

Forecasting Yield Curves in an Adaptive Framework

Ying Chen,* Bo Li†

Submitted: 4.09.2012, Accepted: 1.11.2012

Abstract

Forecasting yield curves with regime switches is important in academia and financial industry. As the number of interest rate maturities increases, it poses difficulties in estimating parameters due to the curse of dimensionality. To deal with such a feature, factor models have been developed. However, the existing approaches are restrictive and largely based on the stationarity assumption of the factors. This inaccuracy creates non-ignorable financial risks, especially when the market is volatile. In this paper, a new methodology is proposed to adaptively forecast yield curves. Specifically, functional principal component analysis (FPCA) is used to extract factors capable of representing the features of yield curves. The local AR(1) model with time-dependent parameters is used to forecast each factor. Simulation and empirical studies reveal the superiority of this method over its natural competitor, the dynamic Nelson-Siegel (DNS) model. For the yield curves of the U.S. and China, the adaptive method provides more accurate 6- and 12-month ahead forecasts.

Keywords: interest rates, functional principal component analysis, local parametric model, Nelson-Siegel model

JEL Classification: C53, E47, G17.

*National University of Singapore; e-mail: stachen@nus.edu.sg

†National University of Singapore; e-mail: stalibo@nus.edu.sg

Ying Chen, Bo Li

1 Introduction

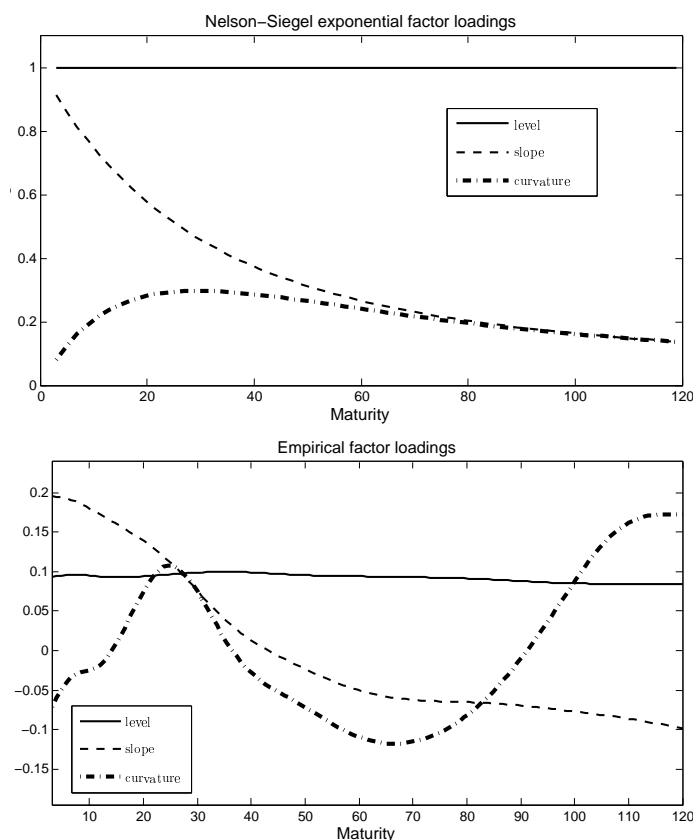
The yield curve depicts interest rates against maturities which is important for not only households, firms and financial institutions when making many economic and financial decisions such as risk management and derivatives pricing, but also for central banks conducting monetary policy for social investment, inflation and unemployment. Although many methods and models predict the interest rate at some particular maturities, such as those of Merton (1973), Vasicek (1977) and Cox, Ingersoll and Ross (1985) as well as Chan, Karolyi, Longstaff and Sanders (1992) and the references therein, accurately forecasting the yield curve remains a challenging task.

The difficulty lies in the curse of dimensionality, for as the number of interest rate maturities increases, accurate estimation of the parameters when some parametric models are applied is difficult, as seen in Härdle, Müller, Sperlich and Werwatz (2004). This inaccuracy creates a non-ignorable risk in applications. Factor models developed by Chen (1996), Schaefer and Schwartz (1984) and Hull and White (1994), among others, address this inaccuracy. By extracting a small number of dominant factors, the problem is converted to a low-dimensional one, see Vetzal (1994). Of all the proposed solutions, the Nelson-Siegel (NS) model (e.g. Nelson and Siegel, 1987; Svensson, 1995) is by far the most popular, obtaining 3 factors based on exponential factor loadings. Figure 1 depicts the factor loadings corresponding to the level, slope and curvature of the yield curve.

Although popular, it is natural to ask whether the NS exponential factor loadings are universally appropriate for any kind of yield curves. Figure 1 also displays 3 empirical factor loadings for the monthly yield curves of China Treasury. These curves are obtained using functional principal component analysis (FPCA), a data-driven method capable of reflecting the empirical evidence of the data. More specifically, the respective factors account for more than 99% of the variations of the raw data and represent well the level, slope and curvature of the yield curves. Section 3.1 details the FPCA method. Interestingly, the shape of the empirical curvature factor loadings, among others, greatly deviates much from the conventional NS curve, with an obvious double-humped shape peaking not only around the medium maturity but also around the long maturity. This possibly is due to unique sovereign credit risks or central bank regulations. The recent development of functional data analysis opens a door to a new approach to obtain factors of yield curves. By considering yield curves as functional data, with each yield curve naturally representing a function of maturities, the FPCA method extracts factors that explain the maximal variation of the curves via orthogonal decomposition. A comprehensive review of theories and applications of FPCA is found in Ramsay and Silverman (2002), Ramsay and Silverman (2005) and Ferraty and Vieu (2006). Compared to the NS model, the FPCA method is appealing as it takes into account the natural functional features of yield curves and extracts factors according to the data's empirical dependence structure. This method also identifies unique features, if they exist, and can be safely used for any kind of yield

 Forecasting Yield Curves in an Adaptive Framework

Figure 1: The empirical factor loadings of the China Treasury yield curves (bottom) and the NS exponential loadings (upper)



In the NS framework: the level loading is 1; the slope loading is $(1 - e^{-\lambda_t \tau})/\lambda_t \tau$ and the curvature loading is $(1 - e^{-\lambda_t \tau})/\lambda_t \tau - e^{-\lambda_t \tau}$, with $\lambda_t = 0.0609$, and τ denoting the time to maturities. Data: monthly yield curves of China treasuries from March 2003 to October 2011, Datastream.

curves. It is worth mentioning that FPCA is different from the multivariate PCA method. The former handles data as curves that correspond to infinite maturities, while the latter considers data as points with multiple but finite maturities, see Müller (2005).

Existing forecasting models are largely based on factor stationarity with a restrictive model set-up. In the widely used Dynamic Nelson-Siegel (DNS) model (Diebold and Li, 2006), the AR(1) specification for each of the three factors is shown to be superior to many competitors, including the random walk model, the slope regression, the Fama-Bliss forward rate regression (Fama and Bliss, 1987), the affine model (see Duffie

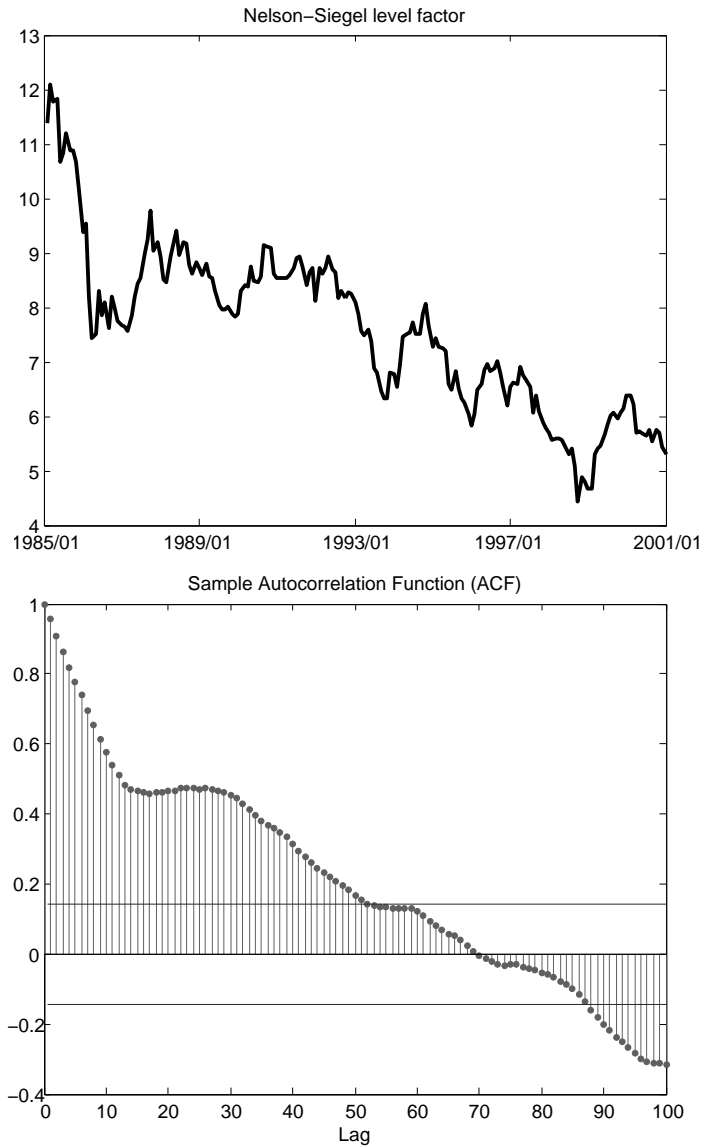
Ying Chen, Bo Li

and Kan, 1996; Egorov, Li and Ng, 2011), the vector autoregressive (VAR) models and the error correction models (Engle and Granger, 1987). The DNS model, however, ignores the structure changes and regime shifts that exist. This non-stationarity, pushed into the factors via the static factor models, is apparent. Figure 2, for example, displays the evolution over time of the resulting NS level factor extracted from data based on the monthly U.S. Treasury yield curves, as in Diebold and Li (2006). The sample autocorrelations are persistent, indicating the existence of non-stationarity. This persistence is addressed in the literature on the non-stationarity of interest rates. Hall, Anderson, Granger (1992) show that interest rates at different maturities are co-integrated and driven by unit-root non-stationary factors. Bansal and Zhou (2002) develop a model where the short rate and the market price of risks are subject to regime shifts. Guidolin and Timmermann (2009) propose a regime-switching VAR model for an aggregated forecast of U.S. short-term interest rates, in which the aggregation weights shift between regimes. The question of the true source of the persistence, however, remains. In fact, Diebold (1986) and Lamoureux and Lastrapes (1990) note that the presence of structural breaks may result in misleading inferences on a long memory diagnosis. The theoretical results in Diebold and Inoue (2001) and Granger and Hyung (2004) further support that this phenomenon can also be spuriously generated by a short-memory model with structural breaks or regime shifts. More generally, Mikosch and St.ari.c.a (2004b) argue, independently of any particular model assumptions, that non-stationarity in the data, such as changes in the unconditional mean or variance, can create long-range dependencies. Such findings led to the development of structural break detection methods (see, for example, Chen and Gupta, 1997; Mikosch and Stărică, 2004a; Liu and Maheu, 2008), time-varying coefficient models using either a Markov-switching approach (see Hamilton and Susmel, 1994; So, Lam and Li, 1998) or using a smooth function of time or other transition variables, (see Baillie and Morana, 2009; Scharth and Medeiros, 2009) as well as local parametric models with a focus on volatility forecasting (see, for example, Čížek, Härdle and Spokoiny, 2009). Chen, Härdle and Pigorsch (2010) propose a local autoregressive (LAR) model, where the time-dependent parameters are estimated under the assumption of local homogeneity. Although a stationary model such as AR(1) fails to handle persistence, the LAR model is universally suitable for both stationary and non-stationary time series. In the application of volatility forecasting for the S&P500 index futures, the local model delivers superior forecasting accuracy to several alternative models including long memory models and some regime-switching models.

We propose a new methodology to adaptively forecast yield curves. First, the FPCA method is used to extract dominant factors to represent the empirical features of yield curves. Then the LAR(1) model is employed to forecast each of the resulting factors, where the time-dependent parameters are estimated over an interval of local homogeneity. The main contribution of our study is not developing a new estimation method or exhibiting related theoretical results; it is to offer a data-driven

Forecasting Yield Curves in an Adaptive Framework

Figure 2: The Nelson-Siegel level factor (upper) and its sample ACF plot (bottom)



Data: monthly U.S. Treasury yield curves from January 1985 to December 2000, see also Diebold and Li(2006).

technology that automatically selects a trustable stationary time interval to improve the accuracy of yield curve forecasting. The application of this local method to the

Ying Chen, Bo Li

prediction of the U.S. and China yield curves is compared to its natural competitor, the DNS model. A simulation study and empirical analysis demonstrate reasonable performance, particularly for forecasting 6- and 12-months ahead yield curves.

A dynamic semiparametric factor model (DSFM) developed by Fengler, Härdle and Mammen (2007) extracts dominant factors of functional data such as the volatility surface via a weighted least squares approach. The DSFM is used in the study of electricity forward curve dynamics, risk-neutral density estimation and the dynamics of hourly electricity prices, see Borak and Weron (2008), Giacomini, Härdle and Krätschmer (2009) and Härdle and Trück (2010), respectively. The semiparametric method represents the evolution of data in a local interval controlled by smoothing parameter(s). By contrast, our method addresses factor extraction and non-stationarity separately. Using time-independent basis functions, it pushes all the dynamics into the time series of the factors, simplifying computations.

The remainder of the paper is structured as follows. Section 2 describes the data. Section 3 presents the proposed forecasting model, detailing the factor extraction by FPCA and the LAR parameter estimation. Section 4 tests the performance of this method in a practically oriented simulation study. Section 5 reports the real data analysis and forecast comparison with the alternative model. Section 6 concludes.

2 Data

Two data sets are considered, the U.S. Treasury and the China Treasury. The U.S. data consist of the end-of-month price quotes (bid-ask average) for U.S. Treasury, from January 1985 to December 2000. There are 192 monthly interest rates at 17 maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. The data is the same as in Diebold and Li (2006). The second set of data are the end-of-month price quotes for China Treasury from March 2003 to October 2011. There are 104 monthly interest rates at 11 maturities of 3, 6, 12, 24, 36, 48, 60, 84, 96, 108 and 120 months.

The underlying yield curve, as a function of maturities, is not directly observable. Parametric and nonparametric methods are available to estimate the yield curve based on the observed discrete interest rates, among which nonparametric methods often provide a better fit. Polynomial splines are the most popular nonparametric technique used to estimate yield curves, see McCulloch (1971), Schaefer (1973), McCulloch (1975), Vasicek and Fong (1982) and Shea (1985). The methods are sensitive to the selection of smoothing parameters such as knots in splines. For the selection issues, we refer to Jarrow, Ruppert and Yu (2004) and Fernández-Rodríguez (2006), among many others.

In our study, we use the B-splines smoothing technique to obtain the functional data. Let $\mathbf{x}_t(\tau_1, \dots, \tau_q)$ be the discrete interest rates at time t , $t = 1, \dots, T$, that contains q maturities. Use $x_t(\tau)$ to denote the yield curve, a function of maturity $\tau \in \mathbb{R}$. The

 Forecasting Yield Curves in an Adaptive Framework

yield curves are estimated as

$$x_t(\tau) = \sum_{k=1}^K c_{tk} \phi_k(\tau) = c_{t1} \phi_1(\tau) + \cdots + c_{tK} \phi_K(\tau),$$

where $\phi_1(\tau), \dots, \phi_K(\tau)$ are K basis functions. We refer to Ramsay and Silverman (2005) to justify the selection of the number of basis functions. The coefficients c_{t1}, \dots, c_{tK} are estimated by minimizing the penalized sum-of-squared errors

$$\text{PENSSE}_\lambda = \|\mathbf{x}_t(\tau_1, \dots, \tau_q) - x_t(\Lambda)\|^2 + \lambda \int [D^2 x_t(\tau)]^2 d\tau,$$

where $x_t(\Lambda)$ contains the function values of the smoothed curve at the discrete maturities $\Lambda = (\tau_1, \dots, \tau_q)$, $D^2 x_t(\tau)$ is the second derivative function of $x_t(\tau)$ and parameter λ is a smoothing parameter that controls the smoothness of the estimated curve \tilde{x}_t . We denote the L^2 norm by $\|\cdot\|$. In Figure 3 and Figure 4, the smoothed yield curves are displayed in both 3D and 2D views. The 3D plots capture the movement of the yield curves as time goes by. At the same time, we have a clear view of the structure and shape of the smoothed curves from the 2D plots. We calculate the percentage errors across all maturities. By assessing the fitted percentage errors, we easily see that the smoothed curves using B-splines serve as a reasonable representation of the underlying yield curves because the average percentage errors for all curves are 1.5309×10^{-5} and 6.5702×10^{-6} for the U.S. and China data sets, respectively.

The smoothed curves are the functional data considered in the following sections.

3 Method

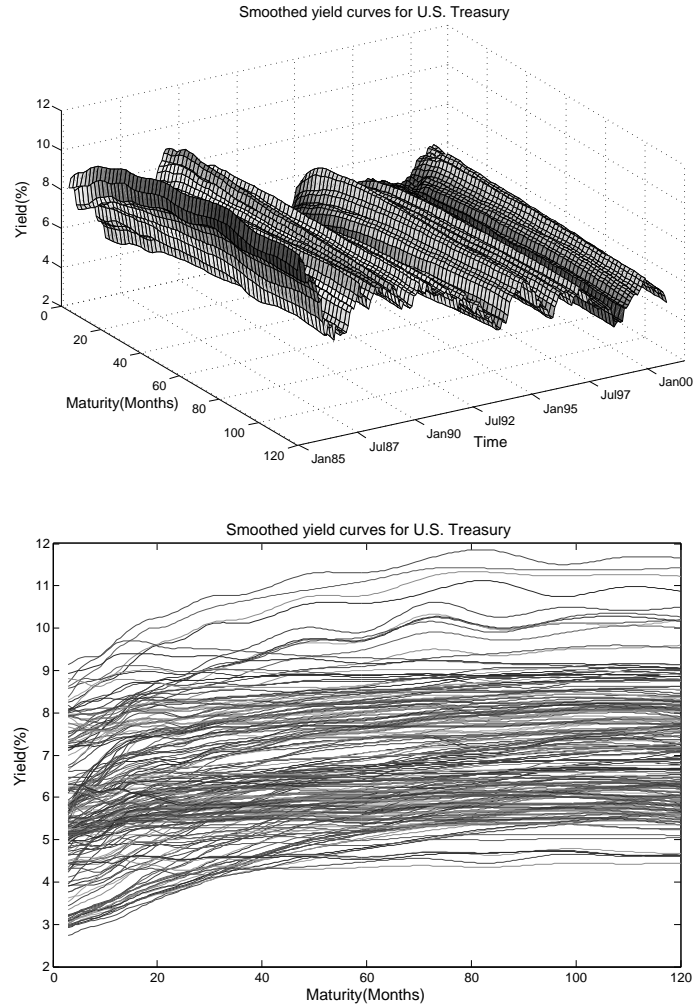
The FPCA method projects yield curves into the directions with the first few largest variations, along which p -dimensional factors are extracted. In fact, the empirical loadings in Figure 1 are obtained using FPCA. Using a linear transformation, any form of non-stationarity in the yield curves is solely attributed to the time series of the resulting factors. A time-varying autoregressive model is used to model and forecast each of the factors, where the parameters are respectively estimated under a local homogeneity assumption. That is, for any particular time point, there exists a past time interval over which the data is represented well by an AR process with constant parameters. It is analogous to the rolling window technique (with fixed window size), though in the local model, the window size changes over time t . The time intervals are identified in a data-driven way.

3.1 Extracting factors via FPCA

The yield curve is denoted as $x_t(\tau)$ at time $t \in [1, T]$, which is a function of time to maturities $\tau \in \mathbb{R}$. Without loss of generality, the yield curves are assumed to be

Ying Chen, Bo Li

Figure 3: Smoothed yield curves via B-splines in 3D(upper) and 2D (bottom) for U.S. Treasury with 15 maturities: 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, from January 1985 to December 2010. The 3D plot captures the movement of the yield curves as time goes by, while the 2D plot provides an overall view of the structure and shape of the smoothed curves.

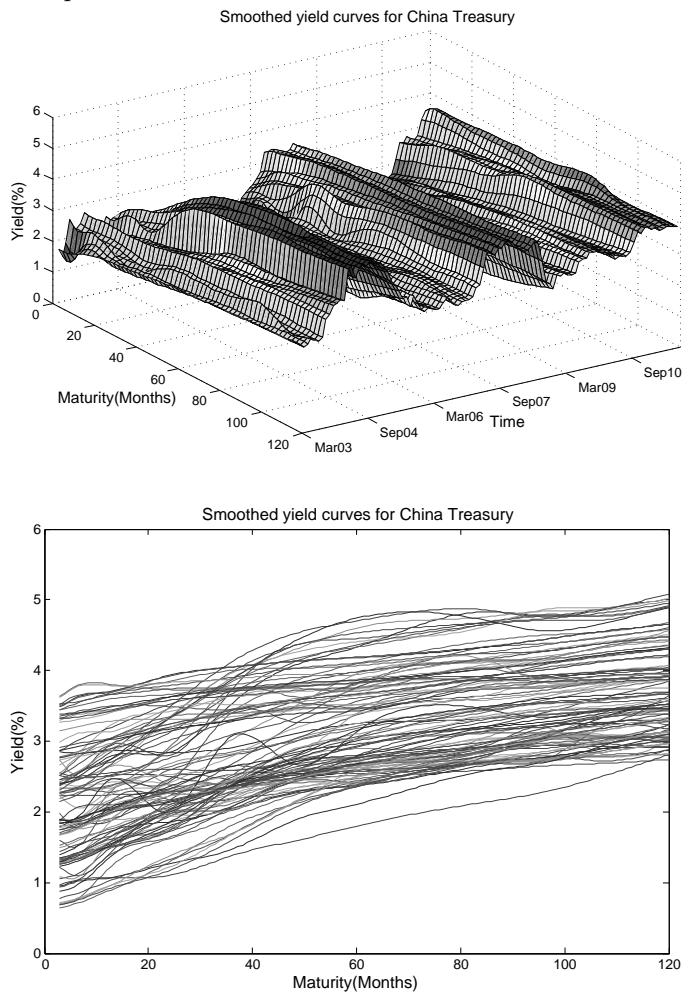


demeaned, i.e., $E[x_t(\tau)] = 0$. We apply FPCA to extract factors, giving

$$f_t = \int \xi(\tau)x_t(\tau)d\tau, \quad (1)$$

 Forecasting Yield Curves in an Adaptive Framework

Figure 4: Smoothed yield curves via B-splines in 3D(upper) and 2D (bottom) for China Treasury with 11 maturities of 3, 6, 12, 24, 36, 48, 60,84, 96, 108 and 120 months, from March 2003 to December 2011. The 3D plot captures the movement of the yield curves as time goes by, while the 2D plot provides an overall view of the structure and shape of the smoothed curves.



where f_t denotes the factor and ξ the corresponding factor loadings. The solution maximizes the variation of the respective factors f_t such that

$$\max \frac{1}{T} \sum_t f_t^2 = \frac{1}{T} \sum_t \left\{ \int \xi(\tau) x_t(\tau) d\tau \right\}^2, \quad \text{subject to } \int \xi^2(\tau) d\tau = 1.$$

Ying Chen, Bo Li

Here, the factor loadings follow the unit-norm condition to guarantee a unique solution. To obtain the other factors, the factor loadings are assumed to be orthogonal with $\int \xi_k(\tau)\xi_m(\tau)d\tau = 0$, for all $k \neq m$ and $k, m \leq p$, where p is the number of selected factors.

The factor loadings are estimated through solving an eigen-decomposition. Define the covariance of yield curves as

$$v(\tau, s) = \frac{1}{T} \sum_t x_t(\tau)x_t(s),$$

where $\tau \in \mathbb{R}$ and $s \in \mathbb{R}$. Figure 5 displays the sample covariance surfaces of the U.S. and China Treasury yield curves respectively. Clearly, the covariance is larger when maturities are closer together, and decays as maturities move farther apart. Factor loading $\xi(\tau)$ is actually an eigenfunction of the covariance

$$\int v(\tau, s)\xi(\tau)d\tau = \rho\xi(s), \quad (2)$$

where ρ denotes the eigenvalue.

Ramsay and Silverman (2005) propose a solution for Equation (2). Express the functional data $\mathbf{x}(\tau)$ with K basis functions $\Phi(\tau) = [\phi_1(\tau), \dots, \phi_K(\tau)]^\top$ as

$$\mathbf{x}(\tau) = \begin{bmatrix} x_1(\tau) \\ x_2(\tau) \\ \vdots \\ x_T(\tau) \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^K c_{1k}\phi_k(\tau) \\ \sum_{k=1}^K c_{2k}\phi_k(\tau) \\ \vdots \\ \sum_{k=1}^K c_{Tk}\phi_k(\tau) \end{bmatrix} = \mathbf{C}\Phi(\tau), \quad (3)$$

where \mathbf{C} is a $T \times K$ matrix of coefficients. Formulate the weight function $\xi(\tau)$ in a basis expansion with the same basis functions, but different coefficients, $\mathbf{b} = (b_1, \dots, b_K)^\top$, such that

$$\xi(\tau) = \Phi^\top(\tau)\mathbf{b}. \quad (4)$$

Substituting the expansions, Equation (3) and Equation (4) into Equation (2), we have

$$\frac{1}{T}\Phi^\top(s)\mathbf{C}^\top\mathbf{C}\int\Phi(\tau)\Phi^\top(\tau)d\tau\mathbf{b} = \rho\Phi^\top(s)\mathbf{b}.$$

Note that this equation applies for all values of s so we can drop $\Phi^\top(s)$. Defining a $K \times K$ matrix $\mathbf{W} = \int\Phi(\tau)\Phi^\top(\tau)d\tau$ and a vector $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{b}$, we have

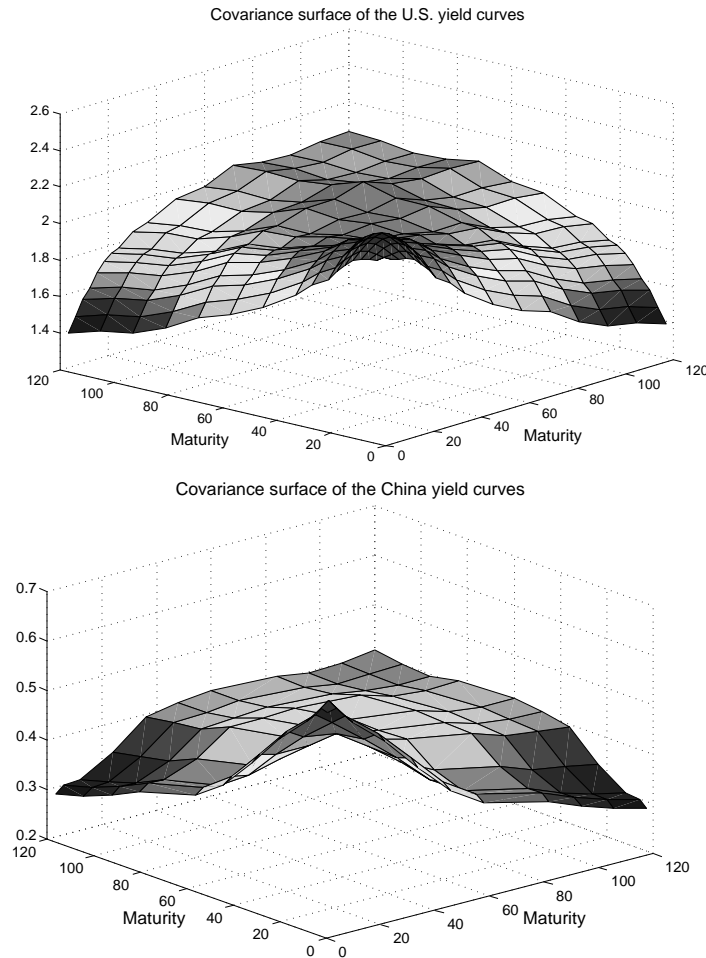
$$\frac{1}{T}\mathbf{W}^{1/2}\mathbf{C}^\top\mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \rho\mathbf{u},$$

and $\mathbf{b}_m^\top\mathbf{W}\mathbf{b}_m = 1$, $\mathbf{b}_k^\top\mathbf{W}\mathbf{b}_m = 0$ for $k \neq m$, where \mathbf{b}_m corresponds to the coefficient vector of the m th eigenfunction $\xi_m(\tau)$. The eigenvalues ρ and eigenvectors \mathbf{u} are solvable for the $K \times K$ matrix $\frac{1}{T}\mathbf{W}^{1/2}\mathbf{C}^\top\mathbf{C}\mathbf{W}^{1/2}$. Consequently, we have

$$\xi(\tau) = \Phi^\top(\tau)\mathbf{b}, \quad \mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}.$$

Forecasting Yield Curves in an Adaptive Framework

Figure 5: The sample covariance surfaces of the yield curves of U.S. Treasury from January 1985 to December 2000 (upper) and of China Treasury from March 2003 to October 2011 (bottom).



How many factors are needed to capture the dynamics of yield curves? In other words, how do you select the number of factors, p . Litterman and Scheinkman (1991) show that three factors are necessary for U.S. yield curves, with the factors representing the level, slope and curvature, while Cochrane and Piazzesi (2005) and Dai, Singleton and Yang (2004) show that up to five factors should be considered for U.S. government bonds. Besides, Egorov *et al.* (2011) find four factors (two common and two local) for the joint yield curves of U.S. and euro interest rates. In addition to these qualitative selection, FPCA provides a natural quantitative selection criterion.

Ying Chen, Bo Li

According to Equation (2), the i th largest eigenvalue, ρ_i , associates with the variation of yield curves defined by $\xi_i(\tau)$, $i = 1, \dots, K$. Therefore, the cumulative proportion of variation explained by all the p factors is

$$\frac{\rho_1 + \dots + \rho_p}{\rho_1 + \dots + \rho_K}.$$

In our study, we select $p = 3$ factors. This selection is motivated by economic meanings. However, more factors may be included should the cumulated sum of the explained variation fall below 99%.

3.2 Fitting a local AR model to the factors

Since the model structure and estimation procedure are the same for all the selected p factors, for notational simplicity we drop the subscript j of the factors f_{jt} with $j = 1, \dots, p$. Therefore, without loss of generality, $\{f_t\}$ now stands for the univariate time series of any of the p factors. For each $\{f_t\}$ taking values in \mathbb{R} , the LAR model of order 1, or LAR(1) model, is defined through a time-varying parameter set $\theta_t = (\theta_{0t}, \theta_{1t}, \sigma_t)^\top$:

$$f_t = \theta_{0t} + \theta_{1t}f_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_t^2).$$

The estimation of the parameters is conducted under the assumption of local homogeneity at each time t . Under local homogeneity, parameter θ is assumed to not deviate substantially from constant in a local interval $I_t = [t - m_t, t)$, and hence the data are (approximately) stationary. The local maximum likelihood estimator $\hat{\theta}_t$ is defined over the local interval:

$$\begin{aligned} \hat{\theta}_t &= \operatorname{argmax}_{\theta \in \Theta} L(I_t, \theta) \\ &= \operatorname{argmax} \left\{ -m_t \log \sigma_t - \frac{1}{2\sigma_t^2} \sum_{s=t-m_t}^{t-1} (f_s - \theta_0 - \theta_1 f_{s-1})^2 \right\}, \end{aligned}$$

where Θ denotes the parameter space and $L(I_t, \theta)$ is the local log-likelihood function. The local estimation method is different from the rolling window technique with a globally constant window size. Here the local window size m_t is time-dependent and, in practice, unknown. The question is, of course, how to select the local interval or the value of m_t . Generally speaking, the optimal selection would be the longest interval where the local homogeneity assumption holds; that is, the time series of the factor can be described well by a model with constant parameters. Under that assumption, long intervals provide accurate estimators with low variation. However, as the interval length further increases, the local homogeneity assumption is likely to be violated. This violation possibly leads to a large modeling bias. Therefore, the optimal selection should be designed to balance the modeling bias and variance tradeoff.

We employ an automatic procedure to select the intervals. At every time point, we

 Forecasting Yield Curves in an Adaptive Framework

start with a small sample size m_0 that defines a local interval $I_t^{(0)} = [t - m_0, t)$. The value of m_0 is small enough to ensure homogeneity, and where a conventional AR(1) model has a reasonable fit. Iteratively, we increase the sample size to m_k , with $k > 0$ and $m_k > m_{k-1}$, which defines a longer interval $I_t^{(k)} = [t - m_k, t)$. Note that at this moment, $I_t^{(k-1)} = [t - m_{k-1}, t)$ has been accepted as an interval of homogeneity. We then check for any deviation from homogeneity with a likelihood ratio:

$$\begin{aligned} T_t^{(k)} &= L(I_t^{(k)}, \hat{\theta}_t^{(k-1)}) - L(I_t^{(k)}, \hat{\theta}_t^{(k)}), \\ \text{where } \hat{\theta}_t^{(k-1)} &= \operatorname{argmax}_{\theta \in \Theta} L(I_t^{(k-1)}, \theta), \\ \hat{\theta}_t^{(k)} &= \operatorname{argmax}_{\theta \in \Theta} L(I_t^{(k)}, \theta). \end{aligned}$$

Note that $\hat{\theta}_t^{(k-1)}$ is an accepted estimate under local homogeneity in the interval $I_t^{(k-1)}$. If the difference is small, indicating that the larger data sample displays similar patterns to the smaller sample, then we accept the longer interval for an improved accuracy of estimation. On the contrary, if the difference is large, it implies modeling changes. We terminate the procedure to avoid substantial modeling bias. The lastly-accepted interval would be the optimal choice. We continue this way until either a change is suspected or the possibly longest interval is screened. Chen *et al.* (2010) have a test to measure the significance level of the difference. We refer readers to their work for further details.

4 Simulation

The proposed forecasting method involves two parts, extracting factors via FPCA and forecasting in an LAR framework. In this section we investigate the performances of the FPCA method in a practical simulation. For the performances of the time-varying coefficient model, we refer to the work of Chen, Hardle, Pigorsch (2010). More specifically, we study the descriptibility of the FPCA method in two scenarios: 1) The U.S. scenario where the yield curves are driven by NS exponential factor loadings, and 2) The China scenario where the empirical loadings (obtained by FPCA) are the true data generating process. Both the FPCA and DNS methods are applied. The root mean square error (RMSE) is used to assess the estimation accuracy for the estimated yield curves.

In the U.S. scenario, yield curves are generated based on the monthly U.S. Treasury from January 1985 to December 2000, following a DNS modeling framework:

$$x_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) + \varepsilon_t(\tau), \quad (5)$$

where $\lambda_t = 0.0609$ that maximizes the curvature loading at a medium maturity of 30 months, see Diebold and Li (2006). The factors β_{jt} with $j = 1, 2, 3$ are the ordinary

Ying Chen, Bo Li

least squares (OLS) estimates and represent the level, slope and curvature of the U.S. yield curves. In the simulation, we generate yield curves containing 17 yields at maturities of 3 months to 10 years for $t = 1, \dots, 192$. The values of the factors are considered to be known and the stochastic innovations $\varepsilon_t(\tau)$ are i.i.d. normal random variables, with mean and standard deviations as calculated in Equation (5). We repeat the process 500 times.

Analogously, the China scenario is designed according to the yield curves of China Treasury from March 2003 to October 2011. The data consist of 104 monthly observations with 11 maturities from 3 months to 10 years. The demeaned yield curves are estimated via FPCA as

$$x_t(\tau) = f_{1t}\xi_1(\tau) + f_{2t}\xi_2(\tau) + f_{3t}\xi_3(\tau) + \varepsilon_t(\tau),$$

where f_{jt} refers to the resulting factor and $\xi_j(\tau)$ refers to the empirical factor loadings, with $j = 1, 2, 3$. Again, the error term $\varepsilon_t(\tau)$ is assumed to be normally distributed. The empirical factor loadings are depicted in Figure 1. In the simulation, we generate 104 curves, each including 11 interest rates. The process is also repeated 500 times. Both the FPCA and DNS methods are used to extract factors. In the DNS modeling, 3 factors are extracted. In the FPCA, we also select 3 dominant factors, as they explain more than 99% of the total variation for each case. The average values of the accumulated proportion of the variance are 94.2%, 99.7% and 99.9% for the U.S. scenario and 89.3%, 98.8% and 99.3% for the China scenario. The DNS exponential factor loadings are fixed yet fail to represent the underlying structure of the data in the China scenario. The FPCA loadings differ among different data sets. This data-driven method adapts according to the actual dependence structure of the data. In the U.S. scenario, the factor loadings replicate the exponential curves well, though with different magnitudes. The scaling deviation is due to the demeaning process and it has no impact on the yield curve forecasts as the mean process is added back. In the China scenario, the empirical factor loadings are good proxies of the underlying curves. As an illustration, we depict the FPCA factor loadings of one randomly selected process for each of the scenarios in Figure 6. The results for the other generated data processes behave in a similar way, and for space considerations are omitted here.

The RMSEs of the fitted yield curves are computed. The smaller the value, the higher the accuracy. Table 1 reports the average value of the RMSEs of the discrete interest rates at various maturities. It reveals that the FPCA method is, indeed, superior to the DNS model for most maturities. In the China scenario, the FPCA works well for maturities between 1 and 10 years. Even in the U.S. scenario, the FPCA performs better for 15 of the 17 maturities, with the 3-month and 10-year maturities the exceptions. In relative terms, the FPCA method improves estimation accuracy by approximately 16% in the U.S. scenario and 29% in the China scenario. There is relatively poor forecasting accuracy for both the shortest and the longest maturities (3 months and 120 months) due to the boundary effect of smoothing,

 Forecasting Yield Curves in an Adaptive Framework

a consequence of converting the original, but discrete, high dimensional data to functional data. Without sufficient observations at the boundaries, we have low accuracy in estimation at the boundary maturities. However, the results show that overall FPCA performs better in both scenarios. It not only improves accuracy but also captures the underlying pattern of data, whereas the alternative method encounters mis-specification in the China scenario.

Table 1: Simulation results

| Maturity | U.S. scenario | | China scenario | |
|----------|---------------|-------------|----------------|-------------|
| | DNS | FPCA | DNS | FPCA |
| 3 | 0.05 | 0.08 | 0.08 | 0.12 |
| 6 | 0.05 | 0.04 | 0.09 | 0.10 |
| 9 | 0.06 | 0.05 | – | – |
| 12 | 0.07 | 0.07 | 0.09 | 0.06 |
| 15 | 0.08 | 0.05 | – | – |
| 18 | 0.06 | 0.04 | – | – |
| 21 | 0.04 | 0.04 | – | – |
| 24 | 0.05 | 0.04 | 0.11 | 0.06 |
| 30 | 0.04 | 0.04 | – | – |
| 36 | 0.06 | 0.04 | 0.08 | 0.08 |
| 48 | 0.06 | 0.05 | 0.09 | 0.08 |
| 60 | 0.08 | 0.05 | 0.10 | 0.03 |
| 72 | 0.07 | 0.07 | – | – |
| 84 | 0.06 | 0.05 | 0.07 | 0.04 |
| 96 | 0.05 | 0.04 | 0.06 | 0.06 |
| 108 | 0.06 | 0.04 | 0.06 | 0.03 |
| 120 | 0.06 | 0.07 | 0.08 | 0.06 |

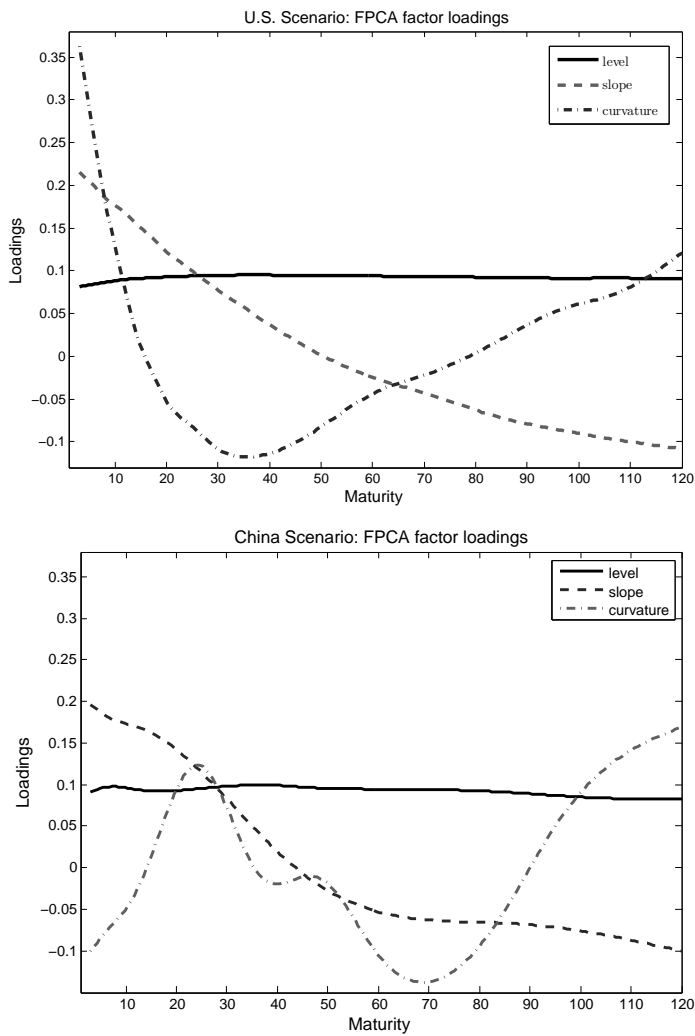
The average value of RMSEs for the fitted and actual interest rates are reported for the two scenarios. Both the DNS and FPCA methods are used. The results with smaller errors are marked in bold to highlight better accuracy.

5 Real data analysis

In this section, we implement the FPCA-LAR method to forecast the yield curves of the U.S. Treasury and China Treasury. We investigate the performance of the adaptive method relative to its natural competitor, the DNS model with an AR(1) specification (Diebold and Li, 2006). Does the proposed forecasting method benefit from using data-driven factor extraction and adaptive modeling? We assess the out-of-sample forecasting accuracy of both methods to determine an answer. To forecast yield curves using FPCA-LAR, at each time t , we apply FPCA to extract factors and fit an LAR(1) model over the selected optimal intervals. The h -step ahead forecasts

Ying Chen, Bo Li

Figure 6: One realization of the FPCA factor loadings for both simulation scenarios



In the U.S. scenario, the factor loadings well represent the underlying NS exponential curves (upper). In the China scenario, the factor loadings are good proxies of the underlying curves displayed in Fig 1.

of the factors as well as the yield curves are

$$\hat{x}_{t+h}(\tau) = \sum_{j=1}^p \hat{f}_{j,t+h} \xi_j(\tau),$$

where $\hat{f}_{j,t+h} = \hat{\theta}_{0jt} + \hat{\theta}_{1jt} f_{j,t}$.

 Forecasting Yield Curves in an Adaptive Framework

To forecast using DNS, we first obtain the NS factors by Equation (5), for which the parameters of the AR(1) model are estimated by using all the data available. The h -step ahead forecasts are thus

$$\hat{x}_{t+h}(\tau) = \hat{\beta}_{1,t+h} + \hat{\beta}_{2,t+h} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \hat{\beta}_{3,t+h} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right),$$

where $\hat{\beta}_{j,t+h} = \hat{\theta}_{0jt} + \hat{\theta}_{1jt} \beta_{j,t}$. We compute the 1-, 6- and 12-month ahead forecasts

Table 2: RMSFE: The average values of the out-of-sample forecast errors for the forecasting horizons, h , of 1-, 6- and 12-month ahead for various maturities, τ , from 3 months to 10 years.

| h | U.S. yield curves | | | | | | China yield curves | | | | | |
|--------|-------------------|------|---------|-------------|----------|-------------|--------------------|-------------|---------|-------------|----------|-------------|
| | 1-month | | 6-month | | 12-month | | 1-month | | 6-month | | 12-month | |
| τ | DNS | F-L | DNS | F-L | DNS | F-L | DNS | F-L | DNS | F-L | DNS | F-L |
| 3 | 0.17 | 0.24 | 0.56 | 0.38 | 0.87 | 0.48 | 0.29 | 0.31 | 0.99 | 0.74 | 1.11 | 0.68 |
| 6 | 0.20 | 0.23 | 0.61 | 0.44 | 0.86 | 0.54 | 0.28 | 0.30 | 0.99 | 0.75 | 1.11 | 0.69 |
| 9 | 0.22 | 0.24 | 0.64 | 0.49 | 0.86 | 0.57 | - | - | - | - | - | - |
| 12 | 0.24 | 0.27 | 0.67 | 0.53 | 0.87 | 0.59 | 0.28 | 0.28 | 0.95 | 0.74 | 1.06 | 0.69 |
| 15 | 0.25 | 0.30 | 0.69 | 0.55 | 0.88 | 0.59 | - | - | - | - | - | - |
| 18 | 0.26 | 0.31 | 0.71 | 0.57 | 0.90 | 0.60 | - | - | - | - | - | - |
| 21 | 0.27 | 0.33 | 0.73 | 0.59 | 0.93 | 0.61 | - | - | - | - | - | - |
| 24 | 0.28 | 0.34 | 0.76 | 0.60 | 0.97 | 0.62 | 0.28 | 0.25 | 0.89 | 0.75 | 0.96 | 0.68 |
| 30 | 0.28 | 0.38 | 0.76 | 0.61 | 1.00 | 0.62 | - | - | - | - | - | - |
| 36 | 0.28 | 0.40 | 0.77 | 0.62 | 1.03 | 0.63 | 0.25 | 0.24 | 0.78 | 0.67 | 0.81 | 0.58 |
| 48 | 0.28 | 0.47 | 0.77 | 0.65 | 1.08 | 0.66 | 0.23 | 0.25 | 0.67 | 0.60 | 0.69 | 0.51 |
| 60 | 0.29 | 0.52 | 0.81 | 0.69 | 1.16 | 0.72 | 0.25 | 0.28 | 0.63 | 0.58 | 0.63 | 0.51 |
| 72 | 0.27 | 0.56 | 0.78 | 0.71 | 1.16 | 0.74 | - | - | - | - | - | - |
| 84 | 0.27 | 0.56 | 0.77 | 0.71 | 1.17 | 0.75 | 0.23 | 0.29 | 0.57 | 0.56 | 0.57 | 0.48 |
| 96 | 0.26 | 0.59 | 0.74 | 0.73 | 1.15 | 0.78 | 0.21 | 0.29 | 0.54 | 0.55 | 0.53 | 0.45 |
| 108 | 0.25 | 0.61 | 0.74 | 0.75 | 1.17 | 0.81 | 0.19 | 0.28 | 0.50 | 0.52 | 0.51 | 0.41 |
| 120 | 0.26 | 0.62 | 0.76 | 0.75 | 1.22 | 0.82 | 0.19 | 0.30 | 0.46 | 0.49 | 0.47 | 0.39 |

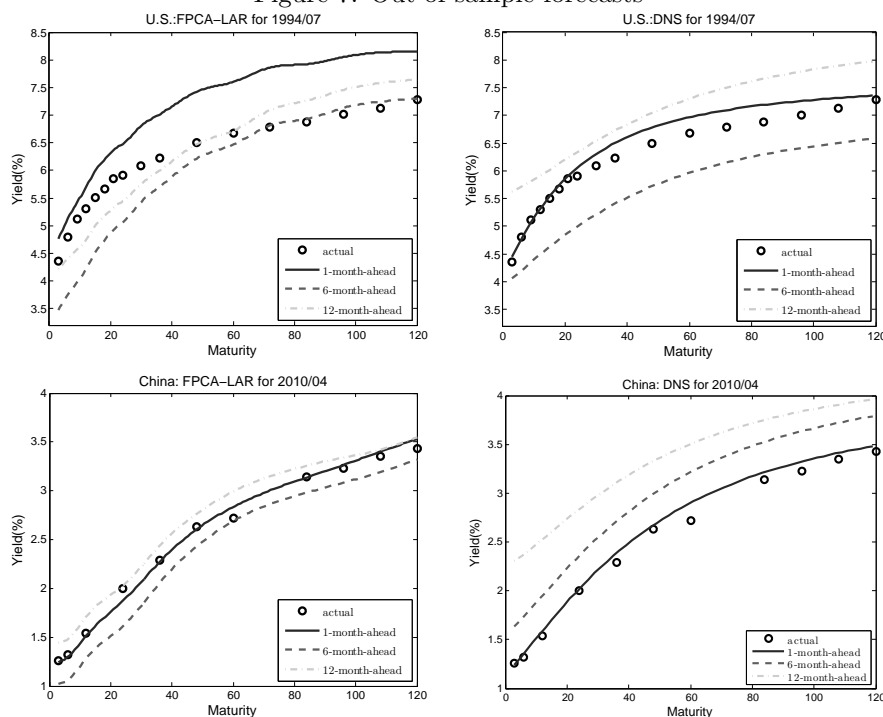
The DNS model and the FPCA-LAR (F-L) method are employed.

for each of the U.S. curves from January 1994 to December 2000 and for China from January 2008 to October 2011. We then perform recursive forecasting with an extending window always starting from the first month. Take the U.S. data as an example. We use the yield curves from January 1985 to December 1993 to obtain the 1-month ahead forecast for January 1994, and the data from January 1985 to July 1993 to get the 6-month ahead forecast for January 1994, and so on. We move forward one period at a time and repeat the forecast until reaching the end of the sample. Therefore, in order to obtain the 1-month (6-month) ahead forecast for February 1994, data curves from January 1985 to January 1994 (January 1985 to August 1993) are utilized. In Figure 7 we depict the 1-, 6- and 12-month ahead out-of-sample

Ying Chen, Bo Li

forecasts against some maturities in July 1994 (U.S.) and in April 2010 (China). For these two months, the yield curves display typical shapes in the market, e.g., a double-humped shape in the China market. This illustrates that for both data sets the FPCA-LAR method performs well for 6- and 12-month ahead forecasts, whereas the DNS model provides accurate results only in the immediate forecast horizon. To

Figure 7: Out-of-sample forecasts



The actual discrete interest rates (dotted), the DNS forecast (right) and the FPCA-LAR forecast (left) for 1-, 6- and 12-month-ahead horizons on dates July 1994 for U.S. market and April 2010 for China market.

measure forecast accuracy, we compute the root mean squared forecast error (RMSFE) between the forecasts and the actual values. Table 2 reports the average values at various maturities. The results also support the superiority of the proposed FPCA-LAR model to the DNS model in the 6- and 12-month ahead forecasts, as shown by the smaller forecasting error values.

The out-of-sample forecasting shows that the proposed FPCA-LAR method attains overall better results than the DNS model. It demonstrates that forecasting yield curves using the adaptive method is more flexible and accurate than the accepted alternative. For some yield curves, particularly those from markets whose underlying

data generation process deviates from exponential basis functions, the FPCA factors are a better choice.

6 Conclusions

When forecasting yield curves, Diebold and Li (2006) propose the DNS model, which performs well for U.S. yield curves. However, its accuracy depends on the NS factor loadings, which are fixed and may misrepresent the underlying structure of the yield curves. It is also observed that the NS factors are persistent, what the stationary AR(1) process is unable to replicate. To counter these inadequacies, we propose the FPCA-LAR model, which is data-driven and able to account for non-stationarity in yield curves. For yield curves in U.S. and China markets, a simulation study and real data analysis illustrate the good performance of the FPCA-LAR model in comparison to the DNS model. In particular, the FPCA-LAR model improves the accuracy of the 6-month ahead and 12-month ahead forecasts. The analysis also reveals that FPCA-LAR is a flexible and adaptive approach, capable of capturing the underlying structure of any type of yield curves in different markets. Especially, for data with a data structure deviating from the exponential basis functions, the proposed method is a better choice for forecasting.

In our study, we developed adaptive forecasting technique on the factors obtained via FPCA. The technique can be also implemented for the factors based on the DNS model, where forecasts are constructed by using NS-LAR combinations. For the implementation, see Chen and Niu (2012). Moreover, we individually forecast the U.S. Treasury and China Treasury yield curves, respectively, ignoring any cross dependence between the two sovereign countries. Nevertheless, the cross dependence is also of great interest to study. Among others, Benko, Härdle and Kneip (2009) outline a common FPCA (CFPCA) method for modeling two or more time series of functional data, a useful reference for future study.

Although we have focused extensively on forecasting yield curves, the proposed method can be easily implemented in other economic research, such as forecasting electricity prices (Weron, 2006), where there is an increasing availability of data, not only large in size, but also in complexity.

Acknowledgement

This research was supported by National University of Singapore, FRC grant R-155-000-117-112.

Ying Chen, Bo Li

References

- [1] Baillie R. T. and Morana C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach, *Journal of Economics Dynamics and Control* 33: 1577–1592.
- [2] Bansal R. and Zhou H. (2002). Term structure of interest rates with regime shifts, *Journal of Finance* 57: 1997–2043.
- [3] Benko M., Härdle W. and Kneip A. (2009). Common functional principal components, *Annals of Statistics* 37: 1–34.
- [4] Borak S. and Weron R. (2008). A semiparametric factor model for electricity forward curve dynamics, *The Journal of Energy Markets* 1: 3.16.
- [5] Chan K. C., Karolyi G. A., Longstaff, F. A. and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate, *The Journal of Finance* 47: 1209. 1227.
- [6] Chen J. and Gupta A. (1997). Testing and locating variance changepoints with application to stock prices, *Journal of the American Statistical Association* 92: 739.747.
- [7] Chen L. (1996). Stochastic mean and stochastic volatility: A three factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives, *Financial Markets, Institutions and Instruments* 5: 1.88.
- [8] Chen Y., Härdle W. and Pigorsch U. (2010). Localized realized volatility modelling, *Journal of the American Statistical Association* 105: 1376.1393.
- [9] Chen Y. and Niu L. (2012). Adaptive dynamic nelson-siegel term structure model with applications. Manuscript.
- [10] Čížek, P., Härdle, W. and Spokoiny, V. (2009). Statistical inference for time-inhomogeneous volatility models, *Econometrics Journal* 12: 248 . 271.
- [11] Cochrane J. and Piazzesi M. (2005). Bond risk premia, *American Economic Review* 95: 138. 160.
- [12] Cox J., Ingersoll J. and Ross S. (1985). A theory of the term structure of interest rates, *Econometrica* 53: 385.407.
- [13] Dai Q., Singleton K. and Yang W. (2004). Predictability of bond risk premia and affine term structure models, *Working paper*, Stanford University.
- [14] Diebold F. X. (1986). Comment on: "modeling the persistence of conditional variance", *Econometric Reviews* 5: 51.56.

Forecasting Yield Curves in an Adaptive Framework

- [15] Diebold F. X. and Inoue A. (2001). Long memory and regime switching, *Journal of Econometrics* 105: 131–159.
- [16] Diebold F. X. and Li C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* 130: 337–364.
- [17] Duffie D. and Kan R. (1996). A yield-factor model of interest rates, *Mathematical Finance* 6: 379–406.
- [18] Egorov A. V., Li H. and Ng D. (2011). A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates, *Journal of Econometrics* 162: 55–70.
- [19] Engle R. F. and Granger C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing, *Econometrica* 55: 251–276.
- [20] Fama E. F. and Bliss R. R. (1987). The information in long-maturity forward rates, *American Economic Review* 77: 680–692.
- [21] Fengler M., Härdle W. and Mammen E. (2007). A semiparametric factor model for implied volatility surface dynamics, *Journal of Financial Econometrics* 5: 189–218.
- [22] Fernández-Rodríguez F. (2006). Interest rate term structure modeling using free-knot splines, *Journal of Business* 79: 3083–3099.
- [23] Ferraty F. and Vieu P. (2006). *Nonparametric Functional Data Analysis*, Springer, New York.
- [24] Giacomini E., Härdle W. and Krätschmer V. (2009). Dynamic semiparametric factor models in risk neutral density estimation, *Advances in Statistical Analysis* 93: 387–402.
- [25] Granger C.W. J. and Hyung N. (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns, *Journal of Empirical Finance* 11: 399–421.
- [26] Guidolin M. and Timmermann A. (2009). Forecasts of us short-term interest rates: A flexible forecast combination approach, *Journal of Econometrics* 150: 297–311.
- [27] Hall A. D., Anderson H. M. and Granger C.W. J. (1992). A cointegration analysis of treasury bill yields, *The Review of Economics and Statistics* 74: 116–126.
- [28] Hamilton J. D. and Susmel R. (1994). Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* 64: 307–333.

Ying Chen, Bo Li

- [29] Härdle W. K., Müller M., Sperlich S. and Werwatz A. (2004). *Nonparametric and Semiparametric Models*, Springer, New York.
- [30] Härdle W. and Trück S. (2010). The dynamics of hourly electricity prices, *Technical report*, SFB 649 Discussion Paper 2010-013.
- [31] Hull J. and White A. (1994). Numerical procedures for implementing term structure models ii: Two factor models, *Journal of Derivatives* 2: 37–48.
- [32] Jarrow, R., Ruppert, D. and Yu, Y. (2004). Estimating the interest rate term structure of corporate debt with a semiparametric penalized spline model, *Journal of the American Statistical Association* 99: 57–66.
- [33] Lamoureux C. G. and Lastrapes W. D. (1990). Persistence in variance, structural change and the GARCH model, *Journal of Business & Economic Statistics* 8: 225–234.
- [34] Litterman R. and Scheinkman J. (1991). Common factors affecting bond returns, *Journal of Fixed Income* 1: 54–61.
- [35] Liu C. and Maheu J. M. (2008). Are there structural breaks in realized volatility?, *Journal of Financial Econometrics* 6: 326–360.
- [36] McCulloch J. H. (1971). Measuring the term structure of interest rates, *Journal of Business* 44: 19–31.
- [37] McCulloch J. H. (1975). The tax adjusted yield curve, *Journal of Finance* 30: 811–829.
- [38] Merton R. C. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4: 141–183.
- [39] Mikosch T. and Stărică C. (2004a). Changes of structure in financial time series and the GARCH model, *REVSTAT Statistical Journal* 2: 41–73.
- [40] Mikosch T. and Stărică C. (2004b). Non-stationarities in financial time series, the long range dependence and the IGARCH effects, *Review of Economics and Statistics* 86: 378–390.
- [41] Müller H.-G. (2005). Functional modelling and classification of longitudinal data, *Scandinavian Journal of Statistics* 32: 223–240.
- [42] Nelson C. and Siegel A. (1987). Parsimonious modeling of yield curve, *Journal of Business* 60: 473–489.
- [43] Ramsay J. and Silverman B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.

Forecasting Yield Curves in an Adaptive Framework

- [44] Ramsay J. and Silverman B. (2005). *Functional Data Analysis*, 2 edn, Springer, New York.
- [45] Schaefer S. M. (1973). On measuring the term structure of interest rates, *discussion paper*, London Business School.
- [46] Schaefer S. M. and Schwartz E. (1984). A two factor model of the term structure: An approximate analytical solution, *Journal of Financial and Quantitative Analysis* 19: 413–424.
- [47] Scharth M. and Medeiros M. C. (2009). Asymmetric effects and long memory in the volatility of dow jones stocks, *International Journal of Forecasting* 25: 304–327.
- [48] Shea G. S. (1985). Interest rate term structure estimation with exponential splines: A note, *Journal of Finance* 40: 319–325.
- [49] So M. K. P., Lam K. and Li W. K. (1998). A stochastic volatility model with markov switching, *Journal of Business & Economic Statistics* 16: 244–253.
- [50] Svensson L. (1995). Estimating forward interest rates with the extended Nelson & Siegel method, *Penning- & Valutapolitik* 3: 13–26.
- [51] Vasicek O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* 5: 177–188.
- [52] Vasicek O. A. and Fong H. G. (1982). Term structure modeling using exponential splines, *Journal of Finance* 37: 339–356.
- [53] Vetzal K. R. (1994). A survey of stochastic continuous time models of the term structure of interest rates, *Insurance: Mathematics and Economics* 14: 139–161.
- [54] Weron R. (2006). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, John Wiley, New York