = E_t \left[ m_{t+1} m_{t+2} \ldots m_{t+n} \cdot 1 \right]
\]

\[
= E_t \left[ \exp \left( -\sum_{i=0}^{n-1} \left( r_{t+i} + \frac{1}{2} \Lambda_{t+i} \Omega \Lambda_{t+i} + \Lambda_{t+i} v_{t+i+1} \right) \right) \right]
\]

\[
= E_t \left[ \exp \left( A_n + B'_n X_t \right) \right] = E_t \left[ \exp \left( -ny_{t,n} \right) \right]
\]

\[
= E_t^Q \left[ \exp \left( -\sum_{i=0}^{n-1} r_{t+i} \right) \right]
\]

\[E_t^Q\] denotes the expectation under the risk-neutral probability measure, under which the dynamics of the state vector \(X_t\) are characterized by the risk-neutral vector of constants \(\mu^Q\) and by the autoregressive matrix \(\Phi^Q\):

\[
\mu^Q = \mu - \Omega \lambda_0
\]

\[
\Phi^Q = \Phi - \Omega \lambda_1
\]

Affine functions of the state variables for yields are:

\[
p_{t,t+n} = \ln P_t^{(n)} = A_n + B'_n X_t
\]

\[
y_{t,t+n} = a_n + b'_n X_t = \frac{1}{n} \left( A_n + B'_n X_t \right)
\]

where the coefficients follow the difference equations:

\[
A_{n+1} = A_n + B'_n \left( \mu - \Omega \lambda_0 \right) + \frac{1}{2} B'_n \Omega B_n + A_1
\]

\[
B_{n+1} = B'_n \left( \Phi - \Omega \lambda_1 \right) + B'_1
\]

with \(a_1 = \delta_0 = -A_1\) and \(b_1 = \delta_1 = -B_1\).

These can be derived from the pricing kernel equation.

\[
P_t^{(n+1)} = \left[ m_{t+1} P_{t+1}^{(n)} \right] = E_t \left[ -r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t - \Lambda'_t v_{t+1} \right] \exp \left\{ A_n + B'_n X_t \right\} E_t \left[ \exp \left\{ -\Lambda'_t v_{t+1} + B'_n X_{t+1} \right\} \right]
\]

\[
= \exp \left\{ -r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t + A_n \right\} E_t \left[ \exp \left\{ -\Lambda'_t v_{t+1} + B'_n X_{t+1} \right\} \right]
\]

\[
\cdot E_t \left[ \exp \left\{ -\Lambda'_t v_{t+1} + B'_n \left( \mu + \Phi X_t + v_{t+1} \right) \right\} \right]
\]

32
\[\begin{align*}
&= \exp \left\{ -\delta_0 - \delta'_2 X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t + A_n + B'_n (\mu + \Phi X_t) \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_2) X_t - \frac{1}{2} \Lambda' \Omega \Lambda_t \right\} \\
5. \textbf{An alternative presentation for the no-arbitrage coefficients.}
\end{align*}\]

In order to understand intuitively how these restrictions are imposed directly on the coefficients in the yield equation, we can write them in the following affined form.

Given that

\[\begin{align*}
\rho_{t,t+n} &= A_n + B'_n X_t \\
y_{t,t+n} &= a_n + b'_n X_t = \frac{1}{n} (A_n + B'_n X_t)
\end{align*}\]

we can derive

\[\begin{align*}
b_{n+1} &= \frac{1}{(n+1)} \left[ \sum_{i=0}^{n} (\Phi' - \lambda'_1 \Omega)^i \right] b_1 \\
a_{n+1} &= a_1 - \frac{1}{(n+1)} \sum_{i=1}^{n} B^{(i)}
\end{align*}\]

where \(B^{(i)} = B'_i (\mu - \Omega \lambda_0) + \frac{1}{2} B'_i \Omega B_i.\)
9.2 Appendix 2. The likelihood function with Chen-Scott (1993) method

(The likelihood function representation follows closely Ang, Piazzesi (2003).)

In order to be able to extract factors under no-arbitrage restrictions, we employ the method by Chen and Scott (1993). Assume that there are \( K \) factors in the state equation and that among them, \( K_2 \) factors are unobserved. When the number of yields \( N \) exceeds number of unobserved factors, \( K_2 \), following Chen and Scott (1993), we assume that \( K_2 \) yields, \( y_{t}^{NE} \), are observed without measurement errors, and that \( N - K_2 \) yields, \( y_{t}^{E} \), are measured with error \( u_{t}^{m} \). The state vector contains both observed variables \( X_{t}^{o} \) and latent factors \( X_{t}^{u} \), thus \( X_{t} = [X_{t}^{o}, \ X_{t}^{u}] \).

The measurement equation can be written as following:

\[
y_{t} = a + b^{o}X_{t}^{o} + b^{u}X_{t}^{u} + b^{m}u_{t}^{m}
\]

where \( y_{t} = \begin{bmatrix} y_{t}^{NE} \\ y_{t}^{E} \end{bmatrix} \), \( a = \begin{bmatrix} a^{NE} \\ a^{E} \end{bmatrix} \), \( b^{o} = \begin{bmatrix} b^{NE,o} \\ b^{E,o} \end{bmatrix} \), \( b^{u} = \begin{bmatrix} b^{NE,u} \\ b^{E,u} \end{bmatrix} \), and \( b^{m} = \begin{bmatrix} 0_{(K_{2} \times (N-K_{2}))} \end{bmatrix} \).

For a given parameter vector \( \theta = (\mu, \Phi, \Omega, \delta_{o}, \delta_{1}, \lambda_{0}, \lambda_{1}) \), the unobserved factors \( X_{t}^{u} \) will be solved from the yields and the observed variables \( X_{t}^{o} \) as:

\[
X_{t}^{u} = (b^{NE,u})^{-1} [Y_{t}^{NE} - a^{NE} - b^{NE,o}X_{t}^{o}].
\]

Denoting the normal density functions of the state variables \( X_{t}^{i} \) and the error \( u_{t}^{m} \) as \( f_{X} \) and \( f_{u^{m}} \) respectively, the joint likelihood \( \mathcal{L}(\theta) \) of the observed data on zero coupon yields \( Y_{t} \) and the observable factors \( X_{t}^{o} \) is given by:

\[
\mathcal{L}(\theta) = \prod_{t=2}^{T} f_{Y_{t}}(y_{t}, X_{t}^{o} | y_{t-1}, X_{t-1}^{o})
\]

\[
\log(\mathcal{L}(\theta)) = \sum_{t=2}^{T} \log |\det (J^{-1})| + \log f_{X}(X_{t}^{o}, X_{t}^{u} | X_{t-1}^{o}, X_{t-1}^{u}) + \log f_{u^{m}}(u_{t}^{m})
\]

\[
= - (T - 1) \log |\det (J)| - \frac{(T-1)}{2} \log(\det(\Omega))
\]

\[
- \frac{1}{2} \sum_{t=2}^{T} (X_{t} - \mu - \Phi X_{t-1})' \Omega^{-1} (X_{t} - \mu - \Phi X_{t-1})
\]

\[
- \frac{(T-1)}{2} \sum_{i=1}^{N-K_{2}} \log (\sigma_{i}^{2}) - \frac{1}{2} \sum_{t=2}^{T} \sum_{i=1}^{N-K_{2}} (\frac{a_{i}^{m}}{\sigma_{i}^{2}})^{2}
\]

(The constant terms like \( \frac{(T-1)}{2} \log(2\pi) \) are ignored.)

The Jacobian term is:

\[
J = \begin{pmatrix}
I_{K - K_{2}} & 0_{(K - K_{2}) \times K_{2}} & 0_{(K - K_{2}) \times (N - K_{2})} \\
B^{o} & B^{u} & B^{m}
\end{pmatrix}
\]

34
Figure 1: US yield curve (1974:4 – 2003:9)
Figure 2: Macro variables

Figure 3: First four macro factor
Figure 4. Comparison of yield factors extracted from different methods

Notes:
1) No-arbitrage factors shown here are extracted under time-varying risk price ($\lambda_i \neq 0$ and diagonal). They are rescaled to make the comparison. No-arbitrage factors extracted under non time-varying risk price ($\lambda_i = 0$ or $\lambda_i$ full matrix) are very similar. The correlations between the corresponding factors are all 0.999.

2) Empirical proxies are directly obtained by yields as follows: Level = $y(120)$; Slope = $y(1) - y(120)$; Curvature = $2*y(24) - (y(120) + y(3))$.

3) Correlation between these factors:

<table>
<thead>
<tr>
<th></th>
<th>NS1</th>
<th>No-ar1</th>
<th></th>
<th>NS2</th>
<th>No-ar2</th>
<th></th>
<th>NS3</th>
<th>No-ar3</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-ar1</td>
<td>0.973</td>
<td>0.977</td>
<td>No-ar2</td>
<td>0.959</td>
<td>0.703</td>
<td>No-ar3</td>
<td>0.964</td>
<td></td>
</tr>
<tr>
<td>Emp1</td>
<td>0.996</td>
<td>0.831</td>
<td>Emp2</td>
<td>0.800</td>
<td>0.830</td>
<td>Emp3</td>
<td>0.820</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5-1. 1-month ahead forecast comparison

Realised yields  A0(3), trp0  Random Walk  A0(3), trp1  Diebold-Li

Realised yields  4 macro, trp0  4 macro, unre.  4 macro, trp1
Figure 5-2. 12-month ahead forecast comparison
Figure 5-3. 24-month ahead forecast comparison
We illustrate the nonlinear and unequal impacts of parameter estimation errors on yield coefficients across the yield curve with a one factor affine no-arbitrage model.

We assume $a_i = 0$, $b_i = 1$, $\omega^2 = 10^{-6}$. Further, we assume $\omega^2 \lambda_i$ is much smaller than $\phi$ so that we can ignore this term in our calculation of the impact kernel $\Delta_b$, i.e. $\Delta_b \approx \frac{1}{n+1} \sum_{i=1}^{n} \phi^i$. We also know that the kernel $\Delta_a = \frac{1}{n+1} \sum_{i=1}^{n} ib_i$. These two kernels measure how much is the impact of parameter estimation errors on the yield coefficients $a_n$ and $b_n$. The shapes of these kernels are closely related to the autoregressive parameter $\phi$. The closer is $\phi$ to 1, the larger is the impacts of the estimation errors on yields with longer maturities.
Table 1 Factor loadings

Factors are extracted from a panel with 162 macro variables (1974:4-2003:9) after transformation to control for stationarity. The first five factors are shown with the five variables with which they are most highly correlated. The first five factors together explain 72.21% of the total variation in the transformed macro panel. The series used are the same as in Giannone, Reichlin and Sala (2004), except that we exclude 9 interest rate variables.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Total variance explained:</th>
<th>$R^2$</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>32.26%</td>
<td></td>
<td>Index of IP: Non-energy, total 0.93, Index of IP: Mfg 0.93, Index of IP: Non-energy excl CCS 0.92, Index of IP: Total 0.92, Index of IP: Non-energy excl CCS and MVP 0.91</td>
</tr>
<tr>
<td>Factor 2</td>
<td>22.25%</td>
<td></td>
<td>PPI: crude materials 0.80, CPI: housing 0.79, CPI: services 0.79, Loans and Securities @ all commercial banks: commercial and Individual loans (in mil of current $) 0.78, CPI: food and beverages 0.77</td>
</tr>
<tr>
<td>Factor 3</td>
<td>9.04%</td>
<td></td>
<td>Loans and Securities @ all commercial banks: Securities, U.S. govt (in mil of current $) 0.55, Loans and Securities @ all commercial banks: Securities, total (in mil of current $) 0.41, ISM mfg index: Employment 0.39, M1 (in bil of current $) 0.38, Mean duration of unemployment 0.38</td>
</tr>
<tr>
<td>Factor 4</td>
<td>4.69%</td>
<td></td>
<td>Nominal effective exchange rate 0.65, Spot Euro/US 0.63, Spot SZ/US 0.56, Depository institutions reserve: Total (adj for rr changes) 0.35, Spot Japan/US 0.34</td>
</tr>
<tr>
<td>Factor 5</td>
<td>3.97%</td>
<td></td>
<td>M3 (in bil of current $) 0.40, M2 (in bil of current $) 0.31, Employment on nonag payrolls: Financial activities 0.21, Loans and Securities @ all commercial banks: Total (in mil of current $) 0.19, Total merchandise exports (FAS value) (in mil of $) 0.19</td>
</tr>
</tbody>
</table>
Table 2. Forecast comparison with RMSFE ratio
RMSFE of the specific model with respect to RMSFE of Random Walk for \( y(n) \) on forecast horizon \( h \) months ahead.

<table>
<thead>
<tr>
<th>Info.</th>
<th>Unrestricted</th>
<th>No-arbitrage Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Const. risk price</td>
<td>Time-varying risk price, ( \lambda_1 ) diagonal</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3} (Diebold-Li)</td>
<td>X: {la1 la2 la3} A0(3)</td>
</tr>
<tr>
<td></td>
<td>X: {la1 la2 la3} A0(3)</td>
<td>X: {la1 la2 la3} A0(3)</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 infl IPgrowth}</td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
</tr>
<tr>
<td></td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 mf1 mf2}</td>
<td>X: {la1 la2 la3 mf1 mf2}</td>
</tr>
<tr>
<td></td>
<td>X: {la1 la2 la3 mf1 mf2}</td>
<td>X: {la1 la2 la3 mf1 mf2}</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 mf1 mf2 mf3 mf4}</td>
<td>X: {y(1) mf1 mf2 mf3 mf4} (Moench)</td>
</tr>
<tr>
<td></td>
<td>X: {y(1) mf1 mf2 mf3 mf4} (Moench)</td>
<td>X: {y(1) mf1 mf2 mf3 mf4} (Moench)</td>
</tr>
<tr>
<td></td>
<td>X: {mf1 mf2 mf3 mf4}</td>
<td>X: {mf1 mf2 mf3 mf4}</td>
</tr>
<tr>
<td></td>
<td>X: {mf1 mf2 mf3 mf4}</td>
<td>X: {mf1 mf2 mf3 mf4}</td>
</tr>
</tbody>
</table>

**Notes:**
1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12.
2) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively.
3) Illustration on the display of RMSFE ratio:

<table>
<thead>
<tr>
<th>RMSFE ratio</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \leq 0.80 )</td>
<td>( 0.80 - 0.90 )</td>
</tr>
</tbody>
</table>

43
Table 3. Forecast comparison with BMAI ratio

BMAI ratio: posterior probability that the model i’s forecast \( \hat{y}(n) \) is included in a realised yield \( y(n) \)’s regression with respect to that of Random Walk forecast \( \hat{y}(n)_{rw} \): 
\[
y(n) = \beta_0 + \beta_1 \hat{y}(n) + \beta_2 \hat{y}(n)_{rw}. \quad h: \text{forecast horizon.}
\]

<table>
<thead>
<tr>
<th>Info.</th>
<th>Unrestricted</th>
<th>No-arbitrage Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Const. risk price</td>
<td>Time-varying risk price, ( \lambda_i ) diagonal</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 (Diebold-Li)}</td>
<td>X: {la1 la2 la3 (A0(3))}</td>
</tr>
<tr>
<td>Small N</td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
</tr>
<tr>
<td>3 0</td>
<td>y(n) \h 1 6 12 24</td>
<td>y(n) \h 1 6 12 24</td>
</tr>
<tr>
<td></td>
<td>y(1)</td>
<td>1.3 0.8 1.2 1.0</td>
</tr>
<tr>
<td></td>
<td>y(24)</td>
<td>1.0 1.2 1.0 1.0</td>
</tr>
<tr>
<td></td>
<td>y(120)</td>
<td>1.1 1.0 1.0 5.3</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 infl IPgrowth}</td>
<td>X: {la1 la2 la3 infl IPgrowth}</td>
</tr>
<tr>
<td>3 2</td>
<td>y(n) \h 1 6 12 24</td>
<td>y(n) \h 1 6 12 24</td>
</tr>
<tr>
<td></td>
<td>y(1)</td>
<td>1.2 0.7 1.3 1.0</td>
</tr>
<tr>
<td></td>
<td>y(24)</td>
<td>1.0 1.0 1.0 1.0</td>
</tr>
<tr>
<td></td>
<td>y(120)</td>
<td>1.1 1.0 1.0 1.2</td>
</tr>
<tr>
<td></td>
<td>X: {NS1 NS2 NS3 mf1 mf2}</td>
<td>X: {la1 la2 la3 mf1 mf2}</td>
</tr>
<tr>
<td>Large N</td>
<td>X: {la1 la2 la3 mf1 mf2} (Moench)</td>
<td>X: {la1 la2 la3 mf1 mf2} (Moench)</td>
</tr>
<tr>
<td>3 4</td>
<td>y(n) \h 1 6 12 24</td>
<td>y(n) \h 1 6 12 24</td>
</tr>
<tr>
<td></td>
<td>y(1)</td>
<td>1.2 0.7 1.4 8.6</td>
</tr>
<tr>
<td></td>
<td>y(24)</td>
<td>1.0 1.0 4.6 9.6</td>
</tr>
<tr>
<td></td>
<td>y(120)</td>
<td>1.1 1.0 10.4</td>
</tr>
<tr>
<td></td>
<td>X: {y(1) mf1 mf2 mf3 mf4} (Moench)</td>
<td>X: {y(1) mf1 mf2 mf3 mf4} (Moench)</td>
</tr>
<tr>
<td>1 4</td>
<td>y(n) \h 1 6 12 24</td>
<td>y(n) \h 1 6 12 24</td>
</tr>
<tr>
<td></td>
<td>y(1)</td>
<td>1.0 1.0 1.1 10.1</td>
</tr>
<tr>
<td></td>
<td>y(24)</td>
<td>1.0 1.0 16.0 16.0</td>
</tr>
<tr>
<td></td>
<td>y(120)</td>
<td>1.0 1.0 8.3 16.0</td>
</tr>
<tr>
<td></td>
<td>X: {mf1 mf2 mf3 mf4}</td>
<td>X: {mf1 mf2 mf3 mf4}</td>
</tr>
<tr>
<td>0 4</td>
<td>y(n) \h 1 6 12 24</td>
<td>y(n) \h 1 6 12 24</td>
</tr>
<tr>
<td></td>
<td>y(1)</td>
<td>1.0 1.0 1.3 7.7</td>
</tr>
<tr>
<td></td>
<td>y(24)</td>
<td>1.0 1.0 10.7 15.1</td>
</tr>
<tr>
<td></td>
<td>y(120)</td>
<td>1.0 1.4 2.8 1.9</td>
</tr>
</tbody>
</table>

Notes:
1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12.
2) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively.
3) Illustration on the display of BMAI ratio:

<table>
<thead>
<tr>
<th></th>
<th>&lt; 1.00</th>
<th>1.00 - 2.00</th>
<th>2.00 - 5.00</th>
<th>&gt; 5.00</th>
</tr>
</thead>
</table>

44
### Table 4. Forecast comparison of Nelson-Siegel 3 factors dynamic models: RMSFE ratio

<table>
<thead>
<tr>
<th>Model</th>
<th>Diebold-Li</th>
<th>Diebold-Li</th>
<th>Diebold-Li</th>
<th>No-ar. Nelson-Siegel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iterated Forecast</td>
<td>Dynamic Forecast</td>
<td>Iterated Forecast</td>
<td>Iterated Forecast</td>
</tr>
<tr>
<td></td>
<td>OLS, $\lambda$ fixed</td>
<td>OLS, $\lambda$ fixed</td>
<td>MLE, $\lambda$ free</td>
<td>MLE, $\lambda$ free</td>
</tr>
<tr>
<td>$y(n)$</td>
<td>1</td>
<td>6</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>$y(1)$</td>
<td>0.84</td>
<td>0.78</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>$y(24)$</td>
<td>1.03</td>
<td>1.01</td>
<td>0.97</td>
<td>1.05</td>
</tr>
<tr>
<td>$y(120)$</td>
<td>1.07</td>
<td>1.07</td>
<td>1.08</td>
<td>1.40</td>
</tr>
</tbody>
</table>

### Table 5. Forecast comparison of Nelson-Siegel 3 factors dynamic models: BMAI ratio

<table>
<thead>
<tr>
<th>Model</th>
<th>Diebold-Li</th>
<th>Diebold-Li</th>
<th>Diebold-Li</th>
<th>No-ar. Nelson-Siegel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iterated Forecast</td>
<td>Dynamic Forecast</td>
<td>Iterated Forecast</td>
<td>Iterated Forecast</td>
</tr>
<tr>
<td></td>
<td>OLS, $\lambda$ fixed</td>
<td>OLS, $\lambda$ fixed</td>
<td>MLE, $\lambda$ free</td>
<td>MLE, $\lambda$ free</td>
</tr>
<tr>
<td>$y(n)$</td>
<td>1</td>
<td>6</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>$y(1)$</td>
<td>1.05</td>
<td>1.03</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>$y(24)$</td>
<td>1.08</td>
<td>1.06</td>
<td>0.98</td>
<td>1.07</td>
</tr>
<tr>
<td>$y(120)$</td>
<td>1.05</td>
<td>1.07</td>
<td>1.05</td>
<td>1.39</td>
</tr>
</tbody>
</table>

**Notes:**
1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12. 2) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1-2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively. 3) Illustration on the display of RMSFE ratio and BMAI ratio:

**RMSFE ratio:**
- < 0.80
- 0.80 – 0.90
- 0.90 – 1.00
- > 1.00

**BMAI ratio:**
- 1.00
- 1.00 – 2.00
- 2.00 – 5.00
- > 5.00
### Table 6. Forecast comparison for macro variables with RMSFE ratio

RMSFE of the specific model with respect to that of AR(1) for the macro variables on forecast horizon $h$ months ahead.

<table>
<thead>
<tr>
<th>Macro</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 Nelson-Siegel factors</td>
<td>1 yield factor: 1-month rate</td>
</tr>
<tr>
<td>$X$: {NS1 NS2 NS3 infl.}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>1.02</td>
<td>1.04</td>
<td>1.06</td>
</tr>
<tr>
<td>$X$: {la1 la2 la3 infl.}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>trp=0</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>trp=1</td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td>trp=2</td>
<td>1.02</td>
<td>1.09</td>
</tr>
</tbody>
</table>

| Inflation | | | |
|-----------|-----------|-----------|
| $X$: \{NS1 NS2 NS3 infl. IPg\} | $X$: \{m1 infl. IPg\} | $X$: \{la1 la2 la3 infl. IPg\} |
| $h$ | 1 | 6 | 12 | 24 |
| 1.07 | 1.28 | 1.41 | 1.70 |
| 1.03 | 1.14 | 1.33 | 1.29 |
| trp=0 | 1.09 | 1.30 | 1.41 | 1.44 |
| trp=1 | 1.10 | 1.37 | 1.48 | 1.35 |
| trp=2 | 1.04 | 1.20 | 1.41 | 1.46 |

| IP growth | | | |
|-----------|-----------|-----------|
| $X$: \{NS1 NS2 NS3 infl. IPg\} | $X$: \{m1 infl. IPg\} | $X$: \{la1 la2 la3 infl. IPg\} |
| $h$ | 1 | 6 | 12 | 24 |
| 0.97 | 0.90 | 0.83 | 0.96 |
| 0.95 | 0.80 | 0.84 | 1.05 |
| 1.01 | 0.98 | 0.92 | 1.05 |
| trp=0 | 0.98 | 0.95 | 0.95 | 1.33 |
| trp=1 | 0.99 | 0.95 | 0.95 | 1.24 |
| trp=2 | 1.05 | 1.02 | 0.91 | 1.07 |

**Notes:**

1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12.
2) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively.
3) trp = 0: constant risk price; trp = 1: time varying risk price, $\lambda_i$ diagonal; trp = 2: time-varying risk price, $\lambda_i$ full matrix.
4) Illustration of the RMSFE ration:

<table>
<thead>
<tr>
<th>&lt;= 1.00</th>
<th>&gt; 1.00</th>
</tr>
</thead>
</table>

46
### Table 7. Forecast comparison for macro variables with BMAI ratio

BMAI ratio: posterior probability that the model $i$’s forecast $\hat{x}_i$ is included in a realised $x$’s regression with respect to that of a simple AR(1) forecast $\hat{x}_{AR}$: $x = \beta_0 + \beta_1 \hat{x}_i + \beta_2 \hat{x}_{AR}$. $h$: forecast horizon.

<table>
<thead>
<tr>
<th>Macro</th>
<th>Unrestricted</th>
<th>Restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 Nelson-Siegel factors</td>
<td>1 yield factor: 1-month rate</td>
</tr>
<tr>
<td></td>
<td>$X$: ${NS1\ NS2\ NS3\ infl.}$</td>
<td>$X$: ${m1\ infl.\ IPg}$</td>
</tr>
<tr>
<td>$h$</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td><strong>1.6</strong></td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.5</td>
</tr>
</tbody>
</table>

| $X$: $\{NS1\ NS2\ NS3\ infl.\ IPg\}$ | $X$: $\{m1\ infl.\ IPg\}$ | $X$: $\{la1\ la2\ la3\ infl.\ IPg\}$ |
| $h$ | 1 | 6 | 12 | 24 | $h$ | 1 | 6 | 12 | 24 | $h$ | 1 | 6 | 12 | 24 |
|       | 0.8 | 1.0 | **1.0** | 0.2 | $trp=0$ | 0.9 | 1.0 | 0.5 | 0.1 |
|       | **1.0** | **1.0** | **1.0** | 0.1 | $trp=1$ | 0.9 | **1.0** | **1.0** | 0.1 |
|       | **1.0** | **1.0** | **1.0** | 0.1 | $trp=2$ | 0.9 | **1.0** | **1.0** | 0.1 |

| $X$: $\{NS1\ NS2\ NS3\ infl.\ IPg\}$ | $X$: $\{m1\ infl.\ IPg\}$ | $X$: $\{la1\ la2\ la3\ infl.\ IPg\}$ |
| $h$ | 1 | 6 | 12 | 24 | $h$ | 1 | 6 | 12 | 24 | $h$ | 1 | 6 | 12 | 24 |
|       | 0.3 | 0.3 | 0.7 | 0.6 | $trp=0$ | 0.4 | 0.6 | 0.9 | **1.0** |
|       | 0.3 | 0.3 | 0.4 | **1.0** | $trp=1$ | 0.3 | 0.2 | 0.5 | 1.0 |
|       | 0.3 | 0.6 | 0.9 | 1.0 | $trp=2$ | 0.3 | 0.6 | 0.9 | 1.0 |

**Notes:**

1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12.

2) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively.

3) $trp = 0$: constant risk price; $trp = 1$: time varying risk price, $\lambda_i$ diagonal; $trp = 2$: time-varying risk price, $\lambda_i$ full matrix.

4) Illustration of the BMAI ratio:

<table>
<thead>
<tr>
<th>&lt; 1.00</th>
<th>1.00 – 2.00</th>
<th>2.00 – 5.00</th>
<th>&gt; 5.00</th>
</tr>
</thead>
</table>

47