State-Space Modelling of Dynamic Factor Structures, with an Application to the U.S. Term Structure

Pieter W. Otter* and Jan P.A.M. Jacobs
University of Groningen

June 2008

Econometrics, Time Series Analysis and Systems Theory:
A Conference in Honor of Manfred Deistler
Vienna, June 18–20, 2009

Abstract

This paper applies the formal test of Jacobs and Otter (2008) [Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach, *Econometric Reviews*, 27, 385–397] for the number of factors and lag order in a dynamic factor model, which is based on canonical correlations and related to entropy and Kullback-Leibler numbers. The testing procedure is used to cast the dynamic factor model in state-space form. We apply the proposed framework to U.S. yield curve data, with special emphasis on prediction and prediction errors.

Keywords: dynamic factor structure, state-space models, Kalman filter, entropy, prediction, U.S. yield curve

JEL-code: C32, C52, C82, E43

*Corresponding author: Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. Tel.: +31 50 363 3782. Email: p.w.otter@rug.nl

1 Introduction

A widely used method to analyse large quantities of data in the social sciences is factor analysis, in which the variation in a large number of observed variables is described in fewer unobserved variables, or movements in a large number of series are driven by a limited set of common 'factors'. In recent years large dimensional factor models have become more and more popular in econometric research too. For an overview of recent developments see Bai and Ng (2008). Especially dynamic factors models, one of research agenda items of Prof. Manfred Deistler (see e.g. Hamann, Deistler and Scherrer (2005) and Deistler and Zinner (2007)), in which the factors may enter the model with leads and lags, are widely applied. One of the issues in factor analysis is the determination of the number of unobserved variables to retain, i.e. the number of factors. Recently many studies have proposed solutions and consistent estimators using different factor model and distributional assumptions. See e.g. Connor and Korajczyk (1993); Bai and Ng (2002, 2007), Amengual and Watson (2007), Kapetanios (2005), Hallin, and Liška (2007) Jacobs and Otter (2008), and Onatski (2009).

In this paper we cast a dynamic factor model in state-space form and apply the Kalman filter for estimation and prediction. This approach has been around for some time—early examples are Otter (1981) and Molenaar (1985)—but has recently been reintroduced by Kapetanios and Marcellino (2006), Doz, Giannone and Reichlin2007, Jungbacker and Koopman (2008), and Aruoba, Diebold and Scotti (2009). To convert the dynamic factor model into state-space from one needs to know the number of factors and the lag order. The formal test of Jacobs and Otter (2008), which is based on canonical correlations and related to entropy and Kullback-Leibler numbers, is especially handy in this respect. Autocovariances are computed for increasing lag parameters, and analysed through singular value decompositions (SVD); the optimal lag is determined at p if the rank of the canonical correlation matrix of the SVD of the $(p-1)$th autocovariance differs from zero whereas the rank of the pth autovariance is equal to zero.

Factor models are quite common in financial econometrics, see e.g. Deistler and Hamann (2005). We apply the proposed framework to daily U.S. term structure data. In modeling yields, the Nelson and Siegel (1987) functional form, which is a convenient and parsimonious three-component exponential approximation, is often used. See for example Bolder and Streblinski (1999), Diebold and Li (2006), Diebold, Li and Yue (2008). Three latent dynamic factors are distinguished, which are called long-term, short-term and medium term or interpreted in terms of level, slope and curvature. The three factors are typically set a priori, and not obtained from first principles. However, Bliss (1997) finds three factors using statistical factor analysis, while recently Bekker and Bouwman (2009) identify and estimate three factors starting from an arbitrage-free model for risk-free interest rates with capital market maturities, where the parameters have a clear economic interpretation.

Unfortunately, the Jacobs-Otter (2008) testing procedure does not work for the daily U.S. term structure data. The yield curve has (almost) the same shape in terms of level, slope, and curvature for all periods, so the observations are strongly correlated at all lags, causing our test to break down. A possible explanation might be the high frequency of the data set;

daily data contain a lot of noise. Instead we determined the number of factors k on the basis of prediction errors of one-step ahead yield curve predictions calculated by the Kalman filter in an estimated State-Space model and found that $k = 4$ was superior to the generally obtained $k = 3$ factors.

The remainder of the paper is structured as follows. Section 2 describes the dynamic factor model, the state-space model and the Kalman filter. Section 3 summarizes the estimation procedure of Jacobs and Otter (2008) for the number of factors and lags, and other estimation issues of the state-space model, while Section 4 shows the application. Section 5 concludes.

2 Method

2.1 The dynamic factor model

Let $\boldsymbol{x}_t$ be an N-dimensional vector of observed data at time t, $t = 1, \ldots, T$, which is driven by q dynamic factors $\boldsymbol{u}_t$ with loadings $\boldsymbol{B}_j$ up to lag p, i.e. $j = 1, \ldots, p$, and idiosyncratic components $\boldsymbol{\varepsilon}_t$

$$\boldsymbol{x}_t = \boldsymbol{B}_0 \boldsymbol{u}_t + \boldsymbol{B}_1 \boldsymbol{u}_{t-1} + \ldots + \boldsymbol{B}_p \boldsymbol{u}_{t-p} + \boldsymbol{\varepsilon}_t. \quad (1)$$

Equation (1) is the (dynamic) factor representation of the data. Note that factors, loadings and idiosyncratic components are not observable. In vector notation the model becomes

$$\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_1 & \dots & \boldsymbol{B}_p \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_t \\ \boldsymbol{u}_{t-1} \\ \vdots \\ \boldsymbol{u}_{t-p} \end{pmatrix} + \boldsymbol{\varepsilon}_t \equiv \boldsymbol{B}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t. \tag{2}$$

We make the following assumptions. First, the q-dimensional vector of factors $\boldsymbol{u}_t$ is Gaussian White Noise (GWN) with $\mathrm{E}(\boldsymbol{u}_t) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{u}_t) = \boldsymbol{I}_q$, the q-dimensional identity matrix. Secondly, the idiosyncratic components $\boldsymbol{\varepsilon}_t$ is GWN with $E(\boldsymbol{\varepsilon}_t) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{V}$ and factors $\boldsymbol{u}_t$ and idiosyncratic components $\boldsymbol{\varepsilon}_t$ are independent. This assumptions imply that the generalized dynamic factor model of Forni, Hallin, Lippi and Reichlin (2000a) and Forni and Lippi (2001), which allows some correlation among idiosyncratic components, can be dealt with too. Thirdly, the matrix of loadings $\boldsymbol{B}$ has full (column-)rank, i.e. $\mathrm{rank}(\boldsymbol{B}) = (p+1)q$ with $k \equiv (p+1)q < N$; this implies $\boldsymbol{F}_t \in \mathbb{R}^k$.

2.2 The State-Space (SS) model and the Kalman filter

The dynamic factor model of Equation (2) can be written in state-space form with $\boldsymbol{F}_t$ as state vector in the following way

$$\boldsymbol{F}_t = \boldsymbol{A}\boldsymbol{F}_{t-1} + \boldsymbol{G}\boldsymbol{w}_t \tag{3}$$

$$\boldsymbol{x}_t = \boldsymbol{B}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{4}$$

where $\boldsymbol{A} \equiv \begin{bmatrix} 0 & \dots & 0 & 0 \\ \boldsymbol{I}_q & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \dots & \boldsymbol{I}_q & \boldsymbol{0} \end{bmatrix}$, $\boldsymbol{A}^{p+1} = 0$ (nilpotent), $\boldsymbol{G}' = \begin{bmatrix} \boldsymbol{I}_q & \boldsymbol{0} & \dots & \boldsymbol{0} \end{bmatrix}$, $\boldsymbol{G} \in \mathbb{R}^{k \times q}$, $\boldsymbol{w}_t \sim \text{GWN}(\boldsymbol{0}, \boldsymbol{I}_q)$, and $\boldsymbol{\varepsilon}_t \sim \text{GWN}(\boldsymbol{0}, \boldsymbol{V}_\varepsilon)$.

Now we can set the Kalman filter to work. The initial estimate of the state vector is $\hat{\boldsymbol{F}}_0 = \boldsymbol{0}$ with variance $\boldsymbol{V}_{\tilde{\boldsymbol{F}}_0} = \boldsymbol{I}_k$, with one-step ahead prediction $\hat{\boldsymbol{F}}_{t+1|t} = \boldsymbol{A}\boldsymbol{F}_t$. The a priori variance is equal to $\boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1|t}} = \boldsymbol{A}\boldsymbol{V}_{\tilde{\boldsymbol{F}}_t}\boldsymbol{A}' + \boldsymbol{G}\boldsymbol{G}' = \boldsymbol{I}_k$.

The Kalman filter algorithm gives $\hat{\boldsymbol{F}}_{t+1} = \hat{\boldsymbol{F}}_{t+1|t} + \boldsymbol{K}_{t+1}\boldsymbol{pe}_{t+1}$, with prediction-error $\boldsymbol{pe}_{t+1} = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t+1|t}$, where $\hat{\boldsymbol{x}}_{t+1|t} = \boldsymbol{B}\hat{\boldsymbol{F}}_{t+1}$. The gain algorithm yields $\boldsymbol{K}_{t+1} = \boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1}}\boldsymbol{B}' \left[\boldsymbol{B}\boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1}}\boldsymbol{B}' + \boldsymbol{V}_\varepsilon\right]^{-1}$ and we obtain the a posteriori variance $\boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1}} = (\boldsymbol{I} - \boldsymbol{K}_{t+1}\boldsymbol{B})\boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1}}$.

The steady-state filter converges to $\boldsymbol{V}_{\tilde{\boldsymbol{F}}_{t+1}} \longrightarrow \boldsymbol{V}_{\tilde{\boldsymbol{F}}}$ where $\hat{\boldsymbol{F}}_{t+1|t} = \boldsymbol{A}\hat{\boldsymbol{F}}_{t|t-1} + \boldsymbol{A}\boldsymbol{K}(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t-1})$ with gain: $\boldsymbol{K} = \boldsymbol{V}_{\tilde{\boldsymbol{F}}}\boldsymbol{B}' \left[\boldsymbol{B}\boldsymbol{V}_{\tilde{\boldsymbol{F}}}\boldsymbol{B}' + \boldsymbol{V}_\varepsilon\right]^{-1}$ and variance: $\boldsymbol{V}_{\tilde{\boldsymbol{F}}} = \boldsymbol{A}\left[\boldsymbol{V}_{\tilde{\boldsymbol{F}}} - \boldsymbol{K}\boldsymbol{B}\boldsymbol{V}_{\tilde{\boldsymbol{F}}}\right]\boldsymbol{A}' + \boldsymbol{G}\boldsymbol{G}'$.

3 Estimation

3.1 The number of lags and dynamic factors[1]

Jacobs and Otter (2008) give a procedure to estimate the number of lags p and the number of dynamic factors q in the dynamic factor model of Equation (1). First the N components of the data matrix $\boldsymbol{x}_t$ are demeaned and

[1]This section draws upon Jacobs and Otter (2008).

unit variances are taken, which produce $\tilde{\boldsymbol{x}}_t$. Let

$$\hat{\boldsymbol{\Gamma}}_i = \frac{1}{T-i} \sum_{t=i}^{T} \tilde{\boldsymbol{x}}_t \tilde{\boldsymbol{x}}'_{t-i}, \quad i = 0, 1, 2, \ldots$$

be a consistent estimate of the autocovariances $\boldsymbol{\Gamma}_i$. Assuming that rank$(\hat{\boldsymbol{\Gamma}}_0)$ has full rank N, we can apply the spectral decomposition

$$\hat{\boldsymbol{\Gamma}}_0 = \boldsymbol{C}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{C}' = \hat{\boldsymbol{\Gamma}}_0^{1/2}(\hat{\boldsymbol{\Gamma}}_0^{1/2})', \tag{5}$$

with $\hat{\boldsymbol{\Gamma}}_0^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'$ and the Singular Value Decomposition (SVD)

$$\hat{\boldsymbol{\Gamma}}_0^{-1/2}\hat{\boldsymbol{\Gamma}}_i(\hat{\boldsymbol{\Gamma}}_0^{-1/2})' = \boldsymbol{H}_i\boldsymbol{S}_i\boldsymbol{Q}_i \tag{6}$$

where the columns of $\boldsymbol{H}$ and $\boldsymbol{Q}$ are orthogonal, i.e. $\boldsymbol{H}'_i\boldsymbol{H}_i = \boldsymbol{H}_i\boldsymbol{H}'_i = \boldsymbol{I}_N$ and $\boldsymbol{Q}'_i\boldsymbol{Q}_i = \boldsymbol{Q}_i\boldsymbol{Q}'_i = \boldsymbol{I}_N$, and $\boldsymbol{S}_i = \text{diag}(s_{i,1}, s_{i,2}, \ldots, s_{i,N})$, an $N \times N$ diagonal matrix with singular values $s_{i,j} \in [0,1]$ and $s_{i,1} > s_{i,2} > \ldots > s_{i,N} \geq 0$. The canonical correlation coefficients (singular values) $s_{i,j}$ are estimates of the equivalent population canonical coefficients $\rho_{i,j}$, see Otter (1990, 1991).

To test the null hypothesis that the $N-k$ smallest population canonical coefficients for the ith autocovariance are equal to zero $H_0 : \rho_{i,k+1} = \ldots = \rho_{i,N} = 0$, we calculate Bartlett's test statistic

$$\chi^2 = -\left[T - \frac{1}{2}(2N+1)\right] \sum_{j=k+1}^{N} \ln(1 - s_{i,j}^2), \tag{7}$$

for all values of $k = 0, 1, 2, \ldots$. This statistic follows a χ^2 distribution under the null with degrees of freedom $df = (N-k)^2$.

The procedure essentially comes down to the linear transformation of $\tilde{\boldsymbol{x}}_t$ and $\tilde{\boldsymbol{x}}_{t-i}$ into canonical vectors $\boldsymbol{y}_t = \boldsymbol{A}_i\tilde{\boldsymbol{x}}_t$ and $\boldsymbol{y}_{t-i} = \boldsymbol{G}_i\tilde{\boldsymbol{x}}_{t-i}$ with $\boldsymbol{A}_i = \boldsymbol{H}_i'\hat{\boldsymbol{\Gamma}}_0^{-1/2}$ and $\boldsymbol{G}_i = \boldsymbol{Q}_i\hat{\boldsymbol{\Gamma}}_0^{-1/2}$ with $E(\boldsymbol{y}_t\boldsymbol{y}_{t-i}') = \boldsymbol{S}_i$ and unit variance matrices. The conditional variance of $\boldsymbol{y}_t$ given $\boldsymbol{y}_{t-1}$ equals $(\boldsymbol{I}_N - \boldsymbol{S}_i^2)$. From Equation (2) and the normalisation we have $\boldsymbol{\Gamma}_i = \boldsymbol{D}^{-1/2}\boldsymbol{B}^{(i)}\boldsymbol{B}'\boldsymbol{D}^{-1/2}$ where $\boldsymbol{D} = \mathrm{diag}(\sigma_1^2, ..., \sigma_N^2)$ with the variances of the components of $\boldsymbol{x}_t$ as elements. The $N \times (p+1)q$ dimensional matrix $\boldsymbol{B}^{(i)} = (\boldsymbol{B}_i \ \boldsymbol{B}_{i+1} \ \ldots \ \boldsymbol{B}_p \ \boldsymbol{0} \ \ldots \ \boldsymbol{0})$ has rank $(p+1-i)q$ and hence the rank of $\boldsymbol{\Gamma}_i$ is $(p+1-i)q$ for lags $i = 1, .., p$ and zero for lags greater than p. The ranks of $\hat{\boldsymbol{\Gamma}}_i$ are estimated by the number of significant singular values using Bartlett's test statistic as follows.

If for a given significance level the hypothesis that all population canonical coefficients for the $(p+i)$th autocovariance $(i > 0)$ are equal to zero, i.e. $H_0 : \rho_{p+i,1} = \rho_{p+i,2} = \ldots = \rho_{p+i,N} = 0$, is accepted whereas the hypothesis that all population canonical coefficients for the (p)th autocovariance are equal to zero, $H_0 : \rho_{p,1} = \rho_{p,2} = \ldots = \rho_{p,N} = 0$ is rejected, but the hypothesis that the $(N-q)$ smallest canonical coefficients are equal to zero, $H_0 : \rho_{p,q+1} = \rho_{p,q+2} = \ldots = \rho_{p,N} = 0$, is accepted, then the estimated lag order is p and the estimated number of factors or $\dim(\boldsymbol{u}_t)$ equals q.

3.2 Estimation and Kalman prediction

Let the $T \times N$ data matrix $\boldsymbol{X}$ consist of normalized realizations of $\boldsymbol{x}_t \in \mathbb{R}^N$, $t = 1, 2, \ldots, T$: $\boldsymbol{X}' = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_T \end{bmatrix}$, where $\boldsymbol{x}_t$ is the yield vector at time t. The singular value decomposition (SVD) to $\boldsymbol{X}$ yields $\boldsymbol{X} =$

$\boldsymbol{USC}'$, with $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{UU}' = \boldsymbol{I}_T$, $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{CC}' = \boldsymbol{I}_N$, $\boldsymbol{S} \equiv \begin{bmatrix} \boldsymbol{S}_1 \\ \boldsymbol{0} \end{bmatrix}$, and $\boldsymbol{S} = \text{diag}(s_1,\ s_2,\ \ldots, s_N)$, $s_1 > s_2 > \ldots > s_N$.

Select $k < kmax \equiv \min(N, T)$ and apply a *least squares decomposition*

$$\boldsymbol{X} = \boldsymbol{U}_k\boldsymbol{S}_k\boldsymbol{C}_k' + \boldsymbol{U}_2\boldsymbol{S}_2\boldsymbol{C}_2' \equiv \hat{\boldsymbol{X}} + \boldsymbol{E} \tag{8}$$

where $\boldsymbol{U}_k = (\boldsymbol{u}_1 \ldots \boldsymbol{u}_k)$, $\boldsymbol{U}_2 = (\boldsymbol{u}_{k+1} \ldots \boldsymbol{u}_{kmax})$, $\boldsymbol{C}_k = (\boldsymbol{c}_1 \ldots \boldsymbol{c}_k)$, $\boldsymbol{C}_2 = (\boldsymbol{c}_{k+1} \ldots \boldsymbol{c}_{kmax})$, $\boldsymbol{S}_k = \text{diag}(s_1 \ldots s_k)$ and $\boldsymbol{S}_2 = \text{diag}(s_{k+1} \ldots s_{kmax})$ with $s_1 \geq s_2 \geq \ldots \geq s_{kmax} \geq 0$, $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{C}_k & \boldsymbol{C}_2 \end{bmatrix}$, $\boldsymbol{C}_k \in \mathbb{R}^{T\times k}$, $\boldsymbol{C}_2 \in \mathbb{R}^{T\times(kmax-k)}$, and $\boldsymbol{C}_k'\boldsymbol{C}_k = \boldsymbol{I}_k$. The Euclidean norm of the errors $\boldsymbol{E}$ is equal to $||\boldsymbol{E}||_E^2 = ||\boldsymbol{X} - \hat{\boldsymbol{X}}||_E^2 = \text{tr}(\boldsymbol{E}'\boldsymbol{E}) = \sum_{j=k+1}^{kmax} s_j^2$, which is the minimum. Note that the least squares properties $\boldsymbol{E}'\hat{\boldsymbol{X}} = \boldsymbol{V}_2\boldsymbol{S}_2\boldsymbol{U}_2'\boldsymbol{U}_k\boldsymbol{S}_k\boldsymbol{V}_k' = 0$ and $\hat{\boldsymbol{X}}'\boldsymbol{E} = 0$ are satisfied because of the orthogonality of $\boldsymbol{U}$.

Taking transposes of the least squares decomposition (8) we get

$$\boldsymbol{X}' = \hat{\boldsymbol{X}}' + \boldsymbol{E}' = \begin{bmatrix} \hat{\boldsymbol{x}}_1 & \ldots & \hat{\boldsymbol{x}}_T \end{bmatrix} + \begin{bmatrix} e_1 & \ldots & e_T \end{bmatrix},$$

where $\hat{\boldsymbol{X}}'$ is equal to

$$\hat{\boldsymbol{X}}' = \begin{bmatrix} \hat{\boldsymbol{x}}_1 & \ldots & \hat{\boldsymbol{x}}_T \end{bmatrix} = \boldsymbol{C}_k\boldsymbol{S}_k \begin{bmatrix} \hat{\boldsymbol{F}}_1 & \ldots & \hat{\boldsymbol{F}}_T \end{bmatrix},$$

with $\boldsymbol{U}_k' \equiv \begin{bmatrix} \hat{\boldsymbol{F}}_1 & \ldots & \hat{\boldsymbol{F}}_T \end{bmatrix}$, and $\boldsymbol{U}_k'\boldsymbol{U}_k = \sum_{t=1}^{T} \hat{\boldsymbol{F}}_t\hat{\boldsymbol{F}}_t' = \boldsymbol{I}_k$. The second component of the least squares decomposition (8), the errors $\boldsymbol{E}'$, can be written as

$$\boldsymbol{E}' = \begin{bmatrix} e_1 & \dots & e_T \end{bmatrix} = \boldsymbol{C}_2 \boldsymbol{S}_2 \boldsymbol{U}_2',$$

with $\hat{\boldsymbol{V}}_\epsilon \equiv \boldsymbol{E}'\boldsymbol{E} = \boldsymbol{C}_2 \boldsymbol{S}_2^2 \boldsymbol{C}_2'$. Note that $\hat{\boldsymbol{V}}_\epsilon$ also captures an approximate factor structure, which is more general than the exact factor structure.

4 Application

4.1 Data

Our data consists of continuously compounded annualized U.S. interest rates, swap rates to be precisely, for three months, one year, two years, three years, four years, five years, seven years, and ten years, for 3468 consecutive days, covering the period September 1988 until May 2005, incomplete days excluded. Figure 1 shows our term structure data in 3D format. We have $N = 8$ interest rates for $T = 3468$ days. The figure shows that the term structure was normal, i.e. the ten year interest rates exceeding the three months rate, for most of the sample, but inverted for the first 650 observations.

[Figure 1 about here.]

Table 1 gives some additional information on the interest rates in the form of descriptive statistics. The 10-year rate, usually referred to as the capital market rate has the highest mean and minimum, but the lowest standard deviation. The 3-month rate, the money market rate attains the highest value, and almost the lowest value, and shows the highest volatility (standard deviation).

[Table 1 about here.]

4.2 Estimation and testing results

First we applied the Jacobs-Otter (2008) testing procedure described in Section 3.1 to the normalized term structure data. Unfortunately, our method did not work well for this data set. The yield curve has (almost) the same shape in terms of level, slope, and curvature for all periods, so the observations are strongly correlated at all lags, and therefore our test breaks down. We decided to use the testing procedure of Section 3.2 and obtained an estimated State-Space model based on $T_1 = 2000$ observations, and to use the Kalman Filter to predict the yield curve one-step ahead for $T = T_1, T_1 + 1, T_1 + 2, \ldots$, where the number of factors is based on the sum of squared prediction-errors.

The estimated state-space model is

$$\begin{aligned} \boldsymbol{F}_{t+1} &= \hat{\boldsymbol{A}}\boldsymbol{F}_t + \hat{\boldsymbol{G}}\boldsymbol{w}_t, \qquad t = T_1 + 1, \ldots, T, \\ \boldsymbol{X}_t &= \hat{\boldsymbol{C}}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \end{aligned}$$

where $\hat{\boldsymbol{A}} = \sum_2^{T_1} \hat{\boldsymbol{F}}_t \hat{\boldsymbol{F}}_{t-1}' = \begin{bmatrix} 0.9994 & 0.0001 & -0.0002 & -0.0005 \\ -0.0006 & 0.9991 & -0.0022 & -0.0001 \\ 0.0012 & 0.0018 & 0.9966 & -0.0019 \\ -0.0003 & -0.0018 & -0.0009 & 0.9712 \end{bmatrix}$, $\hat{\boldsymbol{V}}_\varepsilon = \boldsymbol{C}_2 \boldsymbol{S}_2^2 \boldsymbol{C}_2'$, $\hat{\boldsymbol{C}} = \boldsymbol{C}_k \boldsymbol{S}_k$, and $\hat{\boldsymbol{G}} = \mathrm{diag}(\hat{g}_1, \ldots, \hat{g}_k)$ with $\hat{g}_j = (\max(\boldsymbol{u}_j) - \min(\boldsymbol{u}_j))/6$, $j = 1, \ldots, k$, $\hat{\boldsymbol{G}} = \begin{bmatrix} 0.0035 & 0.0118 & 0.0161 & 0.0186 \end{bmatrix}$ and $\boldsymbol{U}_k =$

$$\begin{bmatrix} \hat{\boldsymbol{F}}_1' \\ \vdots \\ \hat{\boldsymbol{F}}_{T_1}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_k \end{bmatrix},$$ with $\boldsymbol{u}_j \in \mathbb{R}^{T_1}$ and orthonormal. All estimates are based on T_1 observations. In the Kalman filter the initial value for the state vector is $\hat{\boldsymbol{F}}_{T_1}$.

The squared prediction error (PE) can be expressed as

$$||PE||_E^2 = \sum_{i=1}^{N} \sum_{t=T_1+1}^{T} (pe_{it})^2 = \sum_{j=1}^{N} s_j^2 = \text{tr}(\boldsymbol{PE}'\boldsymbol{PE}),$$

with $\boldsymbol{PE}' \equiv \begin{bmatrix} \boldsymbol{pe}_{T_1+1} & \dots & \boldsymbol{pe}_T \end{bmatrix} \in \mathbb{R}^{N\times(T-T_1)}$. Table 2 lists the squared prediction errors for different values of k. As can be seen, the squared prediction errors are very small. As written in the Introduction, three latent factors are generally distinguished. Taking aboard a fourth factor however decreases the prediction error considerably. In the remainder of this section, our analysis will therefore be based on a dynamic factor model with $k = 4$ factors.

[Table 2 about here.]

Figure 2 shows the histograms of the prediction errors of the eight interest rates in our data, which gives an indication of the quality of the predictions produced by our estimated state-space structure. The prediction error histograms for the eight interest rates are not only normally distributed but also small in magnitude. Note that the figures on the x-axes have to be multiplied by 10^{-3}.

[Figure 2 about here.]

Figure 3 presents histograms of the estimated factors, for $t = 1, \dots, 2000$. It is easily seen that the estimated factors are not normally distributed.

[Figure 3 about here.]

Figures 4 and 5 shows the four factors and the factor loadings. Factor loading 1 together with the time series of estimated factor 1 represent the general level which is decreasing throughout the estimation period, see also Figure 1. Factor loading 2 in Figure 4 is the slope, i.e. the spread between the 10-years interest rate and the 3-months interest rate. The graph of factor 2 in Figure 5 illustrates that the yield curve was inverted for the first 650 observations, with a positive sign for the short-term rates for the first 650 observations and a negative sign for the long-term rates. After the first 650 observations the yield curve becomes normal. Note that the slope is flattening in the second period. The curvature, factor loading 3 in Figure 4, is negative for both short-term rates and long-term rates, and positive in between. However, factor 3 in Figure 5 is changing sign over time. The fourth factor loading, the bottom right graph in Figure 4, is positive for the 3-months interest rate and medium-term interest rates, and negative for the 10-years interest rate, but with periodically changing signs and a negative trend at the end of the sample. At the moment, a clear interpretation is lacking.

[Figure 4 about here.]

[Figure 5 about here.]

5 Conclusion

In this preliminary study our main goal was to apply the Jacobs and Otter (2008) testing procedure to yield curve data in order to get a dynamic factor model. However the testing procedure broke down due to the special time-invariant features of the yield curve, namely the level, the slope, and the curvature. The least-squares procedure, based on the Euclidean matrix norm, proved to be useful to gain insight into the time series of the individual weights (factors) of each of the interest rates on the general, time-invariant features of the yield curve (factors loadings).

In future research we will analyze yield curve data of the U.S. and Europe, in particular Germany, and apply our testing procedure to detect Wiener-Granger causality or feedback processes. In addition we will investigate the sampling frequency of the data to detect long-term systematic movements in the yield curve, to overcome the problem that high-frequency data induce a lot of noise.

Acknowledgements

Part of the material of this paper was presented at The Far Eastern Meeting of the Econometric Society, Beijing, China, July 2006 (Otter and Jacobs, 2006). We thank SNS asset management, and in particular Theo Dijkstra and Bob Stoutjesdijk, for the provision of the U.S. term structure data.

References

Amengual, D. and M. Watson (2007), “Consistent estimation of the number of factors in a large N and T panel”, *Journal of Business & Economic Statistics*, **25**, 91–96.

Aruoba, S.B., F.X. Diebold, and C. Scotti (2009), “Real-time measurement of business conditions”, *Journal of Business and Economic Statistics* [in press].

Bai, J. and S. Ng (2002), “Determining the number of factors in approximate factor models”, *Econometrica*, **70**, 191–221.

Bai, J. and S. Ng (2007), “Determining the number of primitive shocks in factor models”, *Journal of Business and Economic Statistics*, **25**, 52–60.

Bai, J. and S. Ng (2008), “Large dimensional factor analysis”, *Foundations and Trends in Econometrics*, **3**, 89–163.

Bekker, P.A. and K.E. Bouwman (2009), “Risk-free interest rates driven by capital market returns”, mimeo, University of Groningen.

Bliss, R. (1997), “Movements in the term structure of interest rates”, *Economic Review Federal Reserve Bank of Atlanta*, **82**, 1633.

Bolder, D. and D. Stréblinski (1999), “Yield curve modelling at the Bank of Canada”, Technical Report 84, Bank of Canada.

Connor, G. and R. Korajczyk (1993), “A test for the number of factors in an approximate factor model”, *The Journal of Finance*, **58**, 1263–1291.

Deistler, M. and E. Hamann (2005), “Identification of factor models for forecasting returns”, *Journal of Financial Econometrics*, **3**, 256–281.

Deistler, M. and C. Zinner (2007), "Modelling high-dimensional time series by generalized linear dynamic factor models: An introductory survey", *Communications in Information and Systems*, **7**, 153–165.

Diebold, F.X. and C. Li (2006), "Forecasting the term structure of government bond yields", *Journal of Econometrics*, **130**, 337–364.

Diebold, F.X., C. Li, and V.Z. Yue (2008), "Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach", *Journal of Econometrics*, **146**, 351–363.

Doz, C., D. Giannone, and L. Reichlin (2007), "A two-step estimator for large approximate dynamic factor models based on Kalman filtering", Discussion Paper 6043, Centre for Economic Policy Research.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The generalized dynamic-factor model: Identification and estimation", *The Review of Economics and Statistics*, **82**, 540–554.

Forni, M. and M. Lippi (2001), "The generalized factor model: Representation theory", *Econometric Theory*, **17**, 1113–1141.

Hallin, M. and R. Liška (2007), "Determining the number of factors in the general dynamic factor model", *Journal of the American Statistical Association*, **102**, 603–617.

Hamann, E., M. Deistler, and W. Scherrer (2005), "Factor models for multivariate time series", in A. Taudes, editor, *Adaptive Information Systems and Modelling in Economics and Management Science*, Springer, Wien, 243 – 251.

Jacobs, J.P.A.M. and P.W. Otter (2008), “Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach”, *Econometric Reviews*, **27**, 385–397.

Jungbacker, B. and S.J. Koopman (2008), “Likelihood-based analysis for dynamic factor models”, mimeo, VU Amsterdam.

Kapetanios, G. (2005), “A testing procedure for determining the number of factors in approximate factor models with large datasets”, Working Paper 551, Queen Mary University of London [forthcoming in the *Journal of Business & Economic Statistics*].

Kapetanios, G. and M. Marcellino (2006), “A parametric estimation method for dynamic factor models of large dimensions”, Discussion Paper 5620, Centre for Economic Policy Research.

Molenaar, P.C.M. (1985), “A dynamic factor model for the analyis of multivariate time series”, *Psychometrika*, **50**, 181–202.

Nelson, C.R. and A.F. Siegel (1987), “Parsimonious modeling of yield curves”, *Journal of Business*, **60**, 473489.

Onatski, A. (2009), “Testing hypotheses about the number of factors in large factor models”, *Econometrica* [forthcoming].

Otter, P.W. (1981), “On statae space modelling, dynamic factor analysis and self-tuning control of (economic) processes”, mimeo, presented at the third International Conference on Mathematical Models, Los Angelos, 1981.

Otter, P.W. (1990), “Canonical correlation in multivariate time series analysis with an application to one-year-ahead and multiyear-ahead macroeconomic forecasting", *Journal of Business and Economic Statistics*, **8**, 453–457.

Otter, P.W. (1991), "On Wiener-Granger causality, information and canonical correlation", *Economics Letters*, **35**, 187–191.

Otter, P.W. and J.P.A.M. Jacobs (2006), "On information in static and dynamic factor models", Working Paper #2006/5, CCSO [presented at the Far Eastern Meeting of the Econometric Society, July 2006, Beijing, China].

Figure 1: U.S. term structure data, period

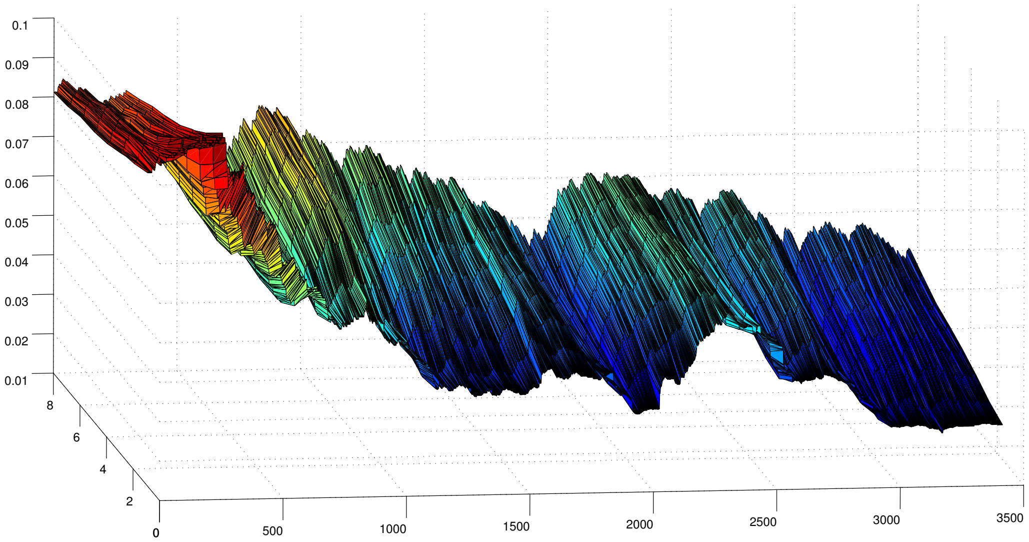

Figure 2: Histograms of prediction errors in interest rates

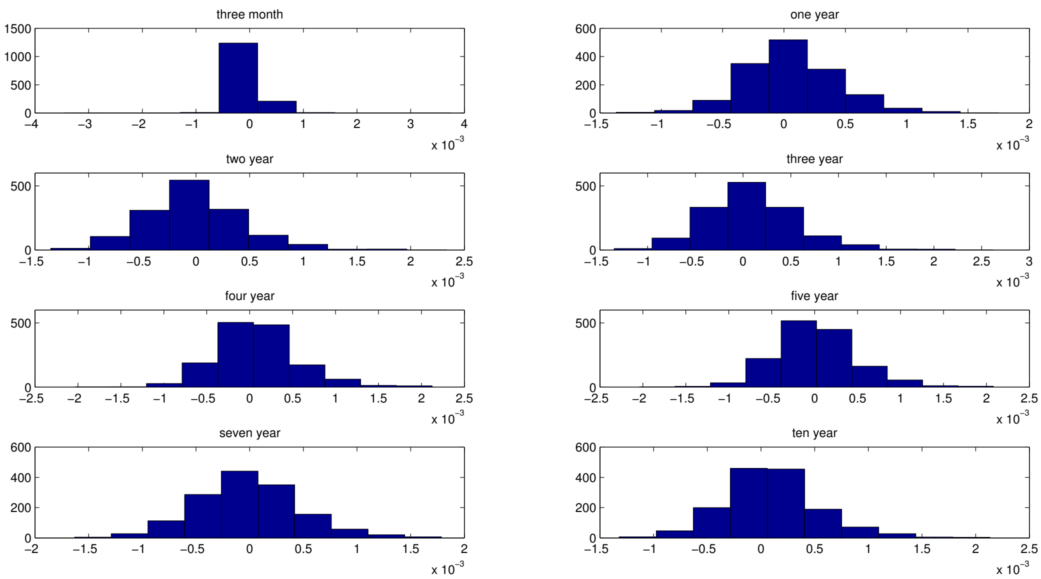

Figure 3: Histogram of estimated factors, $t = 1, \ldots, 2000$

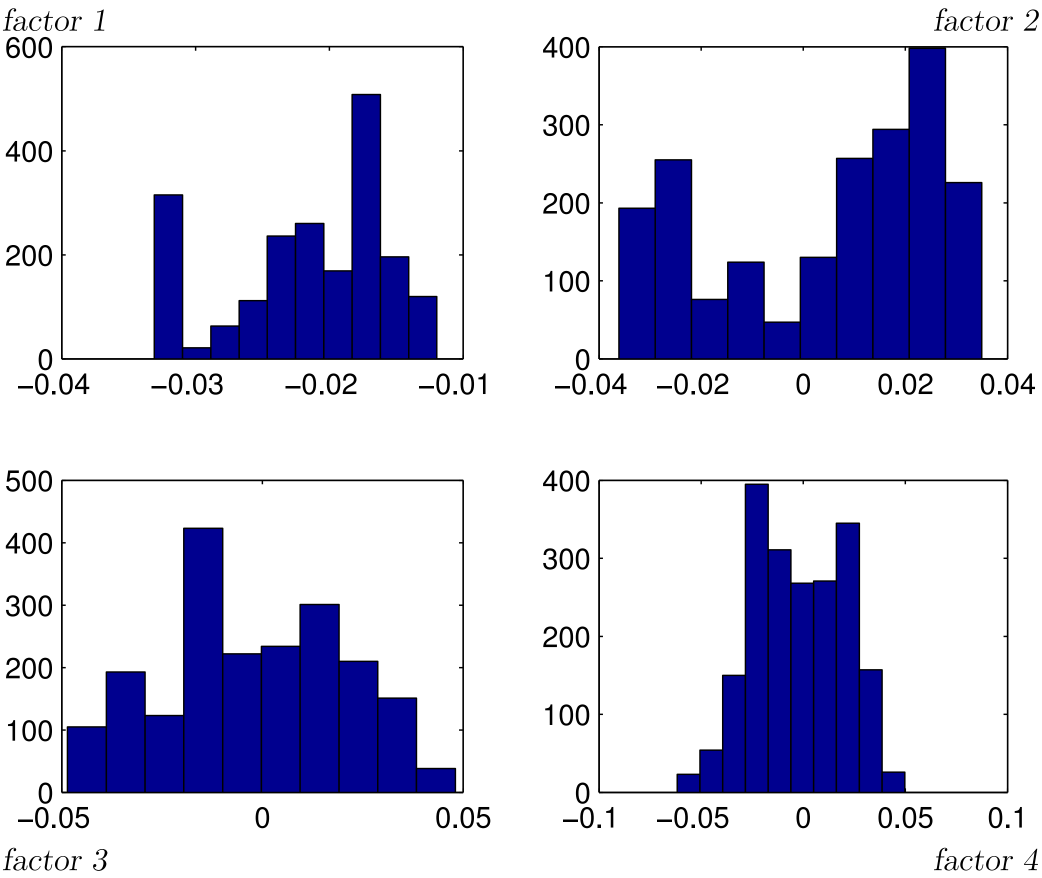

Figure 4: Factor loadings in dynamic factor model with k=4

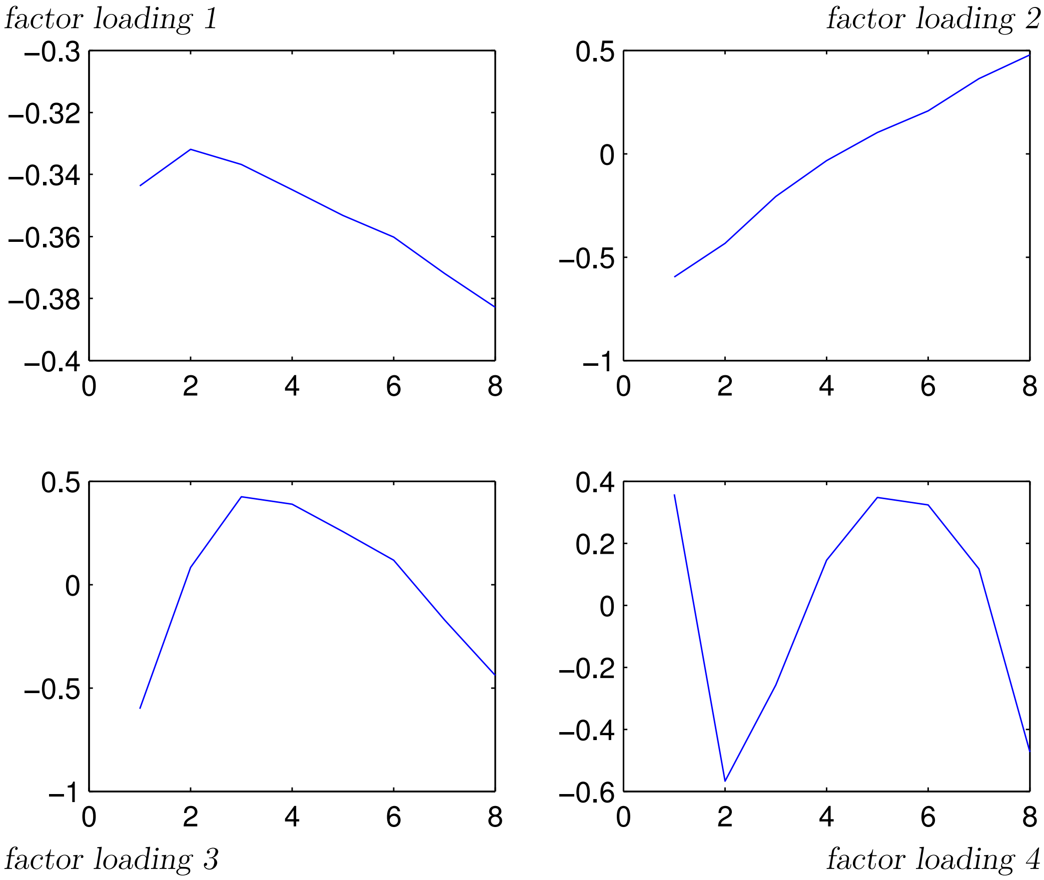

Figure 5: Time series of estimated factors in dynamic factor model with k=4

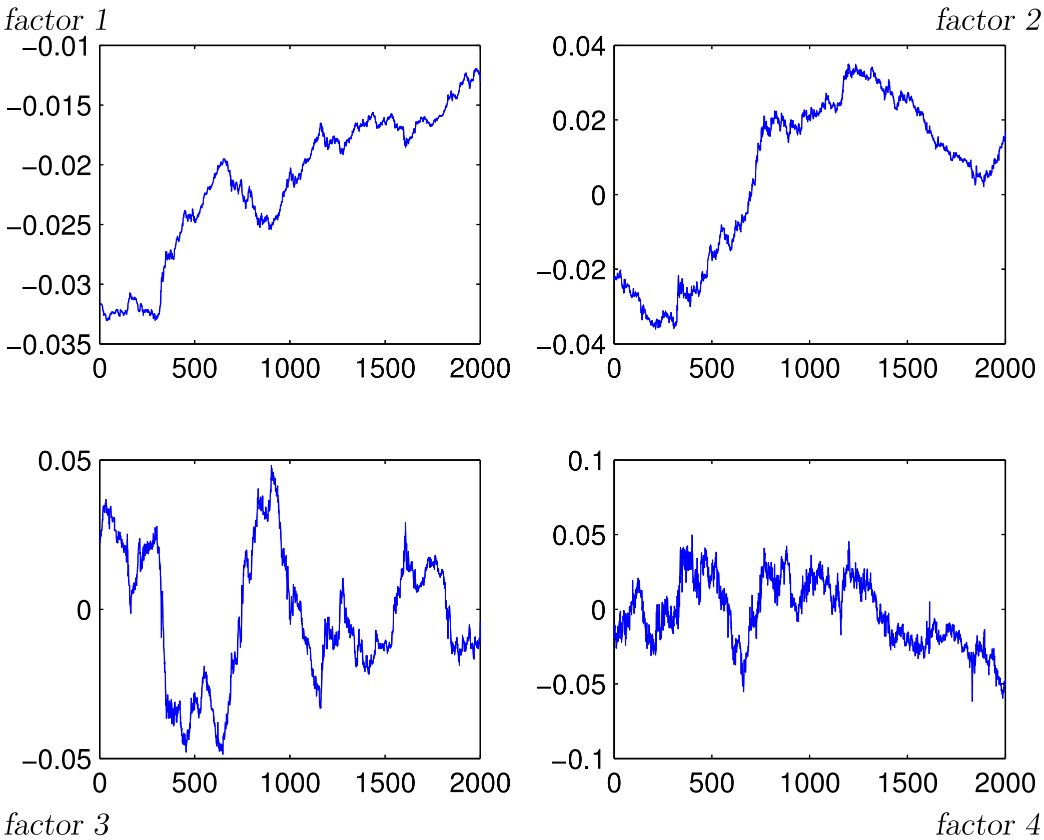

Table 1: U.S. term structure data: descriptive statistics

	mean	st. dev.	min.	max
3-month	0.0447	0.0219	0.0195	0.0988
1-year	0.0444	0.0196	0.0191	0.0947
2-year	0.0463	0.0176	0.0199	0.0916
3-year	0.0484	0.0163	0.0221	0.0888
4-year	0.0504	0.0153	0.0246	0.0878
5-year	0.0520	0.0145	0.0270	0.0869
7-year	0.0547	0.0134	0.0310	0.0852
10-year	0.0573	0.0124	0.0345	0.0845

Table 2: Prediction errors (PEs) for different numbers of factors k

k	1	2	3	4	5	6	7
PE	0.2093	0.0159	0.0035	0.0023	0.0022	0.0022	0.0022