Dynamic Factor Models with Smooth Loadings for Analyzing the Term Structure of Interest Rates

Borus Jungbacker[(a)] *Siem Jan Koopman*[(a,b)] *Michel van der Wel*[(b,c)]

[(a)] Department of Econometrics, VU University Amsterdam

[(b)] Tinbergen Institute

[(c)] Erasmus School of Economics, ERIM Rotterdam and CREATES, Aarhus

Some keywords: Fama-Bliss data set; Kalman filter; Maximum likelihood; Yield curve.
JEL classification: C32, C51, E43.

Acknowledgments

We thank Francis X. Diebold for making the dataset available on his website. More details of our estimation results generated for this paper are available from the authors upon request. Remaining errors are our own. Michel van der Wel acknowledges the support from CREATES, funded by the Danish National Research Foundation.

Address of correspondence: S.J. Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands.
Emails: `bjungbacker@feweb.vu.nl` `s.j.koopman@feweb.vu.nl` `vanderwel@ese.eur.nl`

Dynamic Factor Models with Smooth Loadings for Analyzing the Term Structure of Interest Rates

Borus Jungbacker, Siem Jan Koopman and Michel van der Wel

Abstract

We propose a new approach to the modelling of the term structure of interest rates. We consider the general dynamic factor model and show how to impose smoothness restrictions on the factor loadings. We further present a statistical procedure based on Wald tests that can be used to find a suitable set of such restrictions. We present these developments in the context of term structure models, but they are also applicable in other settings. We perform an empirical study using a data set of unsmoothed Fama-Bliss zero yields for US treasuries of different maturities. The general dynamic factor model with and without smooth loadings is considered in this study together with models that are associated with Nelson-Siegel and arbitrage-free frameworks. These existing models can be regarded as special cases of the dynamic factor model with restrictions on the model parameters. For all model candidates, we consider both stationary and nonstationary autoregressive processes (with different numbers of lags) for the latent factors. Finally, we perform statistical hypothesis tests to verify whether the restrictions imposed by the models are supported by the data. Our main conclusion is that smoothness restrictions can be imposed on the loadings of dynamic factor models for the term structure of US interest rates but that the restrictions implied by a number of popular term structure models are rejected.

1 Introduction

Many time series models for the term structure of interest rates assume that the yield curve for different times to maturity is driven by a small set of unobserved stochastic processes. In this paper we consider specifically the use of dynamic factor models for modelling the time series dynamics of the term structure of interest rates. In a dynamic factor model, the relationships between the yields and the unobserved processes are linear. It is customary in the dynamic factor literature to refer to the unobserved processes as the common factors and to refer to the coefficients that link the factors with the observed time series as the factor loadings. The term structure of interest rates tends to be a smooth function of time to maturity. It is therefore reasonable to assume that the factor loadings are smooth functions of time to maturity as well. Many dynamic factor models for the term structure impose such a smoothness restriction on the factor loadings. The form of the restrictions is often motivated by a no-arbitrage argument. The primary aim of this paper is to find empirical evidence to support the assumption of smooth factor loadings.

To investigate whether smoothness restrictions can be imposed on the factor loadings, we analyze the data set of unsmoothed Fama-Bliss zero yields. As part of the analysis we propose a new type of dynamic factor models: the smooth dynamic factor model. The smooth dynamic factor model allows us to impose smoothness restrictions on the factor loadings in a straightforward manner. We present a statistical procedure based on Wald tests that can be used to find a suitable set of restrictions. Further, we consider a number of popular term structure models that can be seen as dynamic factor models with smoothness restrictions imposed on the factor loadings. Since all these models are nested in the general dynamic factor model, we can test the validity of these restrictions using a likelihood ratio test. The alternative hypothesis of the likelihood ratio test is the unrestricted dynamic factor model as the true data generating process. In the empirical analysis we consider a wide variety of different stationary and nonstationary autoregressive specifications with different numbers of lags.

Our main empirical finding is the high precision level at which the factor loadings of an unrestricted dynamic factor model for the yield curve can be estimated. The high level of precision implies that a restricted model must be sufficiently flexible to match its estimated loadings with those of the unrestricted model. Although the estimated factor loadings for all time series models considered are close to those for the unrestricted model, they are not sufficiently close to be accepted by the likelihood ratio test. Our proposed smooth dynamic factor model is the only exception. Based on the statistical procedure as developed in this paper, we construct a parsimonious time series model that is not rejected by the likelihood ratio test while it imposes considerable smoothness on the factor loadings.

The dynamic factor model plays a central role in this paper. Early contributions to the literature on dynamic factor models can be found in Sargent and Sims (1977), Geweke (1977), Engle and Watson (1981), Watson and Engle (1983), Connor and Korajczyk (1993) and Gregory, Head, and Raynauld (1997). Most of these papers consider time series panels with limited panel dimensions. The increasing availability of high-dimensional data sets has intensified the quest for computationally efficient estimation methods. The strand of literature headed by Forni, Hallin, Lippi, and Reichlin (2000), Stock and Watson (2002) and Bai (2003) led to a renewed interest in dynamic factor analysis. These methods are typically applied to high dimensional panels of time series. Exact maximum likelihood methods such as proposed in Watson and Engle (1983) have traditionally been dismissed as too computationally intensive for such high dimensional panels. Jungbacker and Koopman (2008) however present new results that allow the application of exact maximum likelihood methods to large panels. Examples of recent papers employing likelihood-based methods for the analysis of dynamic factor models are Doz, Giannone, and Reichlin (2006) and Reis and Watson (2007).

The Nelson-Siegel class of factor models for the term structure is based on the seminal paper of Nelson and Siegel (1987) in which a yield curve is approximated by a weighted sum of three smooth functions. The form of these three functions depends on a single parameter. Diebold and Li (2006) use the Nelson-Siegel framework to develop a two-step procedure for the forecasting of future yields. They show that forecasts obtained from this procedure are competitive with forecasts obtained from other standard prediction methods. Diebold, Rudebusch, and Aruoba (2006) integrate the two-step approach into a single dynamic factor model by specifying the Nelson-Siegel weights as an unobserved vector autoregressive process. Generalizations of this state space approach are considered by De Pooter (2007) and Koopman, Mallee, and Van der Wel (2009). The former considers more coefficients for the yield curve while the latter allows the parameter governing the shape of the three Nelson-Siegel functions to vary over time and includes conditional heteroskedasticity in the innovations. A different approach is proposed by Bowsher and Meeks (2008). In their model the term structure is represented as a cubic spline that is observed with measurement noise. The parameters controlling the shape of the spline are modelled by a cointegrated vector autoregressive process. This approach of modelling smooth functions that vary stochastically over time was introduced in Harvey and Koopman (1993).

Many contributions are concerned with the construction of models for the yield curve dynamics that incorporate the restriction that the market is free of arbitrage opportunities, see, for example, Brigo and Mercurio (2006) for an extensive overview. Similarly to the time series models discussed above, the arbitrage-free models are generally specified in terms of a small number of unobserved stochastic processes. However, many of these models imply a nonlinear relation between the unobserved factors and the yields. An exception is the class of affine term structure models presented in Duffie and Kan (1996). The Gaussian specifications contained in this class of models can be shown to be special cases of the dynamic factor model. A closely related model is the arbitrage-free version of the Nelson-Siegel dynamic factor model proposed by Christensen, Diebold, and Rudebusch (2007). We consider both the arbitrage-free version of the Nelson-Siegel model and the Gaussian affine term structure model in our empirical study.

The structure of the paper is as follows. The general dynamic factor model is presented and discussed in section 2. The new methodology to construct dynamic factor models with smooth factor loadings is developed in section 3. Section 4 discusses a selection of existing term structure models that can be regarded as restricted versions of the general dynamic factor model. In section 5 we give a description of the data set and give the results of a preliminary data analysis. Section 6 presents the results of our empirical study. Section 7 concludes and provides suggestions for future research.

2 The dynamic factor model

We consider a monthly time series panel of treasury yields for a set of N different maturities $\tau_1, \dots, \tau_N$. The yield at time t of the treasury with maturity τ_i is denoted by $y_t(\tau_i)$ for $t = 1, \dots, n$. The $N \times 1$ vector of all yields at time t is given by

$$y_t = \begin{bmatrix} y_t(\tau_1) \\ \vdots \\ y_t(\tau_N) \end{bmatrix}, \qquad t = 1, \dots, n.$$

We denote the vector of all observations by $y = (y_1', \dots, y_n')'$.

The general dynamic factor model is given by

$$y_t = \mu_y + \Lambda f_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H), \qquad t = 1, \dots, n, \tag{1}$$

where μ_y is an $N \times 1$ vector of constants, Λ is the $N \times r$ factor loading matrix, f_t is an r-dimensional stochastic process, ε_t is the $N \times 1$ disturbance vector and H is an $N \times N$ variance matrix. Throughout this paper we restrict the variance matrix of the observation disturbances H to be diagonal. This means that covariance between the yields of different maturities is explained solely by the common latent factor f_t. These latent factors are given by

$$f_t = U\alpha_t, \tag{2}$$

where the $r \times p$ matrix U contains appropriate weights that link f_t to a p-dimensional unobserved state vector. This state vector α_t is modelled by the dynamic stochastic process

$$\alpha_{t+1} = \mu_\alpha + T\alpha_t + R\eta_t, \qquad \eta_t \sim N(0, Q), \qquad t = 1, \dots, n, \tag{3}$$

where μ_α is the $p \times 1$ vector of constants, T is the $p \times p$ transition matrix, R is the $p \times q$ selection matrix (typically consisting of ones and zeros), η_t is the $q \times 1$ disturbance vector and Q is an $q \times q$ variance matrix. For the $p \times 1$ initial state vector we assume

$$\alpha_1 \sim N(a_1, P_1), \tag{4}$$

with $p \times 1$ mean vector a_1 and $p \times p$ variance matrix P_1. Generally, we set the mean of the initial state a_1 to zero and choose the initial variance matrix P to be a function of the system matrices. The Gaussian disturbances ε_t and η_s are serially and mutually uncorrelated for $t, s = 1, \ldots, n$ and are assumed independent of α_1. Although dimensions N, p, q and r can be chosen freely we can assume without loss of generality that $r \leq p$ and $p \geq q$. Also, since the motivation of the dynamic factor model is to explain a multivariate time series using a small number of common components, we will generally have $N >> r$. The vectors μ_y and μ_α and the matrices Λ, H, U, T and Q are referred to as system matrices. This general dynamic factor model can be regarded as a specific case of the state space model, see Harvey (1989) and Durbin and Koopman (2001).

The dynamic specification for f_t is general. All vector autoregressive moving average processes can be formulated as (2) and (3) which is known as the companion form; see, for example, Box, Jenkins, and Reinsel (1994). The family of time series processes that can be formulated as (2) and (3) is however much wider and includes a broad range of nonstationary time series processes. In this paper, however, we focus on models where f_t is either a vector autoregression or a cointegrated vector autoregression. We discuss the form that U, T and R take for these two specifications in sections 2.1 and 2.2.

The elements of the system matrices μ_y, Λ, H, μ_α, T and Q will generally contain parameters that need to be estimated from the data. To ensure identification we need to impose two sets of restrictions on respectively the parameters in the means of the yields, determined by μ_y and μ_α, and the parameters in Λ, T and Q that govern the covariance structure.

First, we cannot estimate both vectors μ_y and μ_α without restrictions. Diebold, Rudebusch, and Aruoba (2006) and Bowsher and Meeks (2008), among others, assume that μ_y is zero and proceed by estimating μ_α only. Additional restrictions need to be imposed on

μ_α in case the dynamic process of f_t is nonstationary, see Bowsher and Meeks (2008). In this paper we leave μ_y unrestricted and set μ_α to zero. We choose this more general model because our main concern is inference on the loading matrix Λ and therefore we prefer to avoid additional restrictions on the remaining parameters.

Second, restrictions on Λ are needed because only its column space can be identified uniquely. Several restrictions on Λ can be considered. We choose to set a selection of r rows of Λ equal to the rows of the $r \times r$ identity matrix. In case $r = 3$ and $N > r$, we may have

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \vdots & \vdots & \vdots \\ \lambda_{N,1} & \lambda_{N,2} & \lambda_{N,3} \end{pmatrix}. \quad (5)$$

In this example, we can interpret the elements of f_t as being a set of hypothetical mean-adjusted 'true' yields for the maturities $\tau_1, \ldots, \tau_r$ which are observed at time t subject to measurement noise in ε_t. We do not necessarily have to restrict the first r rows. We can choose to impose the restrictions on each set of r rows of Λ to obtain a dynamic factor model that is observationally equivalent to the model with Λ of the form (5). Since the rows of Λ correspond to fixed maturities we prefer to distribute the rows of the identity matrix evenly over the full range of rows. This allows us to interpret the factors as short-term, medium-term and long-term components. If f_t is a vector autoregression or a cointegrated vector autoregression, as we assume throughout this paper, this choice of restrictions for Λ allows us leave the parameters in T and Q unrestricted.

2.1 Stationary specification

Our stationary dynamic factor model for time series of yields is defined by (1) where the $r \times 1$ vector f_t is modelled by the vector autoregressive process

$$f_{t+1} = \sum_{j=0}^{k-1} \Gamma_j f_{t-j} + \zeta_t, \qquad \zeta_t \sim NID(0, Q_\zeta), \tag{6}$$

with $r \times r$ coefficient matrices Γ_j for $j = 0, \ldots, k-1$ and variance matrix Q_ζ. The dynamic process (6) is commonly known as a VAR(k) model. We will refer to a dynamic factor model with VAR(k) factors as a DFM-VAR(k) model. We denote by $\Gamma(z)$ the characteristic polynomial of the VAR(k) process given by $\Gamma(z) = I - \sum_{j=0}^{k-1} \Gamma_j z^j$. The stationarity of f_t is ensured by imposing the restriction that $|\Gamma(z)| = 0$ has all roots outside the unit circle. The process f_t is straightforwardly written in the form (2) – (3). In case $k = 1$, we have $\alpha_t = f_t$, $U = R = I_r$, $T = \Gamma_0$ and $Q = Q_\zeta$ where I_m is the $m \times m$ identity matrix. In case $k > 1$, we have

$$\alpha_t = \begin{pmatrix} f_t \\ \vdots \\ f_{t-k+1} \end{pmatrix}, \qquad U = R' = \begin{pmatrix} I_r & 0 & \cdots & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} \Gamma_{0:k-2} & \Gamma_{k-1} \\ I_{r(k-1)} & 0 \end{pmatrix}, \tag{7}$$

and $Q = Q_\zeta$ where $\Gamma_{i:j} = (\Gamma_i \ \cdots \ \Gamma_j)$ for $i, j = 0, \ldots, k-1$ and $i < j$. We choose the variance P_1 of the initial state α_1 to be equal to the variance of the invariant distribution of α_t. This implies that P_1 is the solution to the equation $P_1 = TP_1T' + Q$. The mean of the initial state is set to zero.

2.2 Nonstationary specification

For nonstationary dynamic factor models for time series of yields we assume that the factors f_t are generated by a cointegrated vector autoregressive process. In this case the factors are given by the error-correction specification of the VAR(k) process

$$\Delta f_{t+1} = \beta\gamma' f_t + \sum_{j=1}^{k-1} \Gamma_j \Delta f_{t-j} + \zeta_t, \qquad \zeta_t \sim N(0, Q_\zeta), \tag{8}$$

where Δ is the difference operator ($\Delta f_{t+1} = f_{t+1} - f_t$) and $r \times s$ matrices β and γ have full column rank. The remaining matrices are defined as in (6). The matrices β and γ are usually subject to a set of identifying and normalizing restrictions. Let $\Gamma(z)$ denote the characteristic polynomial associated with the process (8). To ensure that the factors are integrated of order one and $\gamma' f_t$ is stationary we need to impose the additional restrictions that all roots of $|\Gamma(z)| = 0$ are outside or on the unit circle and that

$$\det[\beta'_\perp \Gamma(1) \gamma_\perp] \neq 0,$$

where $\beta_\perp$ and $\gamma_\perp$ are $r \times (r - s)$ matrices with their column spaces spanning the null spaces of β' and γ', respectively. A more detailed discussion of error-correction models is given by Johansen (1995). We will refer to (8) as the CVAR(k) model and to the dynamic factor model with CVAR(k) factors as the DFM-CVAR(k) model.

All elements of f_t are nonstationary processes when $s < r$. From a practical perspective it is advantageous to have a specification that decomposes the r factors in f_t into s stationary and $r - s$ nonstationary components. For this purpose we propose an alternative but observationally equivalent specification for f_t via factor rotation. The alternative specification changes the interpretation of the factors but does not alter the dynamic properties of the model. The factors of the new model are given by

$$\bar{f}_t = \begin{pmatrix} \bar{f}_t^N \\ \bar{f}_t^S \end{pmatrix} = \begin{bmatrix} \beta_\perp & \gamma \end{bmatrix}' f_t, \tag{9}$$

where $(r - s) \times 1$ vector $\bar{f}_t^N$ is the nonstationary component and $s \times 1$ vector $\bar{f}_t^S$ is the stationary component. The process $\bar{f}_t$ can be represented by the CVAR(k) model

$$\Delta \bar{f}_{t+1} = A\bar{f}_t + \sum_{j=1}^{k-1} \bar{\Gamma}_j \Delta \bar{f}_{t-j} + \bar{\zeta}_t, \qquad \bar{\zeta}_t \sim N(0, \bar{Q}_\zeta), \tag{10}$$

where the $r \times r$ matrix A is given by

$$A = \begin{pmatrix} 0 & 0 \\ 0 & \bar{\Gamma}_0 \end{pmatrix}, \tag{11}$$

and $\bar{\Gamma}_0, \ldots, \bar{\Gamma}_{k-1}$ and $\bar{Q}_\zeta$ are functions of β, γ, $\Gamma_1, \ldots, \Gamma_{k-1}$ and Q_ζ. To ensure that the model remains observationally equivalent we also need to construct a new loading matrix $\bar{\Lambda}$ by rotating the original matrix Λ into

$$\bar{\Lambda} = \left[\begin{array}{cc} \bar{\Lambda}^N & \bar{\Lambda}^S \end{array} \right], \tag{12}$$

where $N \times (r-s)$ matrix Λ^N and $N \times s$ matrix Λ^S are both of the form (5). We notice that the rotation transfers parameters from the transition equation to the factor loading matrix. The observation equation is given by

$$y_t = \mu_y + \bar{\Lambda}^N \bar{f}_t^N + \bar{\Lambda}^S \bar{f}_t^S + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H), \tag{13}$$

for $t = 1, \ldots, n$. We use this specification to estimate the nonstationary models, but to facilitate comparison with the stationary models we generally present the results for the model with factors given by (8) and loadings given by (5). Note that the maximum likelihood estimators for this second model can be easily obtained from the estimators for the model given by (13). The factors $\bar{f}_t$ can be written in the form (3) by choosing the state vector as follows $\alpha_t = \left(\bar{f}_t' \; \Delta \bar{f}_t' \; \cdots \; \Delta \bar{f}_{t-k+1}'\right)'$ and, for $k > 1$, taking the system matrices $Q = \bar{Q}_\zeta$,

$U = (I_r \; 0 \; \cdots \; 0)$ and

$$T = \begin{pmatrix} I_r + A & \bar{\Gamma}_{1:k-2} & \bar{\Gamma}_{k-1} \\ A & \bar{\Gamma}_{1:k-2} & \bar{\Gamma}_{k-1} \\ 0 & I_{r(k-1)} & 0 \end{pmatrix}, \qquad R = \begin{pmatrix} I_r & I_r & 0 & \cdots & 0 \end{pmatrix}',$$

where $\bar{\Gamma}_{i:j} = \left(\bar{\Gamma}_i \; \cdots \; \bar{\Gamma}_j\right)$ for $i, j = 0, \ldots, k-1$ and $i < j$. The representation for $k = 1$ follows immediately.

If α_t is nonstationary we cannot specify α_1 as in section 2.1. Rosenberg (1973) advocates to consider the nonstationary part of the initial state as an additional set of parameters which can be estimated by maximum likelihood methods. If we choose specification (10) for the factors, only the first $r - s$ elements of α_t are nonstationary. Therefore, if we adopt the approach of Rosenberg (1973) we can set the first $r - s$ rows and columns of the variance of α_1, P_1, to zero. The remaining rows and columns of P_1 are set equal to the variance of the invariant distribution of the stationary elements of α_t. Further, since μ_y is unrestricted, we also set the means of the elements of α_1 corresponding to nonstationary components to zero. We take this approach in the empirical section of this paper. Alternatively, we can choose a reference or diffuse prior for the initial state of the nonstationary components, see the discussion in Durbin and Koopman (2001, Chapter 5). In this case we need to restrict the first $r - s$ elements of μ_y to be zero.

2.3 Parameter estimation and signal extraction

The dynamic factor model consisting of (1), (2) and (3) can be regarded as a special case of the linear state space model. Given the set of system matrices, we can use the Kalman filter and related methods to evaluate minimum mean square linear estimators (MMSLE) of the state vector at time t given the observation sets $\{y_1, \ldots, y_{t-1}\}$ (prediction), $\{y_1, \ldots, y_t\}$ (filtering) and $\{y_1, \ldots, y_n\}$ (smoothing). A detailed treatment of state space methods is given by Durbin and Koopman (2001).

For a given set of system matrices the Kalman filter can also be used to evaluate the loglikelihood function via the prediction error decomposition. The maximum likelihood estimators of the model parameters can then be obtained by numerical optimization. To generate the results in this paper we used the BFGS algorithm to perform the optimization, see for example Nocedal and Wright (1999). An alternative approach would be to use the EM algorithm as developed for state space models by Watson and Engle (1983).

Efficient versions of the Kalman filter have been developed for multivariate models, see for example, Koopman and Durbin (2000). Furthermore, we can achieve considerable computational savings using the methodology of Jungbacker and Koopman (2008). Their method first maps the set of observations y_t into a set of vectors which have the same dimensions as the latent factors f_t in (2). We can then apply the Kalman filter to a typically much lower dimensional set of 'observations'. We have implemented this approach for all models discussed in sections 3 and 4. These efficient Kalman filter methods are also used to evaluate the closed form expressions for the score function given in Koopman and Shephard (1992) and Jungbacker and Koopman (2008). Despite of the large number of parameters involved, this combination of efficient Kalman filter methods and analytical score allows us to estimate all the models considered in this paper in a matter of seconds.

3 Dynamic factor model with smooth factor loadings

The observation equation (1) of the dynamic factor model for the yields $y_t(\tau_i)$ can be written as

$$y_t(\tau_i) = \mu_{y,i} + \sum_{j=1}^{r} \lambda_{ij} f_{jt} + \varepsilon_{it}, \qquad t = 1, \dots T, \qquad i = 1, \dots, N, \tag{14}$$

where λ_{ij} is the factor loading of maturity i and factor j, f_{jt} is the jth element of f_t and ε_{it} is the ith element of ε_t. We propose to specify the model in terms of a set of functions $g_1(\cdot), \dots, g_r(\cdot)$ defined on the interval $[\tau_1, \tau_n]$ and then define the factor loadings as follows

$$\lambda_{ij} = g_j(\tau_i), \qquad i = 1, \dots, N, \qquad j = 1, \dots, r. \tag{15}$$

Since the yield curves tend to be relatively smooth functions of time to maturity and the disturbances ε_{it} are mutually uncorrelated, it is reasonable to assume that the loading functions $g_1(\cdot), \ldots, g_r(\cdot)$ are also smooth functions of time to maturity τ_i. In this section we develop a dynamic factor model that is directly specified in terms of a set of functions $g_1(\cdot), \ldots, g_r(\cdot)$. The model provides means to let the factor loadings be smooth over time to maturity in a straightforward and intuitive manner. We further show how to test for the validity of smoothness restrictions using a series of Wald tests. The resulting model will be referred to as the smooth dynamic factor model (SDFM).

3.1 Model specification

The main assumption of our smooth dynamic factor model is that the loading functions are specified as cubic splines. Specifically, we assume that there is a set of r cubic splines $g_1(\cdot), \ldots, g_r(\cdot)$ defined on $[\tau_1, \tau_n]$ such that $\lambda_{ij} = g_j(\tau_i)$ for $i = 1, \ldots N$ and $j = 1, \ldots, r$. Such cubic splines can capture a wide variety of different shapes. It is therefore reasonable to assume that even if the loading functions of the data generating process are not truly cubic splines, they can still be very closely approximated by functions of this form.

A cubic spline is specified by selecting a set of knots and choosing the function values of the spline at each of these knots. The cubic spline is uniquely defined as the function that is (i) equal to a third-order polynomial between the knots and (ii) twice continuously differentiable at the knots; see, for example, Monahan (2001). It is therefore the location of the knots that determines how the factor loadings behave for varying maturities. In case a small number of knots for a column of Λ is chosen, the corresponding loadings lie on the same cubic polynomial for a considerable number of adjacent maturities. In case we choose the set of knots equal to the set of maturities, no restrictions are imposed on the factor loadings λ_{ij} and the model reduces to the general dynamic factor model of section 2. We can choose a different set of knots for each of the functions $g_1(\cdot), \ldots, g_r(\cdot)$. To ensure a well-defined spline on the interval $[\tau_1, \tau_N]$, the first and last maturities τ_1 and τ_N are taken as knots for all functions $g_1(\cdot), \ldots, g_r(\cdot)$.

In practice, it is convenient to formulate the splines that determine the factor loadings as linear functions of a set parameters which correspond to the unknown values of the splines at their respective knots. Denote the kth knot for the jth column of Λ by s_k^j and suppose that the knots for each column are ordered by time to maturity, that is

$$\tau_1 = s_1^j < \cdots < s_{K_j}^j = \tau_N, \qquad\qquad j = 1, \dots, r,$$

where K_j is the number of knots for the jth column of Λ. Following Poirier (1976), we can specify the loading function $g_j(\tau_i)$ as a linear function, that is

$$g_j(\tau_i) = w_{ij}\delta_j, \qquad \delta_j = \begin{pmatrix} g_j(s_1^j) \\ \vdots \\ g_j(s_{K_j}^j) \end{pmatrix}, \qquad j = 1, \dots, r, \tag{16}$$

where w_{ij} is a $1 \times K_j$ vector that only depends on the location of the knots $s_1^j, \dots, s_{K_j}^j$ and δ_j is treated as a $K_j \times 1$ unknown parameter vector that needs to be estimated. The resulting factor loading matrix Λ of the smooth dynamic factor model is given by

$$\Lambda = \begin{bmatrix} W_1\delta_1 & \cdots & W_r\delta_r \end{bmatrix}, \qquad\qquad W_j = \begin{pmatrix} w_{1j} \\ \vdots \\ w_{Nj} \end{pmatrix}, \tag{17}$$

for $j = 1, \dots, r$. Although the specification (17) of Λ is more parsimonious, we still need to impose restrictions on Λ such as in (5) to ensure that the model is identified.

3.2 Selecting knots via a Wald test procedure

In this section we develop a statistic to test if a subset of knots is significantly contributing to model fit. We use the test statistic to systematically search for a suitable set of restrictions for the loading matrix Λ in the smooth dynamic factor model.

Suppose we have r sets of knots $S_j = \{s_1^j, \dots, s_{K_j}^j\}$ for $j = 1, \dots, r$. We denote the class of all splines associated with the set S_j by G_j. We assume that the sets $S_1, \ldots, S_r$ are sufficiently rich to capture the form of Λ of the data generating process. More formally, if $g_j(\cdot)$ denotes the function that generates the jth column of Λ in the true data generating process, then $g_j(\cdot) \in G_j$ for $j = 1, \ldots, r$. Our aim is to test whether a subset of knots can be removed from a given set S_j. Consider a new set of K_j^* knots denoted by S_j^* such that S_j^* is a subset of S_j. We assume that the set S_j^* is strictly smaller than S_j and therefore $K_j^* < K_j$. Denote the family of splines determined by the knots in S_j^* by G_j^*. It follows that $G_j^* \subset G_j$.

For our purpose, the null-hypothesis H_0 and the alternate hypothesis H_1 are given by

$$H_0 : g_j(\cdot) \in G_j^*, \qquad\qquad H_1 : g_j(\cdot) \notin G_j^*. \tag{18}$$

The null-hypothesis is specifically for the jth spline (or the jth column of Λ) but it can be extended to more general settings and to all r splines jointly. Each spline function in G_j is uniquely determined by the value of δ_j which represents the values of $g_j(\cdot)$ at the knots in S_j. It can therefore be shown that testing the hypotheses of (18) is equivalent to testing linear restrictions on δ_j.

Denote the jth column of Λ by $g_j(\tau) = [g_j(\tau_1), \ldots, g_j(\tau_N)]'$. Then the null-hypothesis can be written as

$$g_j(\tau) = W_j^* \delta_j^*, \tag{19}$$

where W_j^* is the spline weight matrix defined in (17) for set of knots S_j^* and δ_j^* contains the values of the spline at the knots in S_j^*. Since we assumed that $g_j(\cdot)$ is an element of G_j we can also write g_j in terms of W_j, the weight matrix associated with S_j,

$$g_j(\tau) \;=\; \left(\, W_{j\backslash c} \quad W_{j\backslash *} \, \right) \delta_j^\dagger, \qquad\qquad \delta_j^\dagger \;=\; \begin{pmatrix} \delta_j^* \\ \delta_{j\backslash *} \end{pmatrix}, \tag{20}$$

where matrix $W_{j\backslash *}$ consist of columns of the spline weight matrix W_j that correspond to knots that are in S_j but not in S_j^*, matrix $W_{j\backslash c}$ consists of the (remaining) columns in W_j corresponding to knots in S_j^* only and $\delta_{j\backslash *}$ is a vector containing the value of $g_j(\cdot)$ at the knots in S_j that are not in S_j^*. Since a spline is uniquely determined by its value at the knots, the two expressions in (19) and (20) are equivalent if and only if

$$\delta_{j\setminus *} = B_j \delta_j^*,$$

where matrix B_j consists of rows of $W_{j\setminus *}$ corresponding to knots at maturities that are in S_j but not in S_j^*. The hypotheses in (18) reduce to the linear hypotheses

$$H_0 : R_j \delta_j^\dagger = 0, \qquad H_1 : R_j \delta_j^\dagger \neq 0, \qquad R_j = \begin{pmatrix} B_j & -I \end{pmatrix}. \tag{21}$$

Testing linear restrictions of the form (21) is standard in the context of maximum likelihood estimation; see, for example, Engle (1984). For our purposes, a Wald test is particularly convenient. Denote by $\widehat{\delta}_j^\dagger$ the maximum likelihood estimator of $\delta_j^\dagger$ and by $\widehat{V}_j$ a consistent estimator of the asymptotic variance of $\sqrt{nN}\left(\widehat{\delta}_j^\dagger - \delta_j^\dagger\right)$. Under the null-hypothesis we have

$$n \cdot N \cdot \widehat{\delta}_j^\dagger R_j'(R_j \widehat{V}_j R_j')^{-1} R_j \widehat{\delta}_j^\dagger \overset{a.}{\sim} \chi^2\left(K_j - K_j^*\right), \tag{22}$$

where $K_j - K_j^*$ is the number of restrictions imposed under the null-hypothesis. In practice a suitable estimator $\widehat{V}_j$ can be constructed from the Hessian matrix of the log-likelihood function at the maximum likelihood estimator for $\delta_j^\dagger$.

The most important special case of (22) is the situation where $K_j - K_j^* = 1$, meaning that S_j and S_j^* differ by a single knot. We propose to use this test statistic to select the number of knots and their location by means of an iterative general-to-specific approach. At each step we calculate for all the knots in each column a Wald test with the null-hypothesis that the knot is not needed to form the true vector of factor loadings. We then remove the knot that has the smallest non-significant statistic among all the knots used to construct the loading matrix. The procedure is repeated until all selected knots have a statistically significant statistic. We start this iterative testing process with the unrestricted dynamic factor model.

3.3 A more general version

In this section we have focussed on the application of the smooth dynamic factor model to yield curve data. However, this framework has a much wider applicability. We can use this model for any multivariate time series where we observe a smooth function varying over time. Panels of implied volatilities, calculated from call and put contracts on a stock or index with different strikes, are examples of such data sets. These volatility smiles vary over time but tend to be smooth functions of time to maturity. With a slightly more general version of the model discussed in section 3.1 we can also handle a whole different class of problems.

Suppose we have an N dimensional time series $z_1, \ldots, z_n$, where $z_t = (z_{1t}, \ldots, z_{Nt})$ for $t = 1, \ldots, n$ and we model this time series using a dynamic factor model with r underlying latent factors. Even if there is no smooth functional relationship apparent between the elements of $(z_{1t}, \ldots, z_{Nt})$, we might still be able to model the time series very effectively using a SDFM. Suppose that z_t is a very large time series panel containing house prices and let f_{1t} be the factor representing the business cycle. It is likely that houses that are alike have very similar factor loadings for f_{1t}. We can model this by assuming that

$$h_{it} = g_1(p_{it}), \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, n,$$

where h_{it} is the factor loading for house i and factor 1 at time t, p_{it} is a regression variable that indicates the type of house and $g_1(\cdot)$ is a smooth function defined for all values of p_{it}. The variable p_{it} might for example contain the last price at which house i was sold. Just as before we can impose the smoothness restriction on the factor loadings function by assuming that $g_1(\cdot)$ is a cubic spline with a limited set of knots. The general form of this type of SDFM is as follows

$$z_{it} = \mu_{y,i} + \sum_{j=1}^{r} g_j(x_{ijt}) f_{jt} + \varepsilon_{it},$$

where x_{ijt} are regression variables for $i = 1, \ldots, N$, $j = 1, \ldots, r$ and $t = 1, \ldots, n$ and $g_1(\cdot), \ldots, g_r(\cdot)$ are cubic splines. Note that this reduces to the model presented in section 3.1 if we set $x_{ijt} = \tau_i$ for $j = 1, \ldots, r$ and $t = 1, \ldots, n$. This type of model can be especially useful for very large datasets, since it allows us to greatly reduce the number of parameters in the loadings matrix without having to impose the potentially unrealistic assumption that large sets of loadings are equal.

4 Dynamic factor models for the term structure

In this section we review a number of alternative models for the term structure of interest rates that have appeared in the literature. All of these models can be regarded as special cases of the general formulation (1) – (3) with different restrictions imposed on the loading matrix Λ. For some models, restrictions on the dynamics of the factors and the mean vector μ_y are also required. We consider both the stationary specification for f_t as in (6) as well as the nonstationary specification for f_t as in (8).

4.1 Functional Signal Plus Noise Model

The functional signal plus noise (FSN) model is recently proposed by Bowsher and Meeks (2008) as a promising way to model the term structure. The FSN model is also based on cubic splines, just as the model of section 3, but it is used in a different and less flexible way. Consider S_f as a set of r knots and let W_f denote the $N \times r$ spline weight matrix of Poirier (1976). The spline function is then defined by $g_f(\tau) = W_f \delta_f$ where vector δ_f contains the values of the spline function at the knot positions in S_f and is treated as a parameter vector. Instead of using the spline function $g(\cdot)$ to smooth the loadings in each column of Λ, as proposed in the previous section, the spline can also be used to smooth the yield curve directly. In this case, the loading matrix Λ is set equal to the weight matrix W and parameter vector δ_f is replaced by f_t. As a result Bowsher and Meeks (2008) obtain a time-varying cubic spline function for the yield curve. The FSN model is then given by

$$y_t = \mu_y + W f_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, H), \tag{23}$$

where μ_y is the vector of intercepts, f_t is the r-dimensional factor process and H is assumed diagonal. The observed yield curve y_t is now a noisy observation of an unobserved "true" term structure which is modelled by a stochastically time-varying cubic spline function. Finally we notice that by construction the weight matrix W has the same form as Λ in (5). The rows of W that correspond to knots are equal to rows of the identity matrix.

Bowsher and Meeks (2008) consider the CVAR(k) specifications for the unobserved factor f_t with additional restrictions imposed on the cointegration vectors. In this paper we consider both stationary as well as nonstationary specifications for f_t. In case of the nonstationary CVAR(k) specification, we assume that f_t is of the form (10). The decomposition of f_t into stationary and nonstationary components is achieved as in (9) where f_t is transformed to $\bar{f}_t$ which consists of a nonstationary part $\bar{f}_t^N$ and a stationary part $\bar{f}_t^S$. The loading matrix for $\bar{f}_t$ is then given by $\bar{\Lambda} = WL$ where L is the $r \times r$ matrix that transforms the factors $\bar{f}_t$ to the process f_t. This matrix L contains parameters that need to be estimated and is of the same form as $\bar{\Lambda}$ in (12). This decomposition of f_t is useful for interpretation purposes and for the exact handling of the initial state in the implementation of the Kalman filter and related methods.

When certain restrictions are imposed on the smooth dynamic factor model of section 3, it reduces to the FSN model. The key restriction is that all sets of knots S_j for the columns of Λ are set equal and that the number of knots is equal to the number of factors. The restriction that the number of knots equals the number of factors in f_t is strong in practice. For example, Bowsher and Meeks (2008) find that 6 or 7 knots are required to adequately fit the shapes of the term structure typically observed in financial markets. The FSN model therefore requires a vector f_t with at least 6 factors. This number contrasts sharply with empirical studies of, for example, Litterman and Scheinkman (1991) who argue that 3 factors are sufficient to describe the dynamics of the term structure. The SDFM of section 3 has the advantage that the number of factors and the number of knots can be chosen separately and the different sets of knots can be selected more flexibly. Furthermore, a general statistical methodology is provided for the selection of the knots. As a result, we can obtain a better fit using a relatively small number of factors.

4.2 Nelson-Siegel model

In an important contribution Nelson and Siegel (1987) have shown that the term structure can surprisingly well be fitted by a linear combination of three smooth functions. The Nelson-Siegel yield curve, denoted by $g_{ns}(\tau)$, is given by

$$g_{ns}(\tau) = \xi_1 + \lambda^S(\tau) \cdot \xi_2 + \lambda^C(\tau) \cdot \xi_3, \qquad (24)$$

where

$$\lambda^S(\tau) = \frac{1 - e^{-\lambda\tau}}{\lambda\tau}, \qquad \lambda^C(\tau) = \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}, \qquad (25)$$

and where λ, ξ_1, ξ_2 and ξ_3 are treated as parameters. The yield curve depends on these parameters which can be estimated by a least squares method based on the nonlinear regression model

$$y_t(\tau_i) = g_{ns}(\tau_i) + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, n,$$

where u_{it} is noise with zero mean and possibly different variances for different time to maturities τ_i. One of the attractions of the Nelson-Siegel curve is that the ξ parameters have a clear interpretation. The parameter ξ_1 clearly controls the *level* of the yield curve. The parameter ξ_2 can be associated with the *slope* of the yield curve since its loading $\lambda^S(\tau)$ is high for a short maturity τ and low for a long maturity. The loadings $\lambda^C(\tau)$ for different time to maturities τ form an inverted U-shaped function and therefore ξ_3 can be interpreted as the *curvature* parameter of the yield. The decomposition of the yield curve into level, slope and curvature factors has also been highlighted by Litterman and Scheinkman (1991).

The Nelson-Siegel yield curve can also be incorporated in a dynamic factor model by treating the ξ parameters as factors. We obtain

$$y_t = \mu_y + \Lambda_{ns} f_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, H), \qquad (26)$$

where f_t is a 3×1 vector ($r = 3$) and H is a diagonal variance matrix. The loading matrix Λ_{ns} consists of the three columns $(1, \ldots, 1)'$, $\lambda^S(\tau)$ and $\lambda^C(\tau)$ respectively. We note the similarity between the Nelson-Siegel dynamic factor model and the smooth dynamic factor model (14). The slope and curvature loadings in the Nelson-Siegel model both depend on a single parameter λ and this framework is therefore somewhat more restrictive than the SDFM.

The dynamic factor representation of the Nelson-Siegel model is proposed by Diebold, Rudebusch, and Aruoba (2006). Their specification is slightly different as they set μ_y in (26) to zero and include an intercept in the specification of the factors f_t. Furthermore, they specify a stationary vector autoregressive model similar to (6) for the 3-dimensional factor f_t. We will also consider a nonstationary Nelson-Siegel model with f_t of the form (10) and with f_t transformed to $\bar{f}_t$ as in (9). The loading matrix for $\bar{f}_t$ is then given by $\bar{\Lambda} = \Lambda_{ns} L$ where L is the matrix that transforms $\bar{f}_t$ to f_t, see also the discussion of the previous section.

4.3 Arbitrage-free Nelson-Siegel model

Absence of arbitrage opportunities imposes strict restrictions on the stochastic properties of the yield curve; see, for example, the discussion in Cox, Ingersoll, and Ross (1985). The dynamic factor models in this paper so far do not satisfy such restrictions. This is unsatisfactory if we believe that such arbitrage possibilities do not exist in the real world. In this case imposing the no-arbitrage restrictions on the model might improve its performance. This was the motivation for Christensen, Diebold, and Rudebusch (2007) to develop an arbitrage-free version of their Nelson-Siegel dynamic factor model discussed in the previous section.

If the arbitrage-free Nelson-Siegel model is the true underlying data generating process then each $y_t(\tau_i)$ is given by

$$y_t(\tau_i) = \mu_{y,i} + f_{1t} + \lambda_S(\tau_i) f_{2t} + \lambda_C(\tau_i) f_{3t}, \tag{27}$$

where $\mu_{y,i}$ is a correction term that is a deterministic function of the parameters determining the dynamics of the factors, see Christensen, Diebold, and Rudebusch (2007, p. 18) for details. The absence of measurement noise in (27) implies that the corrected yields can be exactly fitted using only f_{1t}, f_{2t} and f_{3t}. Since observed yields never satisfy this restriction in practice it is customary to include measurement errors ε_t in the model. Christensen, Diebold, and Rudebusch (2007) model the factors f_{1t}, f_{2t} and f_{3t} in continuous-time as a multivariate Gaussian process. For evenly spaced observations in discrete time this process can be written as a stationary VAR(1)

$$f_{t+1} = \mu_f^* + \Gamma_0^*(f_t - \mu_f^*) + \zeta_t, \qquad \zeta_t \sim N(0, Q_\zeta^*),$$

where μ_f^* is a 3×1 mean vector, Γ_0^* is the 3×3 autoregressive coefficient matrix and Q_ζ^* is the 3×3 variance matrix. For estimation purposes, it is in practice necessary to formulate the VAR(1) process in terms of the parameters of the original continuous-time process as these parameters appear in $\mu_{y,i}$. We refer the reader to Christensen, Diebold, and Rudebusch (2007) for the functional relationship between these parameters and the VAR(1) matrices. We consider the most general form of the model proposed by Christensen, Diebold, and Rudebusch (2007). This model imposes no restrictions on the intercept μ_f^*, the transition matrix Γ_0^* and the variance matrix Q_ζ^*. Note that this model can be seen as a restricted version of the standard Nelson-Siegel model. Specifically, the AFNS model imposes $N - 3$ restrictions on the intercept μ_y and restricts the factors to be generated by a VAR(1).

4.4 Gaussian Affine Term Structure Model

Let r_t denote the short rate. The short rate can be thought of as the yield of a zero-coupon bond with infinitesimally short time to maturity. For models in the class of affine term structure (AfTS) models, Duffie and Kan (1996) assume that the short rate r_t is an affine function of an unobserved $r \times 1$ dimensional stochastic process f_t

$$r_t = g_1 + g_2' f_t,$$

where g_1 is a scalar parameter and g_2 is a $r \times 1$ vector of parameters. Using a no-arbitrage argument, they proceed to show that if the factors belong to a class of diffusions with affine volatility structure and the market price of risk for each factor is proportional to its volatility, the yields are given by

$$y_t(\tau) = F_1(\tau) + F_2(\tau)r_t, \tag{28}$$

where the functions $F_1(\tau)$ and $F_2(\tau)$ can be obtained from a set of ordinary differential equations, depending on the parameters governing the factor dynamics.

The class of affine term structure models includes a broad range of Gaussian and non-Gaussian specifications. In this paper we focus on the Gaussian case. For the Gaussian specifications it is possible to obtain closed form expressions for $F_1(\tau)$ and $F_2(\tau)$; see equations (3.9) and (3.10) in Duffie and Kan (1996) or equations (9) and (10) in De Jong (2000). In discrete time we can write the factors as a VAR(1) process, after imposing suitable identifying restrictions, see De Jong (2000). This VAR process is of zero mean and has a diagonal transition matrix. Note that this implies that g_1 is the only free parameter in the intercept. Just as for the AFNS model it is unlikely that the observed term structure of interest rates can be fitted exactly by the relation (28). In practice we therefore include a vector of independent Gaussian measurement errors. These measurement errors are allowed to have a different variance for each maturity. The resulting factor model is clearly a restricted version of the DFM model of section 2. For more details on the formulation of the Gaussian AfTS model in dynamic factor form we refer the reader to De Jong (2000).

5 Data description

The empirical study of the next section is based on the same data set considered in Diebold and Li (2006) who constructed a monthly data set of zero yields from the CRSP unsmoothed Fama and Bliss (1987) forward rates. We refer to Diebold and Li (2006) for a detailed discussion of the method that is used for the creation of this data set. We follow Diebold and Li (2006) in considering a subset of the data. Our resulting data set consists of 17 maturities over the period from January 1985 up to December 2000. The maturities we analyze are 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. This dataset has also been considered by Diebold, Rudebusch, and Aruoba (2006), Christensen, Diebold, and Rudebusch (2007) and Bowsher and Meeks (2008), though sometimes for different sample periods and number of maturities.

In Panel A of Figure 1 we present a three-dimensional plot of the data set. The data plot suggests the presence of an underlying factor structure. Although yields vary wildly over time for each of the maturities there is a strong common pattern in the way in which the 17 series develop over time. For most months, the yield curve is an upward sloping function of time to maturity. The overall level of the yield curve is mostly downward trending in our sample period. These findings are supported by the time series plots in Panel B of Figure 1. In these plots we also observe that volatility tends to be lower for the yields of bonds with a longer time to maturity.

Table 1 provides summary statistics for our dataset. For each of the 17 time series we report mean, standard deviation, minimum, maximum and a selection of autocorrelation and partial- autocorrelation coefficients. The summary statistics confirm that the yield curve tends to be upward sloping and that volatility is lower for rates on the long end of the yield curve. In addition, there is a very high persistence in the yields: the first order autocorrelation for all maturities is above 0.95 for each maturity. Even the twelfth autocorrelation coefficient can be as high as 0.57. The partial-autocorrelation function suggests that autoregressive processes of limited lag order will fit the data well since only the first coefficient is significant for most maturities (to preserve space we do not display all coefficients). In Panel B of the Table 1 we present the sample correlations between yields of a selected number of maturities. The correlations are all well above 0.5, in accordance with the strong common pattern in the movements of the different yields that we have observed in Figure 1.

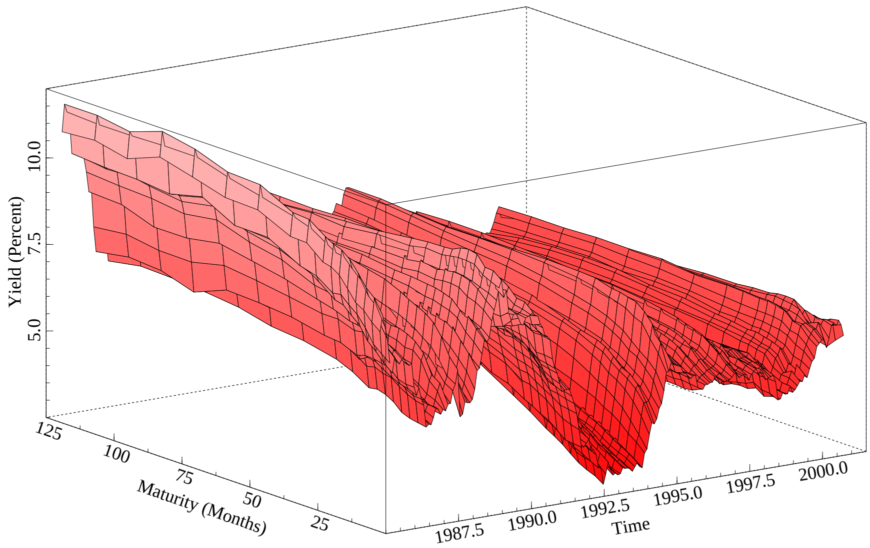
From the data plot in Figure 1, we observe some large breaks in the time series, specifically for the months of May 1985 and October 1987. In both cases, the breaks are apparent in the yield for all maturities and we even observe a drop of more than 1.25% for one of the yields. Also, in both cases the breaks in the yields have lasted for at least a few months.

Finally, Table 1 also provides some information that is relevant for the question whether the yield series are stationary. For this purpose we report Augmented Dickey-Fuller tests for

Figure 1: Yield Curves from January 1985 up to December 2000

In this figure we show the U.S. Treasury yields over the period 1985-2000. We examine monthly data, constructed using the unsmoothed Fama-Bliss method. The maturities we show are 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. Panel A presents a 3-dimensional plot, Panel B provides time-series plots for selected maturities.

(A) 3-Dimensional Term Structure Plot



(B) Time-Series for Selected Maturities

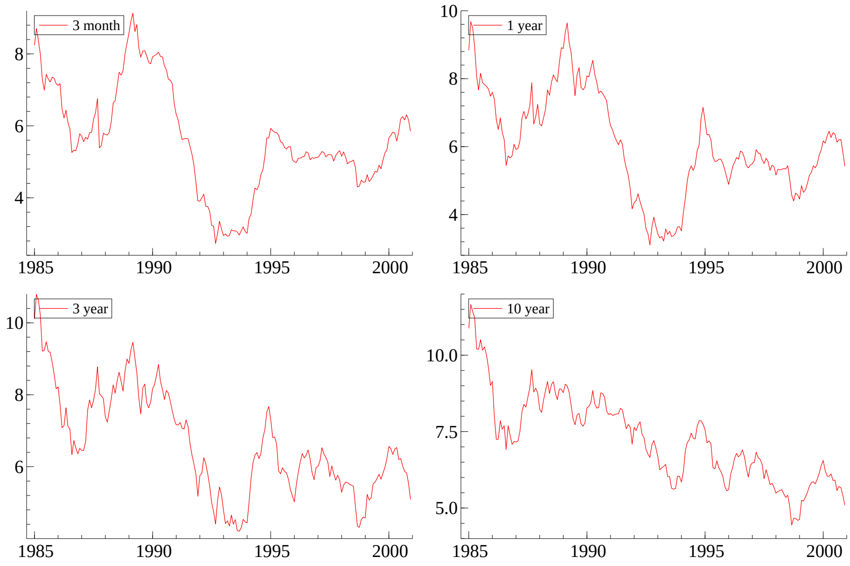


Table 1: Summary Statistics

The table reports summary statistics for U.S. Treasury yields over the period 1985-2000. We examine monthly data, constructed using the unsmoothed Fama-Bliss method. Maturity is measured in months. In Panel A we show for each maturity mean, standard deviation (*Sd*), minimum, maximum and two (1 month and 12 month) autocorrelation (*Acf*, $\hat{\rho}(1)$ and $\hat{\rho}(12)$ respectively) and partial-autocorrelation (*Pacf*, $\hat{\alpha}(1)$ and $\hat{\alpha}(12)$) coefficients. In addition we show the test-statistic and p-value from the Augmented Dickey-Fuller (ADF) unit-root tests. In Panel B we show the correlation matrix for some selected maturities.

Panel A: Summary Statistics

					Acf		Pacf		Unit-root	
Maturity	Mean	Sd	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\alpha}(1)$	$\hat{\alpha}(12)$	ADF	p-value
3	5.63	1.48	2.73	9.13	0.98	0.57	0.98	0.01	-2.06	0.26
6	5.78	1.48	2.89	9.32	0.98	0.55	0.98	0.00	-2.46	0.13
9	5.91	1.49	2.98	9.34	0.97	0.54	0.97	0.01	-2.54	0.11
12	6.07	1.50	3.11	9.68	0.97	0.54	0.97	-0.02	-2.01	0.28
15	6.23	1.50	3.29	9.99	0.97	0.53	0.97	-0.01	-2.75	0.07
18	6.31	1.49	3.48	10.19	0.97	0.51	0.97	0.00	-2.85	0.05
21	6.37	1.48	3.64	10.27	0.96	0.50	0.96	0.00	-2.82	0.06
24	6.40	1.46	3.78	10.41	0.96	0.48	0.96	-0.01	-2.95	0.04
30	6.55	1.46	4.04	10.75	0.96	0.48	0.96	0.03	-3.03	0.03
36	6.64	1.44	4.20	10.79	0.96	0.47	0.96	-0.00	-3.04	0.03
48	6.84	1.44	4.31	11.27	0.95	0.46	0.95	0.00	-2.44	0.13
60	6.93	1.43	4.35	11.31	0.95	0.46	0.95	-0.00	-2.37	0.15
72	7.08	1.45	4.38	11.65	0.95	0.45	0.95	0.01	-2.36	0.16
84	7.14	1.42	4.35	11.84	0.95	0.45	0.95	0.01	-2.53	0.11
96	7.23	1.41	4.43	11.51	0.95	0.47	0.95	0.03	-2.20	0.21
108	7.27	1.42	4.43	11.66	0.95	0.48	0.95	-0.00	-2.27	0.18
120	7.25	1.43	4.44	11.66	0.95	0.47	0.95	0.02	-2.23	0.20

Panel B: Correlation Matrix for Selected Maturities

Maturity	3	12	36	60	120
3	1.00	0.97	0.88	0.80	0.66
12		1.00	0.95	0.88	0.75
36			1.00	0.98	0.91
60				1.00	0.97
120					1.00

each of the series, see Dickey and Fuller (1979). At a 5% significance level we reject the null hypothesis of a unit root for only 3 out of the 17 time series. We reject the null hypothesis for none of the time series if the significance level is lowered to 1%. These findings suggest that a nonstationary dynamic factor model might be a better representation of the yield curve data than a stationary model.

6 Empirical results

In this section we investigate whether the restrictions imposed by the models presented in sections 2 – 4 are supported by the data presented in section 5. For ease of reference we present in Table 2 the most important details of the models discussed thus far. The results of our empirical study are presented as follows. In section 6.1 we review the general assumptions that are applicable to all models. In section 6.2 we discuss the estimation results for the general DFM. In section 6.3 we report the estimation results for the SDFM that is based on a suitable set of smoothing restrictions for the factor loadings as obtained from our Wald test procedure. Section 6.4 discusses the estimation results for the NS and FSN models as well as the arbitrage-free AFNS and AfTS models. In section 6.5 we assess the in-sample fit of the different models by investigating the properties of the residuals. Finally, in section 6.6 we test the validity of the different restrictions by performing a set of likelihood ratio tests. In the remainder of this section we will refer to the models by their acronyms which are listed in Table 2. The dynamic specification for the factors in (6) is referred to as VAR(k) while the nonstationary specification in (10) is referred to as CVAR(k).

6.1 General model assumptions

We start by discussing some general assumptions. To facilitate a fair comparison with the Nelson-Siegel model, we restrict all models to include a total of three latent factors, that is $r = 3$. We can justify this assumption by referring to a growing number of studies that find three factors adequate for explaining most of the variation in the cross-section of yields, see e.g. Litterman and Scheinkman (1991), Bliss (1997) and Diebold and Li (2006). However,

Table 2: Model Summary

This table gives a summary of all the models considered in this paper. We give the acronym used to refer to the model in the text and the section in which the model specification is first discussed. An asterisk in one of the columns with headings μ_y, Λ and T means that respectively the intercept, loading matrix and transition matrix, as defined in section 2, is restricted. An asterisk in the column with heading *VAR(1)* means that not all CVAR(k) and VAR(k) are allowed but only a VAR(1) specification.

Summary of Models and Restrictions						
			Restrictions			
Model	Acronym	Section	μ_y	Λ	T	VAR(1)
Dynamic Factor Model	DFM	2				
Smooth Dynamic Factor Model	SDFM	3.1		*		
Functional Signal plus Noise Model	FSN	4.1		*		
Nelson-Siegel Model	NS	4.2		*		
Arbitrage-Free Nelson-Siegel Model	AFNS	4.3	*	*		*
Gaussian Affine Term Structure Model	AfTS	4.4	*	*	*	*

some other studies have recommended more factors, see the discussion in De Pooter (2007).

For the DFM, SDFM, NS and FSN models the choice of the factor dynamics is arbitrary. To keep the discussion general, we consider VAR(k) as well as CVAR(k) factors for these models. Further, we assume for the CVAR specification that there are two cointegrating vectors for the factors. This means that there is only one random walk present in the cross-section of yields. This is the same assumption as made by Bowsher and Meeks (2008) and is consistent with the findings in Hall, Anderson, and Granger (1992). We will make no assumption on the lag order of the CVAR and VAR processes. Instead, we determine the optimal lag order by minimizing the Akaike Information Criterium (AIC). In this empirical study we will find that the dynamic properties of the factors do not depend on the functional form of the factor loadings.

To account for the large shocks in the yield curve data for the months of May 1985 and October 1987, we include two sets of dummies in each of the models. Since the shocks were persistent and influenced the entire yield curve we included the dummies as intercepts in the unobserved factors. This adds a total of six parameters to each of the model specifications.

6.2 Estimation results for the DFM

In this section we discuss results obtained from the maximum likelihood estimation of the general dynamic factor model. In Table 3 we give the values of the maximized loglikelihood functions for the VAR(k) and CVAR(k) factor specifications together with the corresponding Akaike Information Criterium (AIC) values for $k = 1, 2, 3, 4$. The maxima of the loglikelihood functions for the models with $k = 2$ are considerably higher than the corresponding values for $k = 1$. The improvements from additional lags ($k = 3, 4$) are however much smaller. The VAR(2) and CVAR(2) factor specifications give the smallest AICs values. We will therefore restrict ourselves to these two specifications in the remainder of the section.

Since we are mainly interested in the restrictions imposed on the factor loadings, it is of interest to investigate how the estimated factor loadings change with different choices of k. In Panel A of Figure 2 we plot the estimated factor loadings for the DFM model

Table 3: Likelihoods and AICs for DFM

This table presents maximum likelihood estimation results for the dynamic factor model (DFM) with VAR factors and the DFM with CVAR factors. The models were estimated on the dataset discussed in section 5. We show the value of the loglikelihood evaluated at the maximum likelihood estimates, denoted by $\ell(\widehat{\psi})$, and the value of the Akaike Information Criterium (AIC).

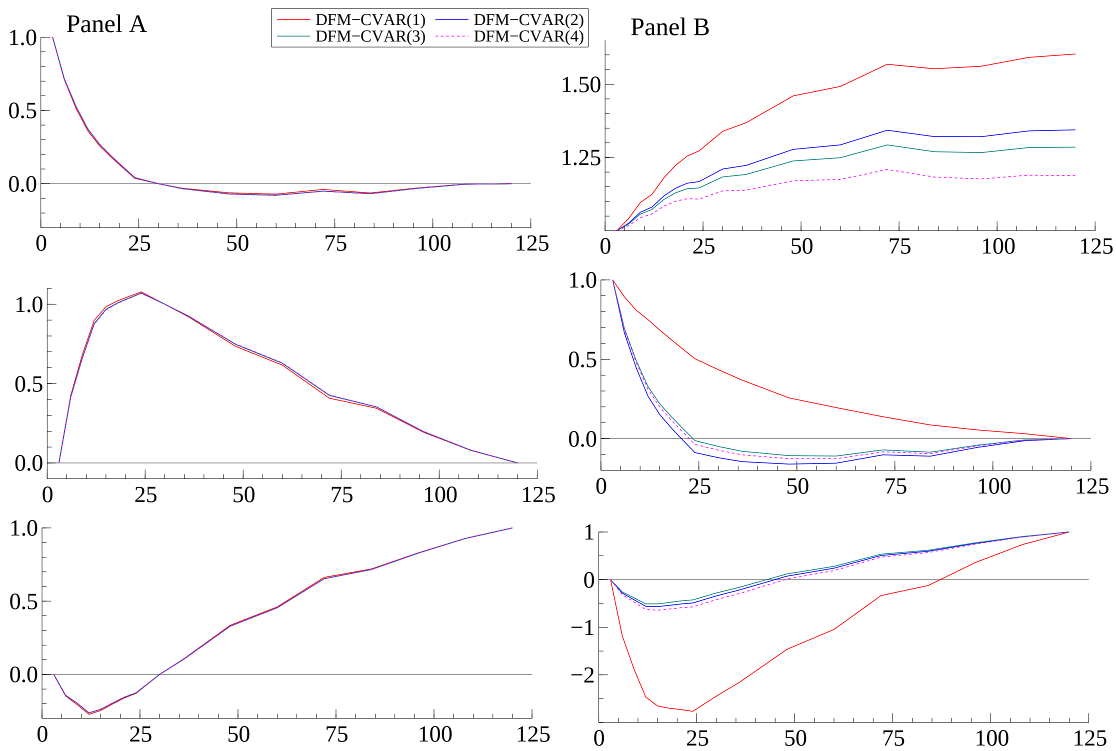
Likelihoods and AICs for DFM					
	VAR			CVAR	
Model	$\ell(\widehat{\psi})$	AIC	Model	$\ell(\widehat{\psi})$	AIC
VAR(1)	3894.5	-7595	CVAR(1)	3899.0	-7606
VAR(2)	3918.5	-7625	CVAR(2)	3923.7	-7637
VAR(3)	3922.6	-7615	CVAR(3)	3927.7	-7627
VAR(4)	3932.2	-7616	CVAR(4)	3937.3	-7628

with CVAR(k) factor specifications, for $k = 1, \ldots, 4$, as functions of time to maturity. The loading matrix Λ is restricted to be of the form (5), with the rows of the identity matrix placed at the rows associated with the maturities of 3, 30 and 120 months. The estimated loadings are nearly identical for different values of k. We have found similar results for the stationary DFM models. We therefore may conclude that the increase of the loglikelihood due to adding extra lags in the VAR and CVAR models for the latent factors is mainly due to a better fit of the time series dynamics. The fit of the yield curve is not affected by different lag orders.

To clarify the effect of an increasing k on the dynamics, we also present the factor loadings for the dynamic factor model with factor specification (10). Here the two stationary factors are separated from the random walk component. The resulting loadings are presented in Panel B of Figure 2. We now obtain clear differences between the factor loadings when the lag order k changes. The loadings for the first (nonstationary) component shifts downward as k increases. It implies that yield variation that is explained in the CVAR(1) model by a nonstationary factor is captured by a highly persistent but stationary component in the CVAR(k) model with $k > 1$. We interpret such shifts as additional evidence that a CVAR(1) model cannot adequately capture the observed yield curve dynamics.

Figure 2: Maximum Likelihood Estimates Loadings for DFM

This figure shows the estimated factor loadings for the DFM-CVAR(k) model, for $k = 1, \dots, 4$ as a function of time to maturity. In panel A we see the results for the model where the factors are modelled by (8). The loadings are now restricted to be of the form (5) with the rows of the identity matrix at the maturities 3 months, 30 months and 120 months. Panel B shows the estimated loadings for the same model but with f_t modelled as (10). In this case the first column of the loading matrix corresponds to the nonstationary factor and is scaled such that the first element is one. The sub-matrix consisting of the second and third columns is now of the form (5).



The autoregressive coefficient matrices Γ_j for VAR(k) and CVAR(k) processes for $j = 0, \ldots, k-1$ are generally difficult to interpret especially when $k > 1$. We therefore choose to report eigenvalues of the estimated transition matrix T of the state space representation (3). In Table 4 the eigenvalues of matrix T for the DFM-VAR(2) and DFM-CVAR(2) are presented. For both models we have two eigenvalues close to one, or equal to one in the nonstationary case, with no imaginary part and two sets of eigenvalues that do have an imaginary component. We can therefore view the factors as a weighted sum of two highly persistent autoregressive (AR) processes (or one AR process and a random walk) and two cyclical components. The presence of two highly persistent factors in the estimated model is in line with our preliminary findings in section 5. Since the highest eigenvalue for the VAR(2) process, 0.992, is very close to one, this process is in practice almost a CVAR process. This explains why the remaining eigenvalues of both models are close to each other as well. Throughout this empirical section we will see that all the VAR specifications are very close to being nonstationary. In practice this means that estimation results for stationary and nonstationary models tend to be very similar.

6.3 Imposing the SDFM restrictions

In the previous section we have concluded that the VAR(2) and CVAR(2) factor specifications are best in representing the yield curve data. Next we apply the methodology of section 3.2 to find a suitable set of smoothness restrictions for the factor loadings of the DFM model with VAR and CVAR specifications.

To ensure that the SDFM specification is identified we need to impose restrictions on the knots and associated parameters. We choose to restrict the loading matrix to be of the form (5) where the rows of the identity matrix are placed in the rows corresponding to the 3, 30, and 120 months of maturities. Since our interpolating cubic spline framework requires knot positions at the begin- and end-points (3 and 120 months), it implies that the knot at 30 months cannot be removed in the course of the selection procedure. Of course, the procedure can be repeated when the knot at 30 months is moved to another time to maturity. After

Table 4: Eigenvalues Estimated Transition Matrices

In these two tables we present the eigenvalues of the estimated transition matrices for the DFM, SDFM, NS and FSN models. In Panel A we show results for the stationary VAR(2) specifications and in Panel B for the models with nonstationary CVAR(2) factors. The columns with heading 'real' contain the real part of the eigenvalues and the columns with heading 'img.' contain the imaginary parts. Eigenvalues are sorted by their norm in ascending order.

Panel A: Stationary models								
	DFM		SDFM		NS		FSN	
	real	img.	real	img.	real	img.	real	img.
1	0.154	0.163	0.164	0.159	0.156	0.166	0.216	0.143
2	0.154	-0.163	0.164	-0.159	0.156	-0.166	0.216	-0.143
3	0.597	0.058	0.607	0.134	0.593	0.056	0.642	0.259
4	0.597	-0.058	0.607	-0.134	0.593	-0.056	0.642	-0.259
5	0.963	-	0.965	-	0.964	-	0.969	-
6	0.992	-	0.992	-	0.992	-	0.993	-

Panel B: Nonstationary models								
	DFM		SDFM		NS		FSN	
	real	img.	real	img.	real	img.	real	img.
1	0.149	0.162	0.155	0.162	0.151	0.165	0.206	0.143
2	0.149	-0.162	0.155	-0.162	0.151	-0.165	0.206	-0.143
3	0.597	0.092	0.601	0.123	0.594	0.099	0.649	0.258
4	0.597	-0.092	0.601	-0.123	0.594	-0.099	0.649	-0.258
5	0.972	-	0.973	-	0.972	-	0.970	-
6	1	-	1	-	1	-	1	-

Table 5: Wald-Statistics of Knots in Unrestricted SDFM specification

This table shows Wald-statistics for the knots in the unrestricted SDFM-CVAR(2) model. The symbol − indicates that for this knot no Wald-statistic was calculated. This is the case for the restricted knots corresponding to 3 months, 30 months and 120 months maturity. We use $*$ respectively $**$ to indicate that a statistic is significant at the 5% and 1% significance level.

SDFM-CVAR(2): Wald-Statistics			
Maturity	Factor 1	Factor 2	Factor 3
3	-	-	-
6	2.65	4.22*	6.08*
9	0.79	2.40	5.59*
12	0.23	1.35	4.28*
15	0.04	0.33	1.51
18	0.01	0.02	0.28
21	0.95	0.74	1.52
24	3.50	2.37	3.98*
30	-	-	-
36	1.14	1.50	6.68**
48	0.44	2.88	13.47**
60	1.20	5.00*	18.04**
72	2.59	5.76*	15.69**
84	2.60	4.59*	8.82**
96	0.77	1.68	1.79
108	0.01	0.06	0.00
120	-	-	-

some experimentation, we have concluded that our main results are not sensitive to moving this knot to maturities in the neighborhood of 30 months.

In Table 5 we present the Wald test-statistics for each knot in the unrestricted model with a CVAR(2) specification for the factors. We only give results for the CVAR(2) specification as the statistics for the VAR(2) model are almost identical for reasons given at the end of section 6.2. At the start of the procedure 12 out of 42 loading coefficients (or knots) are significant at the 5% significance level. This suggests that the number of parameters can be reduced enormously without affecting the fit. However, the test statistics are highly correlated and removing one knot will generally change the statistics of the neighbouring knots considerably. We then proceed by sequentially removing the knot with the lowest

Table 6: Wald-Statistics of Knots in Final SDFM Specifications

This table shows Wald-statistics for the knots in the final SDFM-VAR(2) and SDFM-CVAR(2) models obtained using the iterative procedure discussed in section 3.2. The symbol $-$ indicates that for this knot no Wald-statistic was calculated. This is the case for knots that have been removed and for the restricted knots corresponding to 3 months, 30 months and 120 months maturity. We add a superscript $*$ if the statistic is significant at the 5% significance level and $**$ if the statistic is significant at the 1% level.

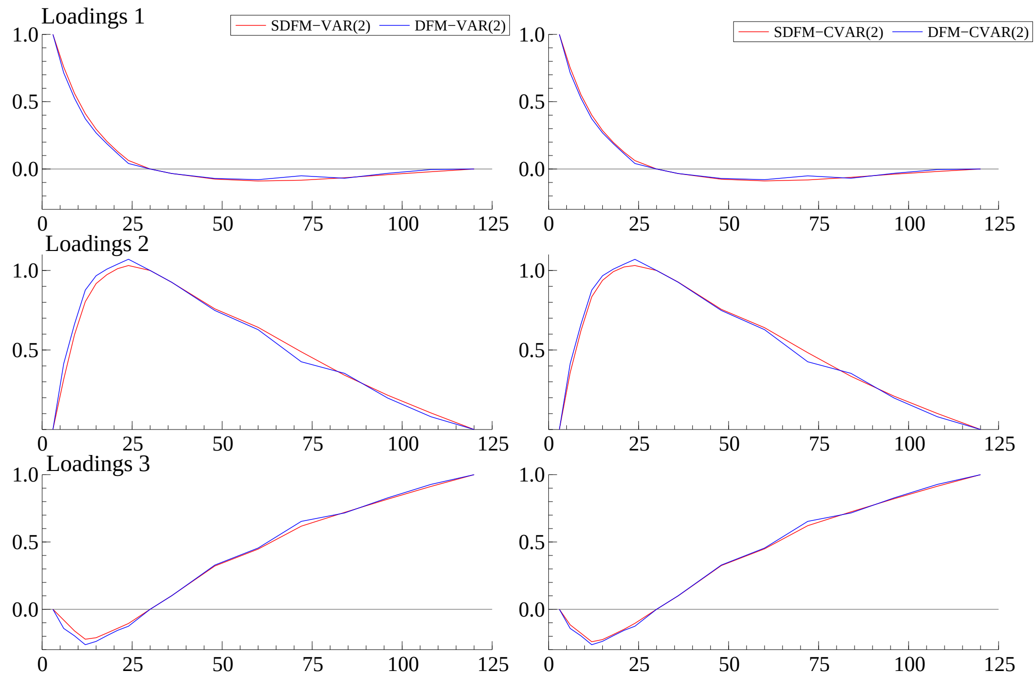
Wald-Statistics of Final Models						
	SDFM-VAR(2)			SDFM-CVAR(2)		
Maturity	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
3	-	-	-	-	-	-
6	55.36^{**}	-	-	59.08^{**}	6.50^{*}	5.24^{*}
9	-	-	17.39^{**}	-	6.58^{*}	8.92^{**}
12	-	53.72^{**}	20.60^{**}	-	16.25^{**}	19.62^{**}
15	-	-	10.20^{**}	-	24.17^{**}	26.83^{**}
18	-	15.86^{**}	-	-	-	-
21	14.41^{**}	-	5.05^{*}	18.55^{**}	-	-
24	16.70^{**}	4.14^{*}	7.55^{**}	23.13^{**}	-	7.35^{**}
30	-	-	-	-	-	-
36	-	-	25.18^{**}	-	-	26.88^{**}
48	-	19.98^{**}	45.68^{**}	-	30.07^{**}	52.87^{**}
60	-	19.00^{**}	47.53^{**}	-	26.79^{**}	54.39^{**}
72	-	15.67^{**}	40.38^{**}	-	22.80^{**}	43.00^{**}
84	-	-	18.04^{**}	-	-	17.85^{**}
96	4.49^{**}	-	5.39^{*}	7.68^{**}	-	5.10^{*}
108	-	-	-	-	-	-
120	-	-	-	-	-	-

Wald-statistic and re-estimating the model after each step. The procedure is terminated when all statistics are significant at the 5% significance level.

In Table 6 we present for both the SDFM-VAR(2) and SDFM-CVAR(2) models the Wald-statistics for the knots that have remained after the final step. The final knot selections for the stationary and nonstationary models are different. However, the distribution of the knots over the interval $[\tau_1, \tau_N]$ is similar for both models. To let a cubic spline fit a certain shape, the distribution of the knots is generally more important than the exact location of the knots. Furthermore, we find that the procedure is especially successful in fitting the first column

Figure 3: Estimated Factor Loadings for SDFM Model

This figure shows the estimated factor loadings for the SDFM-VAR(2) and SDFM-CVAR(2) models, obtained using the procedure of section 3.2, as functions of time to maturity. For ease of comparison we also show the maximum likelihood estimates of the loadings in the DFM model. The loadings are restricted to be of the form (5) with the rows of the identity matrix at the 3 months, 30 months and 120 months maturities.



of factor loadings. The original set of 14 loading parameters is reduced to four remaining knot parameters. In total we reduced the number of parameters in loading matrix Λ by 20 for the nonstationary and 21 for the stationary model, a reduction of, say, 50 percent. In Figure 3 we show the spline estimates for the factor loadings of the final smooth dynamic factor models. For both SDFMs the factor loadings are smooth and close to the estimated loadings for the general DFM model. We have shown that we can achieve almost identical loadings using a much smaller set of parameters. It confirms our prior believe that the true factor loadings are subject to smoothness restrictions. The results for the CVAR and VAR specifications are almost identical.

The construction of the two SDFM specifications in this section required us to estimate respectively 20 and 21 different dynamic factor models each containing around 100 parameters.

This may appear computationally intensive from the outset. However, the computationally efficient methods, discussed in section 2.3, make such a procedure computationally feasible even for larger models.

6.4 Estimation results for other term structure models

In this section we present estimation results for the term structure models discussed in section 4. In Table 7 we report loglikelihood and corresponding AIC values for the NS and FSN models with both VAR(k) and CVAR(k) specifications for $k = 1, \ldots, 4$. To generate these results we first need to specify a set of knots to construct the loading matrix W for the FSN model. The location of the knots are selected using the same methodology as adopted in Bowsher and Meeks (2008). For each possible set of three knots, we fitted a spline through all observed yield curves. We then chose the knot configuration that produced the lowest average mean square error. In our setting we only need to choose one knot, since the other two knots are fixed at the first and last maturities. The results of Table 7 are consistent with the results for the general dynamic factor model reported in section 6.2. For both models, the AIC criteria favour both the VAR and CVAR specifications with $k = 2$. This is also in accordance with the results of Bowsher and Meeks (2008) who, working with a similar data set for a larger number of maturities, find that the FSN model with CVAR(2) factors performs best in their class of models.

In Table 8 we report the maxima of the loglikelihood functions and their corresponding AIC values for the AFNS and AfTS models. These loglikelihood values can only be compared with their corresponding values in Table 3 and Table 7 for the stationary VAR specifications of the factors since AFNS and AfTS models are defined as strictly stationary models. The maximized loglikelihood values for the arbitrage-free models are considerably smaller when compared to the dynamic factor model, with and without smooth factor loadings, and compared to the Nelson-Siegel model. The difference with the FSN model is however less pronounced, especially for the AfTS.

In Table 4 we present the eigenvalues of the estimated transition matrices for the NS and

Table 7: Likelihoods and AICs for Term Structure Models

This table presents maximum likelihood estimation results for the Nelson-Siegel Model (NS) and the Functional Signal plus Noise (FSN) model. All models were estimated on the data set discussed in section 5. We show the value of the loglikelihood evaluated at the maximum likelihood estimates, denoted by $\ell(\widehat{\psi})$, and the value of the Akaike Information Criterium (AIC).

Panel A: NS

	VAR			CVAR	
Model	$\ell(\widehat{\psi})$	AIC	Model	$\ell(\widehat{\psi})$	AIC
VAR(1)	3784.0	-7456	CVAR(1)	3788.7	-7467
VAR(2)	3808.4	-7487	CVAR(2)	3813.5	-7499
VAR(3)	3812.5	-7477	CVAR(3)	3817.5	-7489
VAR(4)	3822.2	-7478	CVAR(4)	3827.3	-7491

Panel B: FSN

	VAR			CVAR	
Model	$\ell(\widehat{\psi})$	AIC	Model	$\ell(\widehat{\psi})$	AIC
VAR(1)	3446.9	-6784	CVAR(1)	3452.0	-6796
VAR(2)	3479.0	-6830	CVAR(2)	3483.7	-6841
VAR(3)	3483.4	-6821	CVAR(3)	3488.1	-6832
VAR(4)	3494.8	-6826	CVAR(4)	3499.6	-6837

Table 8: Arbitrage-Free Term Structure Models: Maximum Likelihood Results
This table presents the maximum likelihood estimation results for the two arbitrage-free term structure models: the AFNS and AfTS models. We show the maximum of the loglikelihood function in the column with heading $\ell(\widehat{\psi})$. Further we give the number of parameters in the model n_ψ and the Akaike information criterium (AIC). Finally we present the eigenvalues of the estimated transition matrix T. The eigenvalues are presented in descending order.

Arbitrage-Free Term Structure Models						
				Eigenvalues T		
Model	$\ell(\widehat{\psi})$	n_ψ	AIC	1	2	3
AFNS	3253.3	42	-6423	0.986	0.952	0.884
AfTS	3429.4	36	-6786.9	0.9997	0.998	0.984

FSN models. The nonstationary and stationary specifications for the factor produce similar results. This finding is consistent with the results of the DFM and SDFM models. The eigenvalues for the DFM, NS and FSN specifications are almost identical. Table 8 presents the eigenvalues for the two arbitrage-free models. In both cases the estimated parameters imply a high level of persistence in the factor dynamics.

Next we investigate whether the choice of the lag order in the autoregressive factor dynamics influences the estimates of the factor loadings. This matter only applies to the Nelson-Siegel model. The factor loadings of the FSN model only depend on the selection of the knots and are therefore by definition independent of the dynamics of the factors. The two arbitrage-free models are only defined for a VAR(1) process and therefore we do not consider this issue for these models. For the Nelson-Siegel model, the factor loadings depend on a single parameter λ. In Table 9 we report the maximum likelihood estimates of this parameter λ for different lag orders in the dynamic process of f_t. We find that the estimates of λ are almost identical for stationary and nonstationary factor specifications. Also the estimates of λ vary little for different choices of the lag order. We conclude that maximum likelihood estimation of the factor loadings (here all functions of λ) is not influenced by the dynamic specification of the factors.

In Figure 4 we present the estimated factor loadings for the NS and FSN models with

Table 9: Maximum Likelihood Estimates of Nelson-Siegel Parameter λ
This table presents maximum likelihood estimates of the Nelson-Siegel parameter λ, defined in (24), for models where the factors are given by VAR(p) and CVAR(p) processes and for varying values of p.

Nelson-Siegel Factor Model Parameter λ				
Model	$p = 1$	$p = 2$	$p = 3$	$p = 4$
VAR(p)	0.07303	0.07211	0.07216	0.07193
CVAR(p)	0.07302	0.07210	0.07213	0.07191

CVAR(2) factors. We have rotated the loadings such that the loading matrix Λ is of the form (5). The rotation facilitates easier comparison with the estimated loadings of the general DFM. It is clear that the rotated Nelson-Siegel loadings are similar to the smoothed versions of the DFM loadings. This finding is surprising given that the loadings in the Nelson-Siegel model depend on a single parameter λ while the dynamic factor model relies on 42 factor loadings. For the FSN model, the factor loadings have generally the same shape and form of the loadings obtained from the DFM and NS models. However, the individual factor loadings are quite different.

In Figure 5 we display the factor loadings for the arbitrage-free models. As the loadings of the AFNS are similar to the Nelson-Siegel model we show these of the AfTS model. These loadings have been rotated in the same way so they have the same form as the DFM-VAR(1) loadings. As in the case of the Nelson-Siegel model, the factor loadings in the AfTS are close to the unrestricted estimates. It is revealing to see how close the AfTS loadings are compared to the DFM loadings while the eigenvalues of its transition matrix (its VAR(1) coefficient matrix) presented in Table 8 are quite different from those of the DFM, SDFM, NS and FSN models (based on VAR(2) and CVAR(2) coefficient matrices) reported in Table 4. Also, the factor loadings are estimated simultaneously with the parameters that govern the factor dynamics. These findings suggest that the penalty on an incorrect yield curve specification (determined by the loadings) is much larger than the penalty on an incorrect dynamic specification of the factors. During the search for the optimum of the likelihood function,

Figure 4: Estimated Factor Loadings Nonstationary Models

This figure shows the estimated factor loadings as functions of time to maturity for the DFM-CVAR(2), NS-CVAR(2) and FSN-CVAR(2) models. The factor loadings are rotated such that loading matrix is of the form (5), where the rows of the identity matrix are at the maturities of 3 months, 30 months and 120 months. In panel A we show the loadings for the DFM and NS model and in panel B the loadings for the FSN model.

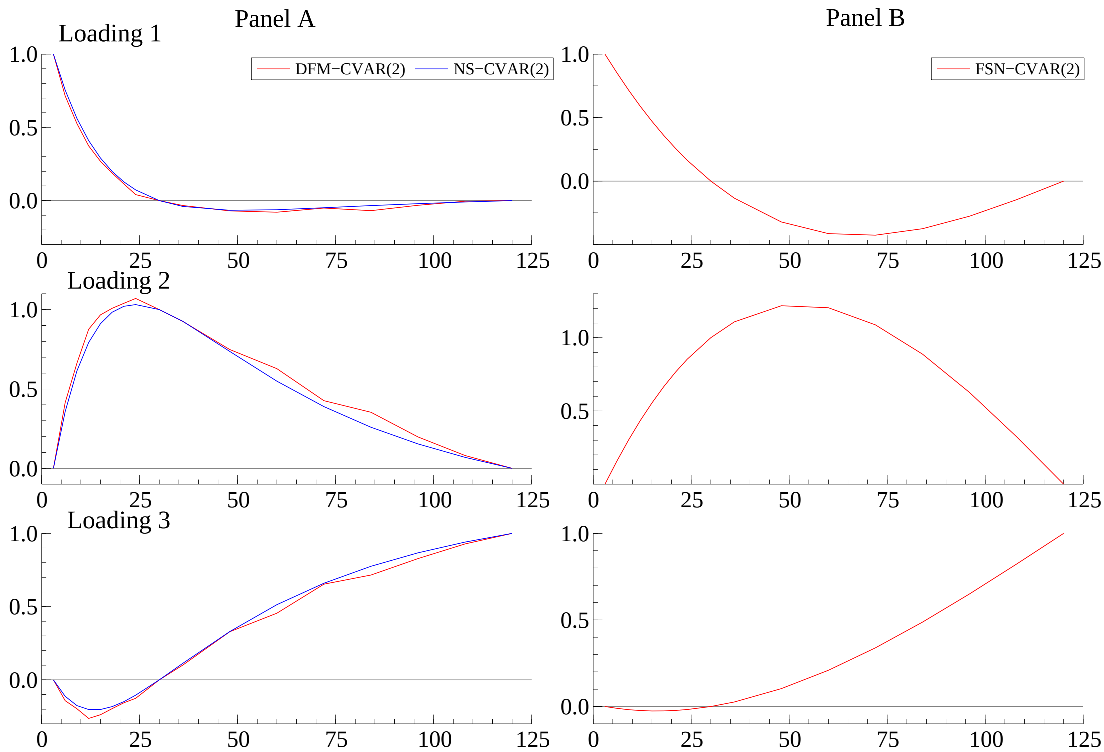
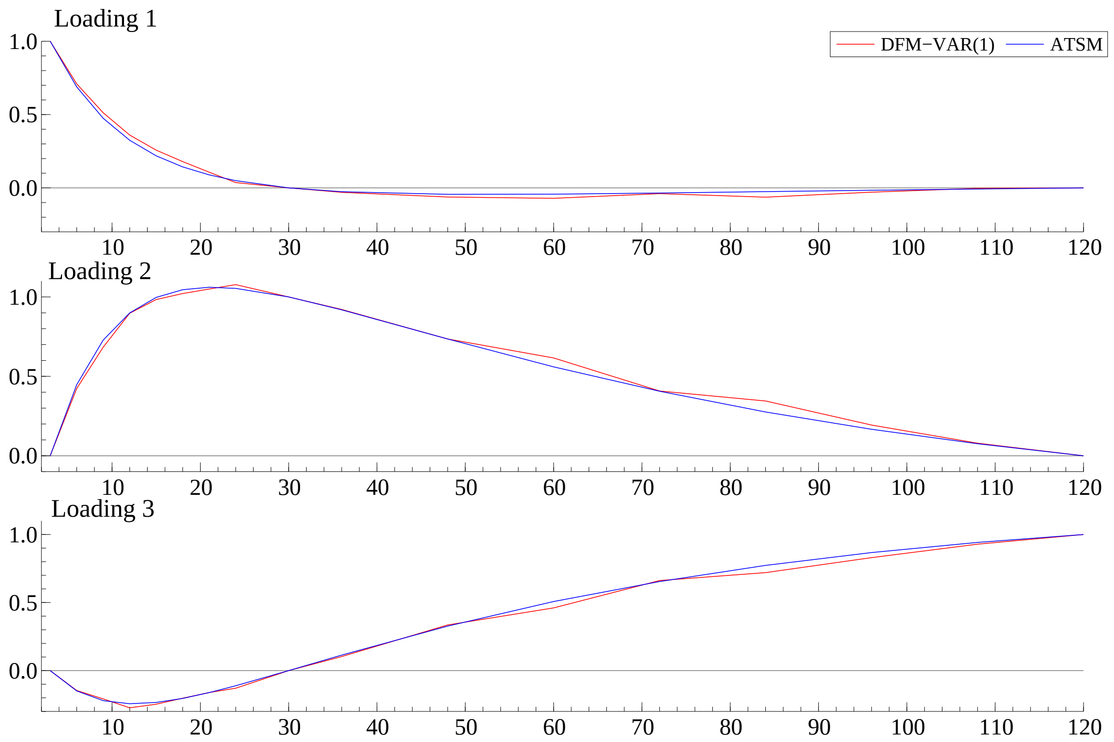


Figure 5: Estimated Factor Loadings Gaussian Affine Term Structure Model
This figure shows the estimated factor loadings for the Gaussian affine term structure model (AfTS) and the DFM-VAR(1) model as functions of time to maturity. The factor loadings for the AfTS are rotated such that the loading matrix is of the form (5), where the rows of the identity matrix are at the maturities of 3 months, 30 months and 120 months.



the optimization algorithm can therefore almost ignore the time series dimension. The parameters are then chosen such that the fit of the cross-section is as good as possible. The same conclusion can be drawn from Tables 3 and 7. The difference between the maximized loglikelihood values of the DFM-CVAR(1) and DFM-CVAR(2) models is much smaller than the difference between the values for the DFM-CVAR(2) and NS-CVAR(2) models while the factor loadings are very similar in all specifications, see Figure 4.

6.5 Model fit

In case all considered dynamic factor models are good approximations of the data generating process, we expect that the residuals are not serially correlated. To verify this, we computed Ljung-Box statistics for all models considered and for all 17 standardized residual series. The null hypothesis of the Ljung-Box tests is that residuals are a white noise sequence. If the null is rejected, dynamic variation in the residuals remains to be explained by a linear process. For a selection of the models we present the results of this procedure in Table 10. We see that the Nelson-Siegel model and to a lesser extent the DFM model are less successful in fitting the dynamics in the interest rates associated with times to maturity of 9 and 12 months. For these yields the FSN model seems to outperform the two other models. The FSN appears to explain the variation in the yields quite well for all maturities with the exception of the shortest 3 month maturity. It is not surprising that the Ljung-Box test statistics for the SDFM model are very similar to those for the general dynamic factor model. The arbitrage-free models perform a lot worse than the DFM, NS and FSN models, especially for the maturities from 6 to 24 months. For both the AFNS and AfTS models we reject the null hypothesis of the Ljung-Box test for 6 out of 17 residual series, at the 5% significance level. This lack of fit is surprising for the AFNS model since it is very similar to the standard Nelson-Siegel model. We can partly explain this from the restrictions of a VAR(1) process for the factors of the AFNS model. For comparison, we also present the Ljung-Box statistics for the standard Nelson-Siegel with VAR(1) factors in Table 10. In this case, the model only differs from the AFNS model in its restrictions imposed on the intercept. The NS-VAR(1) model also performs significantly worse in capturing the dynamics compared to the NS-CVAR(2) model. We stress however that it is not caused by the stationarity restriction of a VAR process. Ljung-Box statistics for the models with CVAR(2) factors and the VAR(2) specifications are very similar. In case of the AfTS model, we can explain the poor Ljung-Box statistics from the VAR(1) restriction and the estimates of the parameters in the transition matrix, see also the discussion at the end of section 6.4.

Table 10: Ljung-Box Statistics

This table shows Ljung-Box statistics calculated for the scaled residuals of some of the models discussed in this paper. Separate statistics are calculated for each maturity. We chose a number of 12 lags to calculate the test-statistics. The superscript $*$ is used to indicate rejection of the null hypothesis at the 10% significance level and $**$ is used for rejection at the 5% significance level. The headings 'CVAR(2) factors' and 'VAR(1) factors' indicate the specifications chosen for the factors.

Ljung-Box Statistics							
	CVAR(2) factors				VAR(1) factors		
Maturity	DFM	NS	FSN	SDFM	NS	AFNS	AfTS
3	5.8	6.2	84.3^{**}	6.0	11.6	10.5	17.9
6	7.1	7.4	11.2	7.4	12.6	11.8	33.1^{**}
9	19.2^{*}	19.3^{*}	11.6	18.7^{*}	31.8^{**}	39.9^{**}	55.1^{**}
12	22.5^{**}	29.2^{**}	16.7	23.1^{**}	36.2^{**}	53.1^{**}	52.6^{**}
15	15.9	17.8	16.0	15.6	25.7^{**}	36.9^{**}	28.5^{**}
18	12.8	13.0	13.2	12.7	22.2^{**}	28.1^{**}	22.2^{**}
21	12.2	12.0	13.8	12.2	18.8^{*}	22.4^{**}	19.1^{*}
24	10.2	11.2	15.3	10.6	18.9^{*}	21.6^{**}	22.0^{**}
30	9.3	9.4	10.5	9.1	17.2	15.8	16.0
36	8.7	9.1	10.3	8.3	16.1	14.8	15.2
48	6.2	6.0	7.7	5.4	12.6	11.1	11.1
60	5.9	5.7	9.2	5.6	11.5	10.3	11.2
72	5.7	5.5	8.2	5.9	11.7	10.5	10.8
84	8.4	9.4	10.4	9.1	15.3	12.9	17.4
96	7.7	7.6	8.6	7.9	11.8	10.8	14.1
108	9.2	8.6	9.0	9.2	10.5	9.7	11.6
120	10.1	9.7	6.9	11.0	9.7	9.3	12.2

Table 11: Results Likelihood Ratio Tests
This table presents the likelihood-ratio statistics for the null-hypothesis that the restrictions of the considered model are valid. The column k contains the number of restrictions imposed by the model. In panel A we show the VAR(2) versions of the NS, FSN and SDFM models. Panel B gives the CVAR(2) versions of the NS, FSN and SDFM models. Finally, we give the likelihood ratio statistics for the arbitrage-free models in Panel C.

Panel A: Stationary Models			
Model	LR	k	p-value
NS	220.2	41	0.000
FSN	879.0	42	0.000
SDFM	23.4	21	0.32

Panel B: Nonstationary Models			
Model	LR	k	p-value
NS	220.4	41	0.000
FSN	879.8	42	0.000
SDFM	20.2	20	0.45

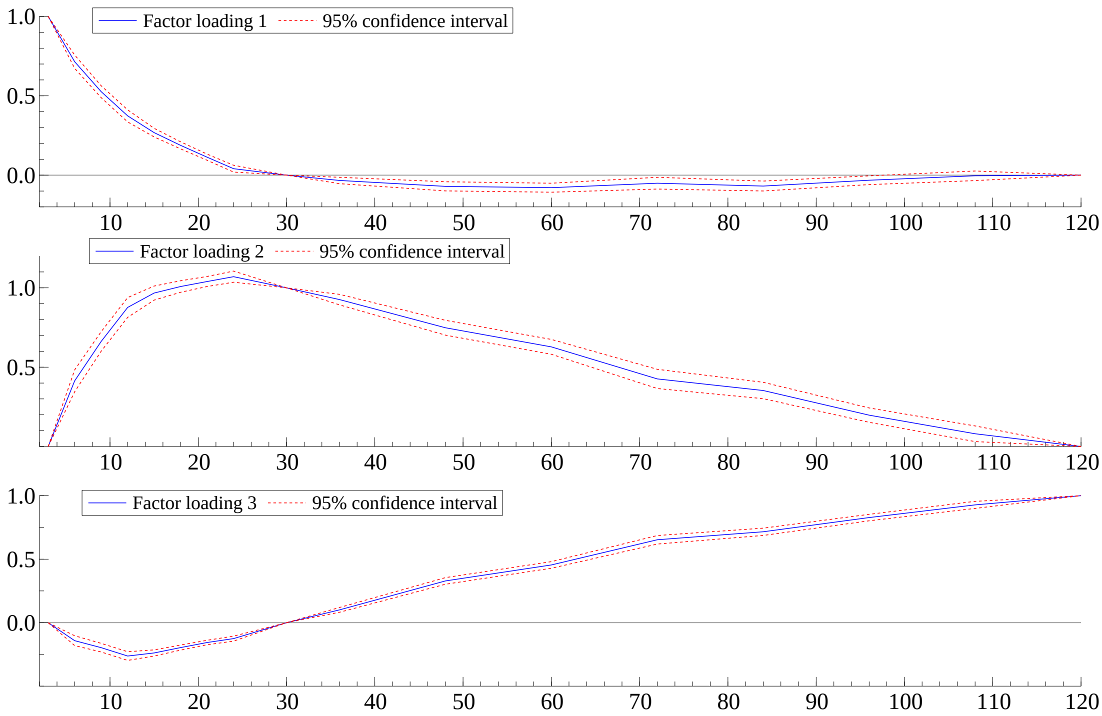
Panel C: Arbitrage-Free Models			
Model	LR	k	p-value
AFNS	1282.4	64	0.000
AfTS	930.2	76	0.000

6.6 Testing the restrictions on the factor loadings

In section 4 we have argued that all existing models considered in this paper can be viewed as dynamic factor models with smoothness restrictions imposed on the parameters. Since all models are nested in the general dynamic factor model, we can test the validity of these restrictions by means of a likelihood ratio test. For each model we test the null hypothesis that the restrictions are correct versus the alternative hypothesis that the true model is a general DFM model with the same dynamic specification for the factors. In case of the arbitrage-free models we indirectly test the restrictions on the intercepts. In Table 11 we present the established likelihood ratio tests. For the NS and FSN models, we focus on the VAR(2) and CVAR(2) specifications which we have selected by minimizing the AIC. Similar results are obtained for the other dynamic specifications. For the SDFM models we consider the final specifications presented in section 6.3.

The likelihood ratio tests suggest that we should strongly reject the restrictions implied by NS and FSN and those implied by the arbitrage-free specifications. For the FSN model,

Figure 6: Estimated Factor Loadings DFM with Confidence Bounds
This figure shows the estimated factor loadings for the DFM-CVAR(2) model as functions of time to maturity with 95% confidence intervals.



the differences between the estimated FSN and DFM loadings may have been indicative. However, the results reported in section 6.4 and 6.5 have shown that the loadings for the standard Nelson-Siegel model are very close to those estimated for the general DFM model. It is therefore somewhat surprising that the Nelson-Siegel restrictions are so strongly rejected. To find a possible explanation, we take a close look at the factor loading estimates for the DFM-CVAR(2) model as presented in Figure 6 together with their 95% confidence intervals. It is clear that the factor loadings of the DFM are estimated very precisely. The confidence intervals are very narrow, especially for the third column of the factor loading matrix Λ. It implies that small perturbations in the maximum likelihood estimates will cause large changes in the the loglikelihood value. This may explain why the Nelson-Siegel model is rejected by the likelihood ratio test although it is seemingly similar to the DFM model.

The strong rejection of these model specifications might suggest that we cannot impose smoothness restrictions on the factor loadings in dynamic factor models for the term structure of interest rates. However, for both the stationary as well as the nonstationary specifications, we cannot reject the restrictions imposed by the SDFM at any reasonable significance level. This finding confirms our believe that we can impose a certain amount of smoothness on the factor loadings for the yield curve. We do however require more flexibility in specifying the factor loadings than provided by the NS, FSN, AFSN and AfTS models. We have shown that the SDFM is sufficiently flexible for this purpose. A limited number of knots is required for each column of the factor loading matrix Λ. The exception is the third column of Λ. A possible explanation is the very narrow confidence intervals for the third factor loadings presented in Figure 6. It implies that a small misspecification in the third column of Λ will severely penalize the loglikelihood value.

7 Conclusion

In this paper we have discussed term structure modelling by means of dynamic factor models with smooth factor loadings. We introduced a new methodology to construct dynamic factor models with smooth factor loadings and proposed a statistical procedure to find suitable smoothness restrictions. For the data set of unsmoothed Fama-Bliss zero yields for US treasuries we show that, using our new methodology, it is possible to construct a parsimonious dynamic factor model with smooth factor loadings. The number of parameters in the loading matrix of our dynamic factor model is almost 50% smaller than the number of parameters in the loading matrix of the unrestricted dynamic factor model. Despite of this large reduction in the number of parameters we find that the fit of our model is qualitatively the same as for the most general dynamic factor model. The restricted model is not rejected by a likelihood ratio test with the alternative hypothesis that the true data generating process is an unrestricted dynamic factor model. We also investigated the validity of the restrictions imposed by a number of popular term structure models. For all of these models the restrictions were strongly rejected by the likelihood ratio test. We conclude that there is clear evidence for the validity of smoothness restrictions on the factor loadings of dynamic factor models for the term structure. However, a number of popular models is not flexible enough to capture the correct shape of the loadings.

Although the analysis of interest rate term structures is of key interest in finance and economics, we emphasize that dynamic factor models with smooth factor loadings can be used in many different settings. In many applications of the dynamic factor model it is possible to identify variables of which the factor loadings can reasonably be assumed to be smooth functions. In this way we can build parsimonious dynamic factor models even for high dimensional time series panels. We plan to explore this methodology in future work.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–72.

Bliss, R. (1997). Movements in the term structure of interest rates. *Federal Reserve Bank of Atlanta Economic Review 82*, 16–33.

Bowsher, C. and R. Meeks (2008). The dynamics of economic functions: modeling and forecasting the yield curve. *J. American Statistical Association 103*, 1419–37.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis, Forecasting and Control* (3rd ed.). San Francisco: Holden-Day.

Brigo, D. and F. Mercurio (2006). *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit (Springer Finance)*. Springer.

Christensen, J., F. Diebold, and S. Rudebusch (2007). The affine arbitrage-free class of Nelson-Siegel term structure models. Manuscript University of Pennsylvania.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *Journal of Finance 48*(4), 1263–91.

Cox, J. C., J. Ingersoll, Jonathan E, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *J. Business and Economic Statist. 18*(3), 300–14.

De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models. Tinbergen Institute Discussion Paper.

Dickey, D. C. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. American Statistical Association 74*, 427–31.

Diebold, F. and C. Li (2006). Forecasting the term structure of government bond yields. *J. Econometrics 130*, 337–64.

Diebold, F., S. Rudebusch, and S. Aruoba (2006). The macroeconomy and the yield curve. *J. Econometrics 131*, 309–38.

Doz, C., D. Giannone, and L. Reichlin (2006). A quasi maximum likelihood approach for large approximate dynamic factor models. Discussion paper, CEPR.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Journal of Mathematical Finance 6*(2), 379–406.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Engle, R. F. (1984, July). Wald, likelihood ratio, and lagrange multiplier tests in econometrics. In *Handbook of Econometrics*, Volume 2 of *Handbook of Econometrics*, Chapter 13, pp. 775–826. Elsevier.

Engle, R. F. and M. W. Watson (1981). A one-factor multivariate time series model of metropolitan wage rates. *J. American Statistical Association 76*, 774–81.

Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review 77*, 680–92.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic factor model: Identification and estimation. *Rev. Economics and Statistics 82*, 540–54.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam: North-Holland.

Gregory, A., A. Head, and J. Raynauld (1997). Measuring world business cycles. *International Economic Review 38*, 677–701.

Hall, A. D., H. M. Anderson, and C. W. J. Granger (1992). A cointegration analysis of treasury bill yields. *Review of Economics and Statistics 74*(1), 116–26.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *J. American Statistical Association 88*, 1228–36.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Jungbacker, B. and S. J. Koopman (2008). Likelihood-based analysis for dynamic factor models. Tinbergen Institute Discussion Paper.

Koopman, S. J. and J. Durbin (2000). Fast filtering and smoothing for multivariate state space models. *J. Time Series Analysis 21*, 281–96.

Koopman, S. J., M. Mallee, and M. Van der Wel (2009). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *J. Business and Economic Statist. 27*, forthcoming.

Koopman, S. J. and N. Shephard (1992). Exact score for time series models in state space form. *Biometrika 79*, 823–6.

Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income 1*(1), 54–61.

Monahan, J. F. (2001). *Numerical methods of statistics*. Cambridge: Cambridge University Press.

Nelson, C. and A. Siegel (1987). Parsimonious modelling of yield curves. *Journal of Business 60-4*, 473–89.

Nocedal, J. and S. J. Wright (1999). *Numerical Optimization*. New York: Springer Verlag.

Poirier, D. J. (1976). *The Econometrics of Structural Change: with Special Emphasis on Spline Functions*. Amsterdam: North-Holland.

Reis, R. and M. W. Watson (2007). Relative goods prices and pure inflation. NBER Working paper.

Rosenberg, B. (1973). Random coefficients models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*, 399–428.

Sargent, T. J. and C. A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. A. S. et al. (Ed.), *New methods in business cycle research*. Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, J. H. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *J. Business and Economic Statist. 20*, 147–62.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. *J. Econometrics 23*, 385–400.