

# Adaptive Dynamic Nelson-Siegel Term Structure Model with Applications

Ying Chen<sup>a</sup> Linlin Niu<sup>b,c\*</sup>

<sup>a</sup>Department of Statistics & Applied Probability, National University of Singapore

<sup>b</sup>Wang Yanan Institute for Studies in Economics (WISE), Xiamen University

<sup>c</sup>MOE Key Laboratory of Econometrics, Xiamen University

## Abstract

We propose an Adaptive Dynamic Nelson-Siegel (ADNS) model to adaptively forecast the yield curve. The model has a simple yet flexible structure and can be safely applied to both stationary and nonstationary situations with different sources of change. For the 3- to 12-months ahead out-of-sample forecasts of the US yield curve from 1998:1 to 2010:9, the ADNS model dominates both the dynamic Nelson-Siegel (DNS) and random walk models, reducing the forecast error measurements by between 30 and 60 percent. The locally estimated coefficients and the identified stable subsamples over time align with policy changes and the timing of the recent financial crisis.

*Keywords:* Yield curve, term structure of interest rates, local parametric models, forecasting

*JEL Classification:* C32, C53, E43, E47

---

\*Corresponding author. Linlin Niu, Rm A301, Economics Building, Xiamen University, 361005, Xiamen, China. E-mail: llniu@xmu.edu.cn.

# 1 Introduction

Interest rates are important in economies. They are not only used by households, firms and financial institutions as primary input factors in making many economic and financial decisions, but they are also used by central banks in their conduct of monetary policy in order to fulfill policy goals for social investment, inflation and unemployment. The yield curve, which is also called the term structure of interest rates, illustrates the relationship between interest rates and time to maturities, i.e., the remaining time until a bond expires, which can vary from a few days to many years. Compared with individual interest rates, the yield curve is more important as it provides rich information about interest rates for different maturities. The long-term interest rate is the weighted expectation of future short-term interest rates adjusted for risks and risk compensation. The shape of the yield curve reflects the market's expectation on monetary policy and economic conditions. The average yield curve is upward sloping, i.e., the longer the time to maturity, the higher the interest rate. An inverted slope of the yield curve often anticipates economic recession, as happened in each occasion in the US economy during the last four decades. Moreover, the constantly growing derivatives market also make it necessary to understand the dynamics of the yield curve over time in order to price the interest rate derivatives. Hence yield curve forecasting has gained the attention of scholars and practitioners alike. Nevertheless, it is not an easy task, and it is often found that the theoretically sophisticated models such as the no-arbitrage or equilibrium term structure models forecast poorly, compared with a simple random walk model (see Duffee, 2002; Duffee, 2011).

As opposed to the class of no-arbitrage and equilibrium models with theoretical underpinnings, there are reduced-form models based on a statistical approach, where interest rates are often modeled with a univariate time series or a multivariate time series. The univariate class includes the random walk model, slope regression model, and the Fama-Bliss forward rate regression model (Fama and Bliss, 1987), where interest rates are modeled for each term of maturity individually. One cannot, however, efficiently explore the cross-sectional dependence of interest rates of different maturities for estimation and forecasting purposes. The multivariate class includes the vector autoregressive (VAR) models and the error correction models (ECMs), where interest rates of several maturities are considered simultaneously to utilize the dependence structure and cointegration. However, this class is subject to the curse of dimensionality, requiring the inclusion of more maturities and, hence, cannot provide a complete view of the whole curve.

Alternatively, the recent advancement of factor models enables the modeling and forecasting of the yield curve as a whole, where essential factors are extracted from the available yield curves, and the forecast is based on these resultant low dimensional factors. The factor approach retains the dependence information and is relatively easy to implement.

Among others, Diebold and Li (2006) propose a dynamic model based on the Nelson-Siegel factor interpolation (Nelson and Siegel, 1987), and show that the model not only keeps the parsimony and goodness-of-fit of the Nelson-Siegel interpolation, but also forecasts well compared with the traditional statistical models. This dynamic Nelson-Siegel (DNS) model has since become a popular forecasting tool of the yield curve. Instead of directly forecasting the yield curve, the DNS model extracts three factors via Nelson-Siegel factor loadings. The factors represent the level, slope and curvature of the yield curve. The evolutions of the factors are represented with either an autoregressive (AR) model for each individual factor or with a vector autoregressive (VAR) model for the three factors simultaneously. Given the forecasts of the factors and the fixed loadings, one can obtain the yield curve forecasts. Diebold and Li (2006) show that the AR(1) model specification performs better than many alternatives, including the VAR specification, as well as the univariate and multivariate time series models mentioned above.

However, we find that the recommended AR (1) model is deficient once the empirical features of the factors are considered. The factors, like many other financial variables, are very persistent. As an illustration, Figure 1 displays the time evolution of the Nelson-Siegel factor of level and its sample autocorrelations. The factor is extracted from the US yield curve data between 1983:1 to 2010:9, which we will describe in the next section. It shows that the level factor has a long memory, with the sample autocorrelations hyperbolically decaying up to lag 100. The AR(1) model, albeit a good forecast performance compared with the alternative models, is unable to capture the persistence characteristics of the time series.

[Figure 1. Time Plot of the Nelson-Siegel Level Factor and its Sample Autocorrelation Function]

Naturally, persistence may be modeled with a long memory view using, for example, fractionally integrated processes, see Granger (1980), Granger and Joyeux (1980) and Hosking (1981). Note that these models rely on a stationary assumption in that the model parameters are assumed to be constant over time. Alternatively, a short memory view can be employed, which says persistence can also be spuriously generated by a short memory model with heteroscedasticity, structural breaks or regime switching, see Diebold and Inoue (2001) and Granger and Hyung (2004). In other words, the fitted model is assumed to hold only for a possibly short time period and the evolution of the local subsample can be represented by a simple structural model, though the dynamics are changing from one time period to another. In the literature about interest rates, several nonstationary models have been proposed for yield curve modeling, see Ang and Bekaert (2002) and Bansal and Zhou

(2002) for regime shifts, Guidolin and Timmermann (2009) for the regime switching VAR model, and Ang, Boivin, Dong and Loo-Kung (2011) for a monetary policy shift in a no-arbitrage quadratic term structure model. The question is, which view is more appropriate: a stationary model with long memory or a nonstationary model with a simple structure but time-varying parameters?

It appears that the short memory view is quite realistic and easily understood in the context of business cyclical dynamics, policy changes and structural breaks. Figure 2 plots the monthly data, from 1983:1 to 2010:9, of three interest rates, with maturities of 3 months, 24 months and 10 years. The gray dashed line with circles shows the CPI inflation, which is an important macroeconomic factor affecting nominal interest rates. The shaded bars mark the three NBER recessions during the sample period: namely 1990:7-1991:3, 2001:3-2001:11 and 2007:12-2009:6. Over the whole sample period, interest rates show an obvious downward trend. However, once we look at each subsample, separated by the three recessions, they appear more stationary. Although the recession periods may not be the best divisions to isolate stationary periods, statistically or economically, the point here is to show that there may exist subsamples where the interest rates are approximately stationary and that can be modeled sufficiently well with simple autoregressive processes. Our study employs the methodology in Chen, Härdle and Pigorsch (2010) to detect divisions in real time.

[Figure 2. Three Yields and the CPI Inflation in the US (1983.1-2010.9)]

In the volatility forecasting literature, Chen et al. (2010) propose a local AR (LAR) model, that allows for time-varying parameters, to model the seemingly persistence of volatility with a short memory view. With time-varying parameters, the LAR model is flexible and able to account for various changes of different magnitudes and types including both sudden jumps or smoothed changes. Moreover, in terms of out-of-sample prediction, it performs better than several alternative models including long memory models and some regime switching models. A direct application of the LAR(1) model for yield curve forecast is apt. However, as interest rate movements are closely related to macroeconomic conditions (see Taylor, 1993; Mishkin, 1990; Ang and Piazzesi, 2003), it is more appropriate to widen the scope of the LAR model by introducing macroeconomic variables such as inflation, growth rate, etc.

In our study, we develop a new model to adaptively forecast the yield curve. It extracts Nelson-Siegel factors cross-sectionally as in the DNS model. Then a local autoregressive model with exogenous variables (LARX) is adopted to fit each factor dynamically. The rationale is that, to conduct an estimation at any point of time, we take all available past

information and find a subsample where the observations included can be well represented by a local stationary model with approximately constant parameters. In other words, there are not any kind of significant changes in the estimation window. We then forecast the factors as well as the yield curve based on the fitted model. We name it the adaptive dynamic Nelson-Siegel (ADNS) model. At each point of time, this resulting model is a DNS model over the detected stable subsample, where the dynamics of each factor are described by an AR(1) process with exogenous variables. The adaptive procedure of stable subsample detection is crucial for this model to distinguish it from the DNS model with predetermined sample length. We investigate the performance of this proposed model in real data analysis, and compare it with popular alternative models including the DNS and the random walk models. We show the substantial advantage gained from using the adaptive approach in forecasting, and bring useful insights in diagnosing yield dynamics with the produced stable intervals and parameter evolvments.

The rest of the paper is organized as follows. Section 2 describes the data used. Section 3 presents the ADNS forecast model with a detailed discussion on the local model for each factor. Section 4 reports the real data analysis and forecast comparison with alternative models. Section 5 concludes.

## 2 Data

We use US Treasury zero-coupon-equivalent yield data with 15 maturities: 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, from 1983:1 to 2010:9. The short-term yields of 3 and 6 months are converted from the 3- and 6-month Treasury Bill rates on a discount basis, available from the Federal Reserve's H.15 release of selected interest rates. The remaining yields with maturities of integer years are taken from research data of the Federal Reserve Board, as released by Gürkaynak, Sack and Wright (2007). We add the 9-, 18-, and 30-month yields interpolated according to the parameters provided in their data file. Both data are of daily frequency, updated constantly. We choose the data at the end of each month to form a monthly data set for our empirical analysis. Figure 3 shows the yield curve dynamics of the data set.

[Figure 3. US Yield Curve (1983.1-2010.9)]

For macroeconomic variables, we choose the monthly CPI annual inflation rate as an exogenous variable in determining the factor dynamics of the LARX model. Inflation is not only important for determining the level of interest rates in the long run, as described by

Fisher’s equation, but is also highly correlated with interest rates as can be seen from Figure 2. Although it is interesting to explore the joint dynamics of the yield curve and inflation, as our primary goal is to forecast the yield curve, we opt for the simple and effective method of taking inflation as an exogenous variable in the state dynamics. It is also beneficial and possible to include other macroeconomic variables or factors in the states. We relegate this investigative task to future work regarding defining the optimal set of variables or factors to be included. Our focus remains using inflation as a relevant macroeconomic variable to illustrate the effectiveness of the ADNS model.

### 3 Adaptive dynamic Nelson-Siegel model

In the adaptive dynamic Nelson-Siegel (ADNS) model, the cross section of the yield curve is assumed to follow the Nelson and Siegel (1987) framework, based on which three factors are extracted via the ordinary least squares (OLS) method. For each single factor, a local autoregressive model with exogenous variables (LARX) is employed to forecast. In the LARX approach, the parameters are time dependent, but each LARX model is estimated at a specific point of time under *Local Homogeneity*. Local homogeneity means that, for any particular time point, there exists a past subsample over which the parameters, though time varying, are approximately constant. In this situation the time-varying model does not deviate much from the stationary model with constant parameters over the subsample, and hence the local maximum likelihood estimator is still valid. The interval of local homogeneity is automatically selected for each time point by sequentially testing the significance of the divergence.

#### 3.1 Extract Nelson-Siegel factors

In the framework of Nelson and Siegel (1987), the yield curve can be formulated as follows:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau), \quad \epsilon_t(\tau) \sim N(0, \sigma_\epsilon^2) \quad (1)$$

where  $y_t(\tau)$  denotes the yield curve with maturity  $\tau$  (in months) at time  $t$ . Parameter  $\lambda$  controls the exponentially decaying rate of the loadings for the slope and curvature factors, and a smaller value produces a slower decay. However, within a wide range of values, Nelson and Siegel (1987) find that the goodness-of-fit of the yield curve is not very sensitive to the specific value of  $\lambda$ . In our study, we follow Diebold and Li (2006) and set  $\lambda = 0.0609$  which maximizes the curvature loading at a medium maturity of 30 months. Under fixed  $\lambda$ , if there exists any form of non-stationarity in the observations  $y_t(\tau)$ , then it is solely attributed to changes in the sequences of the factors.

The three factors,  $\beta_{1t}$ ,  $\beta_{2t}$  and  $\beta_{3t}$  are denoted as level, slope and curvature, respectively. Their loadings on the yield curve are displayed in Figure 4. The loadings of  $\beta_{1t}$  are 1 for all maturities, implying that the higher the factor, the higher the level of the whole yield curve. Empirically it is close to the long-term yield of 10 years. The loadings of  $\beta_{2t}$  start from 1 at the instantaneous spot rate and decay towards zero as maturities increase. The average value of  $\beta_{2t}$  is negative, indicating a normal upward-sloping yield curve; when  $\beta_{2t}$  is positive, it leads to a downward slope of the yield curve which often anticipates recessions. The slope factor,  $\beta_{2t}$ , is highly correlated with the negative spread of the yield curve, e.g., the difference between the short-term and long-term yields,  $-[y(120) - y(3)]$ . The loadings of  $\beta_{3t}$  display a humped shape, which peak around the maturities of medium duration. Thus  $\beta_{3t}$  is named a curvature factor. A positive value of  $\beta_{3t}$  indicates that the yield curve is concave with a hump, while a negative value means that the yield curve is convex with an inverted hump. Empirically, it is approximated by twice the medium yield minus the sum of short- and long-term yields, e.g.,  $2 \times y(24) - [y(120) + y(3)]$ .

[Figure 4. Nelson-Siegel Factor Loadings]

The time evolution of the three factors is displayed in Figure 5, where the empirical proxies of level, slope and curvature are also depicted in dotted lines. It shows that the three factors are consistent with the empirical proxies, demonstrating similar values and patterns. The correlations between the three factors and their proxies are as high as 0.976, 0.991 and 0.997, respectively. The level factor has a downward trend, which indicates that the yield curve, on average, has been decreasing along time. The slope factor is negative in most cases, which means that the yield curve is normally upward sloping. Nevertheless, a few positive values are observed at three locations, specifically leading the NBER recessions. The curvature factor fluctuates wildly around zero, indicating that both the hump and inverted hump shapes occur frequently. Nonetheless the curvature has been persistently negative during the past decade.

[Figure 5. Time Evolution of Nelson-Siegel Factors Extracted from US Yield Curve]

### 3.2 Local autoregressive model with exogenous variables

Now, for each one-dimensional factor  $\beta_{jt}$ , with  $j = 1, 2$  and  $3$ , a local model is adopted to perform forecast. For notational simplification, we drop the subscript  $j$  in the following elaboration. Given a univariate time series  $\beta_t \in \mathbb{R}$ , we identify a past subsample for any

particular time point, over which all the included observations can be well represented by a local model with approximately constant parameters, i.e., local homogeneity holds. More specifically, given all available past observations for the time point, we seek the longest time interval, beyond which there is a high possibility of a structural change occurring which means the local homogeneity assumption no longer holds. The likelihood ratio test helps to select the local subsample. Among many possible modeling candidates, we focus on an AR(1) model, motivated by its simplicity, parsimony and hence a good out-of-sample forecasting ability. Moreover, to increase modeling flexibility, we allow relevant exogenous variables in addition to the one lagged autoregressive component. The proposed model is a local AR(1) model with exogenous variables, or a LARX(1) model.

The LARX(1) model is defined through a time-varying parameter set  $\theta_t$ :

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

where  $\theta_t = (\theta_{0t}, \theta_{1t}, \theta_{2t}, \sigma_t)^{\top}$ . The innovations  $\mu_t$  have a mean of zero and a variance of  $\sigma_t^2$ . In our study, the three Nelson-Siegel factors are individually represented by the LARX(1) model and share the inflation rate as the exogenous variable,  $X$ . The model parameters are time dependent and can be obtained by the (quasi) maximum likelihood estimation, once the local subsample is specified.

Suppose the subsample is given for time point  $s$ , denoted as  $I_s = [s - m_s, s - 1]$ , over which the process can be safely described by an autoregressive model with exogenous variables (ARX) with constant parameter  $\theta_s$ . Then suppose that the local maximum likelihood estimator  $\hat{\theta}_s$  is defined as:

$$\begin{aligned} \hat{\theta}_s &= \arg \max_{\theta \in \Theta} L(\mathbf{f}; I_s, \theta, X) \\ &= \arg \max_{\theta \in \Theta} \left\{ -(m_s - 1) \log \sigma_s - \frac{1}{2\sigma_s^2} \sum_{t=s-m_s+1}^{s-1} (f_t - \theta_0 - \theta_1 f_{t-1} - \theta_2 X_{t-1})^2 \right\} \end{aligned}$$

where  $\Theta$  is the parameter space and  $L(\mathbf{f}; I_s, \theta, X)$  is the local log-likelihood function. In practice, the stable subsample is unknown and the number of possible candidates is large, e.g. as many subsamples as there are sample periods. Our goal is to select an optimal subsample among a finite set of alternatives. The finest search would be on all possible candidates contained in the sample. However, to alleviate the computational burden, we divide the sample into discrete intervals with each interval containing  $M$  periods, such that the resulting number of subsamples,  $K_s$ , is roughly the available sample length divided by  $M$ . Suppose there are  $K_s$  candidate subsamples for any particular time point  $s$ , or

$$I_s^{(1)}, \dots, I_s^{(K_s)} \text{ with } I_s^{(1)} \subset \dots \subset I_s^{(K_s)},$$

which starts from a short subsample,  $I_s^{(1)}$ , where the ARX(1) model with constant parameters provides a reasonable fit. The selection procedure then iteratively extends this subsample with the next interval of  $M$  periods and sequentially tests for possible structural change in the next longer subsample. This is performed via an  $\ell - 1$  type of likelihood ratio test. The test statistic is defined as

$$T_s^{(k)} = |L(I_s^{(k)}, \hat{\theta}_s^{(k)}) - L(I_s^{(k)}, \hat{\theta}_s^{(k-1)})|^{1/2},$$

where  $L(I_s^{(k)}, \hat{\theta}_s^{(k)}) = \max_{\theta \in \Theta} L(\mathbf{f}; I_s^{(k)}, \theta, X)$  denotes the fitted likelihood under *hypothetical* homogeneity and  $L(I_s^{(k)}, \hat{\theta}_s^{(k-1)}) = L(\mathbf{f}; I_s^{(k)}, \hat{\theta}_s^{(k-1)}, X)$  refers to the likelihood with the estimate under *accepted* local homogeneity in the testing subsample. The test statistic measures the difference between these two estimates. If the difference is significant, it indicates that the model changes more than what would be expected due to sampling changes. If this is the case, then the procedure terminates and the latest accepted subsample is selected. Otherwise, we accept the longer subsample and continue the procedure until a change is found or the longest subsample,  $I_s^{(K_s)}$ , is reached under local homogeneity. The significance level is controlled by critical values obtained by Monte Carlo simulations. The basic idea is to generate a stationary ARX(1) time series based on the available data, i.e., where local homogeneity is guaranteed, and then compute the cut-off value that leads to a decision of not rejecting the null hypothesis of homogeneity. The cut-off values tend to be critical values. The details of the computation of the critical values can be found in Chen et al. (2010).

The local model is then fitted over the selected subsample and can be safely used for predictions. Since the selection of subsamples is time dependent, the optimal subsamples are changing over time. A related assumption is that the selected model holds with high probability for the forecast horizon. We may wish to predict structural change out of sample, such as Pesaran, Pettenuzzo and Timmermann (2006) who use a Bayesian procedure to incorporate the future possibility of breaks. In this regard, our approach is less sophisticated, however it is less computationally demanding and, as will be shown in the results, is very effective.

### 3.3 Modeling, forecasting factors and the yield curve

Essentially, the ADNS model can be summarized in a state-space representation as follows:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau) \quad (3)$$

$$\beta_{it} = \theta_{0t}^{(i)} + \theta_{1t}^{(i)} \beta_{it-1} + \theta_{2t}^{(i)} X_{t-1} + v_t^{(i)}, \quad v_t^{(i)} \sim (0, \sigma_{it}^2), \quad i = 1, 2, 3. \quad (4)$$

At each point of time, once the most recent optimal subsample is selected, the resulting model amounts to a DNS model with possible exogenous factors. This is in spite of our ADNS model taking time-varying coefficients. However, we will show that, with adaptive selection of subsample length, the forecast accuracy is substantially improved.

We estimate two specifications of the ADNS model. One is denoted as LAR and has no exogenous factor, and the other denoted as LARX and has an exogenous factor, i.e., CPI inflation. The LAR specification provides a pure comparison to the DNS model with an AR(1) process for each factor, and shows the stark difference in forecasts resulting from the adaptive feature. The LARX specification then shows additional benefits gained by including inflation.

The yield curve forecast  $h$ -step ahead is directly based on the  $h$ -step ahead forecast of Nelson-Siegel factors:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \hat{\beta}_{3,t+h/t} \left( \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (5)$$

where, for the LAR specification, the factor forecasts are obtained by

$$\hat{\beta}_{i,t+h/t} = \hat{\theta}_{0t} + \hat{\theta}_{1t} \hat{\beta}_{i,t}, \quad i = 1, 2, 3, \quad (6)$$

and, for the LARX specification, by

$$\hat{\beta}_{i,t+h/t} = \hat{\theta}_{0t} + \hat{\theta}_{1t} \hat{\beta}_{i,t} + \hat{\theta}_{2t} X_t, \quad i = 1, 2, 3. \quad (7)$$

The coefficient  $\hat{\theta}_{jt}$  ( $j = 0, 1, 2$ ) is obtained by regressing  $\hat{\beta}_{i,t}$  on an intercept and  $\hat{\beta}_{i,t-h}$  for LAR, and by regressing  $\hat{\beta}_{i,t}$  on an intercept,  $\hat{\beta}_{i,t-h}$  and  $X_{t-h}$  for LARX. We follow Diebold and Li (2006) to directly predict factors at  $t+h$  with factors at time  $t$ , instead of using the iterated one-step ahead forecast, to obtain a multistep ahead forecast.

## 4 Empirical application of the ADNS model

In this section, we apply the ADNS model to perform out-of-sample forecast of the yield curve and elaborate on its usefulness in detecting structural breaks and monitoring parameter changes in the state dynamics.

### 4.1 Out-of-sample forecast performance of ADNS

Diebold and Li (2006) show that the DNS model does well for out-of-sample forecasts in comparison with various other popular models. So, compared with the DNS model, how much can the ADNS model benefit from this adaptive feature for forecasting yields? We

perform an out-of-sample forecast comparison to answer this question.

#### 4.1.1 Forecast procedure

We report the factor and yield curve forecasts of the ADNS model with the two specifications: LARX(1) and LAR(1).

We use the yield data as described in Section 2 from 1983:1 to 2010:9. We make an out-of-sample forecast in real time starting the estimation from 1997:12, and make 1-month, 3-months, 6-months and 12-months ahead forecasts for 1998:1, 1998:3, 1998:6 and 1998:12, respectively. We move forward one period at a time to redo the estimation and forecast until reaching the end of the sample. The 1-month forecast is made for 1998:1 to 2010:9, a total of 153 months; the 3-, 6- and 12-months ahead forecasts are made for 1998:3 to 2010:9 with 151 months, 1998:6 to 2010:9 with 148 months, and 1998:12 to 2010:9 with 142 months, respectively.

At each point of estimation, we use data up to that point. We first extract the three Nelson-Siegel factors using OLS for the available sample. For each factor, we assume an LAR(1) or LARX(1) process for the different forecast horizons, and test the stability backward using our method as described in the previous section. To reduce computational burden, we use a universal set of intervals for each time point, starting with the shortest subsample covering the past 6 months, with the interval between two adjacent subsamples equal to 6 months, ending with the longest subsample. Once an optimal subsample is selected against the next longer subsample, the change is understood as having taken place in the 6 month' period in between. In doing so, we trade off precision in break identification with computational efficiency. However, since our primary goal is to estimate from the identified sample with relatively constant parameters, the lack of a few observations are not of major concern. Also, as many changes can happen smoothly, point identification of breaks may not be precise.

We choose the most recent stable subsample as the optimal sample to estimate the parameters, based on which we do out-of-sample forecasts 1-, 3-, 6- to 12- months ahead as in Equation (6) for LAR and in Equation (7) for LARX, respectively. The predicted Nelson-Siegel factors at the specific horizon will then be used to form a yield forecast as in Equation (5).

#### 4.1.2 Alternative models for comparison

For comparison, we choose the DNS model and the random walk model. The DNS model is a direct and effective comparison for our ADNS model. The random walk model serves as a very natural and popular benchmark for financial time series with non-stationary characteristics.

### 1) DNS model

We forecast the Nelson-Siegel factors and yield curve with the DNS model using three strategies of sample length selection: a five-year rolling estimation, a ten-year rolling estimation and a recursive estimation. With a rolling estimation, we fix the sample length at each point of time. For the five-year or ten-year rolling estimation, we include in the regression the most recent 60 or 120 months' observations, respectively. With the recursive estimation, we expand the sample length one point at a time as we move forward, with the initial sample being about 15 years (180 months). These three cases correspond to a comparison with increasing sample lengths.

These rolling and recursive strategies are popular in forecast practice. The recursive estimation aims to increase information efficiency by extending the available sample as much as possible, such as that used in Diebold and Li (2006) with an initial sample length of about 9 years and extending it forward. The rolling estimation reflects the idea of making a trade-off between information efficiency and possible breaks in the data generating process, such as in the 15-year rolling window used in Favero, Niu and Sala (2012). Note that a predetermined sample length is unable to simultaneously serve both purposes: ensure stable estimation under stationarity and have a quick reaction to structural changes.

The forecast of factors and yields are similar to the ADNS model with the selected sample length, as shown in Equations (5) and (6). The only difference is that parameters  $\hat{\theta}_j$  ( $j = 0, 1, 2$ ) are obtained by regressing  $\hat{\beta}_{i,t}$  on an intercept and  $\hat{\beta}_{i,t-h}$  with a sample of predetermined length.

### 2) Random walk model

With the random walk model of no drift, yields are directly forecasted by its current value, or

$$\hat{y}_{t+h/t}(\tau) = \hat{y}_t(\tau).$$

#### 4.1.3 Measures of forecast comparison

We use two measures as indicators of forecast performance, the forecast root mean squared error (RMSE) and the mean absolute error (MAE). We calculate these measures for the Nelson-Siegel factors and yields in the DNS model, and for yields in the random walk model.

#### 4.1.4 Forecast results

Table 1 reports the forecast RMSE and MAE for the three Nelson-Siegel factors for the forecast horizons (denoted by  $h$ ) of 1, 3, 6 and 12 months, respectively. Each column reports the results of the ADNS model with the two specifications of LARX and LAR, and of the DNS models using three different sample length strategies: 5-year and 10-year

rolling window with fixed lengths and the recursive forecast with expanding window length. In each column, the number in bold indicates the best forecast for each horizon and each measure, and the underlined number the second best.

[Table 1. Forecast Comparison of Factors]

Table 1 shows that, without exception, the LAR and LARX specifications are always the best two across all factors, horizons and RMSE/MAE measures. The ADNS model, with the adaptive selection of sample lengths, always forecasts better than the DNS model with various predetermined window lengths. The advantage grows as the forecast horizon increases. At the 12-month forecast horizon, the RMSE/MAE of the ADNS are, in general, 40-50 percent less than those produced by the DNS models. The LARX model tends to outperform the LAR model, indicating that the inflation factor helps to predict the yield factors in the ADNS model, in addition to their lagged information.

Among the DNS models, there is no clear pattern as to which window length performs the best. Although in most cases the 10-year rolling window produces better forecasts than the 5-year rolling window, the recursive window, which incorporates more observations, tends to do better for the first two factors but fare worse for the third factor than the 10-year rolling window.

To visualize the forecast differences, we plot the factor forecasts from the LARX of the ADNS model and the 10-year rolling window of the DNS model together with the real data in Figure 6. Each row displays the forecast for one factor, for the forecast horizons of 1-, 6- and 12 months, respectively. Factors extracted from real yields are displayed with solid lines; LARX forecasts of the ADNS are displayed with dotted lines, and the 10-year rolling forecasts of the DNS with dashed lines. Although it is difficult to distinguish the three series for the 1-month ahead forecast because they are so close to each other, as the forecast horizon lengthens, it is clear that the dotted line of the LARX forecast tends to trace the factor dynamics much more closely than does the DNS. What is dramatic is its capability of predicting the large swings of slope and curvature during the financial crisis. From 2007 to 2008, the slope reduced sharply reflecting an ever decreasing short rate, while initially the LARX reacts insufficiently to the decline in the 6- and 12-month ahead forecast, it quickly catches up and correctly predicts the sharp fall during 2008. The DNS with the 10-year rolling window, however, not only reacts with a persistent lag, but also wrongly predicts the change in direction for a persistent period.

[Figure 6. Plots of Factor Forecasts with LARX of ADNS and 10-year Rolling of DNS]

Table 2 reports the forecast RMSE and MAE for yields with the selected maturities of 3 months, 1 year, 3 years, 5 years and 10 years. In addition to the forecasts from the ADNS and the DNS models, we also report the random walk forecast.

[Table 2. Forecast Comparison of Yields]

Again, the ADNS model dominates the DNS model across yields, forecast horizons and forecast measures. Even if we compare the second best result of the ADNS model and the best result among the DNS models, the reduction in the RMSE/MAE of the ADNS model reaches about 30 percent for the 3-month ahead forecast, and even to about 60 percent for the 12-month ahead forecast. The advantage is tremendous! Compared with the random walk model, the ADNS model does not always have an absolute advantage, for the short-to-medium yields at the 1-month ahead forecast. The outperformance begins to be obvious from the 3-month ahead forecast. Also, the reduction in RMSE/MAE when compared with the random walk is fairly substantial ranging between 30 to 60 percent in multistep ahead forecasts.

In Figure 7 we plot the yield forecast from the LARX of the ADNS model and the 10-year rolling window of the DNS model together with the real data. We select yields with maturities of 3 months, 36 months and 120 months to represent the short-, medium- and long-term yields. Each row displays the forecast for one yield, for the forecast horizons of 1, 6 and 12 months, respectively. Real yields are displayed by solid lines; LARX forecasts are displayed by dotted lines, and 10-year rolling forecast by dashed lines. As the forecast horizon lengthens, it is clear that the dotted line of the LARX forecasts tends to follow the yield dynamics much more closely than the DNS forecasts. The excellence of LARX in predicting dramatic changes and large swings in factors now follows through to superiority in predicting yield changes. The ADNS model can not only accurately predict the large swings in short- and medium-term yields, but also captures the subtle movement in the 10-year yields 6 months or even 12 months earlier.

[Figure 7. Plots of Yields Forecast with LARX and 10-year Rolling]

#### 4.1.5 Lengths of stable subsamples

Table 3 summarizes the average lengths detected for sample stability along the forecast for each factor and each horizon. When  $h = 1$ , it is a standard AR(1)/ARX(1) model of lag 1. When  $h > 1$ , it is an AR(1)/ARX(1) model with lag  $h$ . The overall length of

the stable subsample is surprisingly low compared to the conventional sample size used in the literature for yield curve forecasts at a monthly frequency, with a maximum of 73 months (about six years) for the first Nelson-Siegel factor and for the one-month ahead forecast from the LAR specification. For other cases, the average lengths of the stable subsample is around 30 months, or two-and-a-half years, which is much shorter than the conventionally-used 5-year, 10-year rolling or recursive samples.

The dominating performance of the adaptive approach indicates that the instability problems inherent in yield factor dynamics outweighs the information inefficiency with short samples, such that shorter samples are preferred to longer samples with more observations. The question is how short the selected subsample should be. Instead of relying on a rule-of-thumb selection of a predetermined window length, our approach detects the optimal subsamples at each point of time employing the adaptive procedure.

[Table 3. Average Lengths of Stable Sample Intervals Detected by LAR and LARX]

## 4.2 Detecting breaks and monitoring parameter evolution

We have shown that the ADNS model predicts yield dynamics remarkably better than the DNS model, benefiting from the adaptive feature of the model. There must be useful information hidden in the detected breaks and parameter evolutions that drive this excellence in prediction. Understanding the time-varying features of the state dynamics in real time would be very valuable not only for forecasts, but also for real-time diagnosis of the yield curve and for policy investigation.

Figure 8 plots the detected stable subsamples and breaks along the forecasting time for each factor. We report the results of the 1-step ahead forecast of the LARX(1) model, i.e., a standard AR(1) model with the most adjacent lag and lagged inflation. The vertical axis denotes the time when the estimation or forecast is made. At each point of time, an optimal subsample is detected which is shown as a (light yellow) line along the horizontal axis of time. At the end of the line, a (dark blue) dotted line indicates the period, in this case six months, during which the most recent break happens, i.e., the immediate interval beyond the identified stable subsample. When these (dark blue) dotted lines are stacked along the forecast time, i.e., the vertical dimension, some common areas of detected breaks become evident.

[Figure 8. Breaks of Factors Detected in the LARX(1)]

Observing Figure 8, we find that the stable subsamples of the level factor are typically longer than those of slope and curvature. For level, the intervals of 1994-1998, 1998-2003 and 2003-2008 are identified as relatively stable within the whole sample. For slope and curvature, breaks are constantly detected during the forecast exercise, indicating ever-evolving parameters. Moreover, breaks within the level factor often lead breaks within the other two factors. There are three occasions of sequential breaks led by level breaks in 1994, 1998 and 2003, with breaks in slope and curvature following in 1995, 1998 and 2003. Overall, there seems to be a common break in all three factors just at the beginning of the financial crisis and recession in 2008.

Those major breaks in level align with some big events and debated policy change. For example, the 1998 and 2008 breaks correspond well to the 1997-1998 financial market turmoil and the 2008 financial crisis. The break at the beginning of 2003 follows the internet bubble bursting, and when the US Federal Reserve had begun to conduct relatively loose monetary policy for a prolonged period. Taylor (2011), among many others, admonishes that “interests rates were held too low for too long by the Federal Reserve”. Measured by the Taylor rule (Taylor, 1993), the monetary reaction to inflation was remarkably low between 2003 and 2005. Ang et al. (2011) estimate a time-varying Taylor rule using the quadratic term structure model and confirm similar findings with richer data of the whole yield curve. The policy change is criticized as having accommodated the financial bubble which eventually ends in the recent financial crisis.

When breaks are detected, it is because parameters have changed so dramatically that the most recent sample estimates are not valid when extending the sample further back. Therefore a close monitoring of the evolvement of parameters is helpful to understand the origins of the breaks.

Figure 9 presents the parameter estimates along the forecasting path. For each of the three factors, the ARX process involves four parameters as in Equation (2): the intercept,  $\theta_{0t}$ , the autoregressive coefficient,  $\theta_{1t}$ , the coefficient of inflation,  $\theta_{2t}$ , and the standard deviation of the error term,  $\sigma_t$ . We report in each column the evolvement of the four parameters for each factor.

[Figure 9. Parameter Evolvement in the LARX Model]

This figure shows that the coefficients of the ARX(1) model of each Nelson-Siegel factor vary wildly through time, which justifies the application of an adaptive approach. Within the optimal interval detected at the each point of time, the autoregressive coefficient mostly lies within the stable range, indicating that our short memory view, indeed, can be justified as long as our identification procedure works reasonably well.

We observe that the CPI inflation coefficient on the level factor changed to become negative after 2000, which corresponds to the loose monetary policy much criticized by academia as having precipitated the recent financial crisis. Although our model is not as structural as those of Taylor (2011) and Ang et al. (2011), the discovery of this simple pattern is very revealing about an important break in the relationship between interest rates and inflation.

Simple as it is, our results provide a vivid picture of the evolution of the volatilities of the three factors along time. The bottom row shows that the volatilities of all factors have surged during the recent financial crisis; other coefficients also show dramatic changes in this period. These results indicate both higher risks and a regime switch against the outbreak of the financial crisis.

## 5 Conclusion

By optimally selecting the sample period over which parameters are approximately constant for the factor dynamics, the adaptive dynamic Nelson-Siegel (ADNS) model we propose produces superior forecasts of the yield curve compared with both the dynamic Nelson-Siegel model with predetermined sample lengths and the random walk model. The ADNS model not only improves forecast precision substantially by reducing forecast errors by between 30 to 60 percent, but also accurately predicts the large swings in the yield curve, six to twelve months ahead of the financial crisis. This forecasting excellence originates from its capacity to adaptively and swiftly detect structural breaks. The detected breaks and the evolution of the estimated parameters are revealing about the structural changes in economic conditions and monetary policy. We believe this model and the method used are of great value to policy makers, practitioners and academics alike for drawing ever more precise and timely determinations from the yield curve information about imminent recessions.

Moreover, for monitoring and forecasting purposes, this adaptive modeling method is data-driven and can be widely applied to other macroeconomic and financial time series, either uni- or multivariate, stationary or nonstationary, with different sources of changes including heteroskedasticity, structure break, and regime switching.

### Acknowledgement

Linlin Niu acknowledges the support of the Natural Science Foundation of China (Grant No. 70903053).

## References

- Ang, A. and Bekaert, G. 2002. Regime switches in interest rates, *Journal of Business and Economic Statistics* **20**: 163–182.
- Ang, A. and Piazzesi, M. 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**(4): 745–787.
- Ang, A., Boivin, J., Dong, S. and Loo-Kung, R. 2011. Monetary policy shifts and the term structure, *Review of Economic Studies* **78**: 429–457.
- Bansal, R. and Zhou, H. 2002. Term structure of interest rates with regime shifts, *Journal of Finance* **57**(5): 1997–2043.
- Chen, Y., Härdle, W. and Pigorsch, U. 2010. Localized realized volatility modelling, *Journal of the American Statistical Association* **105**: 1376–1393.
- Diebold, F. X. and Inoue, A. 2001. Long memory and regime switching, *Journal of Econometrics* **105**: 131–159.
- Diebold, F. X. and Li, C. 2006. Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.
- Duffee, G. R. 2002. Term premia and interest rate forecasts in affine models, *Journal of Finance* **57**: 405–443.
- Duffee, G. R. 2011. Forecasting with the term structure: The role of no-arbitrage restrictions, *working paper*, Johns Hopkins University.
- Fama, E. and Bliss, R. 1987. The information in long-maturity forward rates, *American Economic Review* **77**: 680–692.
- Favero, C. A., Niu, L. and Sala, L. 2012. Term structure forecasting: No-arbitrage restrictions versus large information set., *Journal of Forecasting*. Forthcoming.
- Granger, C. W. and Joyeux, R. 1980. An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**: 5–39.
- Granger, C. W. J. 1980. Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* **14**: 227–238.
- 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.

- Guidolin, M. and Timmermann, A. 2009. Forecasts of US short-term interest rates: A flexible forecast combination approach, *Journal of Econometrics* **150**: 297–311.
- Gürkaynak, R. S., Sack, B. and Wright, J. H. 2007. The U.S. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* **54**(8): 2291–2304.
- Hosking, J. R. M. 1981. Fractional differencing, *Biometrika* **68**: 165–176.
- Mishkin, F. S. 1990. The information in the longer maturity term structure about future inflation, *The Quarterly Journal of Economics* **105**: 815–828.
- Nelson, C. and Siegel, A. 1987. Parsimonious modeling of yield curve, *Journal of Business* **60**: 473–489.
- Pesaran, M. H., Pettenuzzo, D. and Timmermann, A. G. 2006. Forecasting time series subject to multiple structural breaks, *Review of Economic Studies* **73**: 1057–1084.
- Taylor, J. B. 1993. Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* **39**: 195–214.
- Taylor, J. B. 2011. Origins and policy implications of the crisis, in: R. Porter (Ed.), *New Directions in Financial Services Regulation*, MIT Press, Cambridge, MA, pp. 13–22.

**Table 1. Forecast Comparison of Factors**

NS1	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.281</b>	<b>0.204</b>	<b>0.334</b>	<b>0.247</b>	<b>0.374</b>	<b>0.273</b>	<u>0.462</u>	<b>0.342</b>
<b>LAR</b>	<u>0.307</u>	<u>0.216</u>	<u>0.364</u>	<u>0.266</u>	<u>0.430</u>	<u>0.338</u>	<b>0.453</b>	<u>0.344</u>
5-year Rolling	0.333	0.239	0.462	0.356	0.623	0.496	0.738	0.540
10-year Rolling	0.330	0.232	0.463	0.350	0.617	0.495	0.780	0.622
Recursive	0.331	0.232	0.461	0.346	0.610	0.490	0.729	0.559

NS2	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.381</b>	<b>0.278</b>	<b>0.530</b>	<b>0.414</b>	<b>0.704</b>	<b>0.495</b>	<b>0.873</b>	<b>0.620</b>
<b>LAR</b>	<u>0.399</u>	<u>0.291</u>	<u>0.568</u>	<u>0.426</u>	<u>0.726</u>	<u>0.552</u>	<u>1.024</u>	<u>0.698</u>
5-year Rolling	0.453	0.343	0.893	0.714	1.570	1.922	3.026	2.568
10-year Rolling	0.434	0.321	0.782	0.599	1.234	1.025	1.889	1.619
Recursive	0.431	0.320	0.772	0.600	1.203	1.013	1.853	1.601

NS3	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.722</b>	<b>0.518</b>	<b>1.017</b>	<b>0.750</b>	<b>1.179</b>	<b>0.860</b>	<b>1.124</b>	<b>0.858</b>
<b>LAR</b>	<u>0.777</u>	<u>0.550</u>	<u>1.103</u>	<u>0.809</u>	<u>1.292</u>	<u>0.945</u>	<u>1.299</u>	<u>0.989</u>
5-year Rolling	0.889	0.665	1.497	1.125	2.171	1.693	3.204	2.683
10-year Rolling	0.863	0.644	1.394	1.039	1.913	1.496	2.529	2.085
Recursive	0.868	0.647	1.442	1.075	1.992	1.537	2.660	2.206

Notes for Table 1 and 2:

- 1) The 1-month ahead forecast period is 1998:1 to 2010:9, a total of 153 months. The 3-month ahead forecast period is 1998:3-2010:9 with 151 months. The 6-month ahead forecast period is 1998:6-2010:9 with 148 months and the 12-month ahead forecast period is 1998:12-2010:9 with 142 months.
- 2) For each column, the best forecast, i.e., smallest RMSE or MAE, is marked in bold-face; the second best is underlined.

**Table 2. Forecast Comparison of Yields**

y (3)	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	0.331	0.231	<b>0.429</b>	<u>0.316</u>	<u>0.612</u>	<b>0.408</b>	<b>0.662</b>	<b>0.452</b>
<b>LAR</b>	<u>0.297</u>	<u>0.199</u>	<u>0.436</u>	<b>0.309</b>	<b>0.589</b>	<u>0.427</u>	<u>0.771</u>	<u>0.514</u>
5-year Rolling	0.346	0.237	0.722	0.529	1.345	1.066	2.871	2.339
10-year Rolling	0.326	0.219	0.637	0.447	1.077	0.860	1.814	1.568
Recursive	0.318	0.208	0.626	0.460	1.048	0.878	1.778	1.577
Random Walk	<b>0.254</b>	<b>0.153</b>	0.547	0.363	0.927	0.673	1.650	1.306

y (12)	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	0.259	<u>0.196</u>	<b>0.392</b>	<u>0.296</u>	<u>0.576</u>	<b>0.397</b>	<b>0.602</b>	<b>0.460</b>
<b>LAR</b>	<b>0.244</b>	<b>0.176</b>	<u>0.408</u>	<b>0.293</b>	<b>0.551</b>	<u>0.412</u>	<u>0.685</u>	<u>0.523</u>
5-year Rolling	0.280	0.216	0.675	0.508	1.264	1.020	2.608	2.150
10-year Rolling	0.260	0.197	0.601	0.458	1.024	0.833	1.722	1.511
Recursive	0.269	0.206	0.622	0.498	1.042	0.884	1.732	1.544
Random Walk	<u>0.251</u>	<b>0.176</b>	0.561	0.401	0.927	0.699	1.582	1.284

y (36)	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.296</b>	0.229	<b>0.434</b>	<b>0.331</b>	<u>0.542</u>	<u>0.420</u>	<b>0.494</b>	<b>0.378</b>
<b>LAR</b>	<u>0.300</u>	<b>0.211</b>	<u>0.463</u>	<u>0.348</u>	<b>0.533</b>	<b>0.405</b>	<u>0.532</u>	<u>0.413</u>
5-year Rolling	0.330	0.256	0.640	0.497	1.045	0.802	1.954	1.559
10-year Rolling	0.326	0.253	0.627	0.478	0.955	0.754	1.478	1.234
Recursive	0.330	0.256	0.650	0.507	0.981	0.781	1.486	1.260
Random Walk	<u>0.300</u>	<u>0.228</u>	0.573	0.448	0.845	0.679	1.237	1.026

y (60)	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.298</b>	<u>0.230</u>	<b>0.414</b>	<b>0.322</b>	<b>0.486</b>	<b>0.382</b>	<b>0.459</b>	<b>0.360</b>
<b>LAR</b>	0.309	<b>0.228</b>	<u>0.442</u>	<u>0.341</u>	<u>0.497</u>	<u>0.393</u>	<u>0.469</u>	<u>0.378</u>
5-year Rolling	0.335	0.259	0.575	0.453	0.873	0.673	1.487	1.166
10-year Rolling	0.340	0.265	0.592	0.459	0.859	0.680	1.258	1.063
Recursive	0.342	0.266	0.608	0.471	0.878	0.694	1.252	1.044
Random Walk	<u>0.301</u>	0.231	0.532	0.429	0.753	0.607	1.005	0.829

y (120)	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
<b>LARX</b>	<b>0.276</b>	<u>0.219</u>	<b>0.343</b>	<b>0.269</b>	<b>0.387</b>	<b>0.304</b>	<u>0.425</u>	<u>0.330</u>
<b>LAR</b>	<u>0.284</u>	<b>0.217</b>	<u>0.363</u>	<u>0.281</u>	<u>0.404</u>	<u>0.329</u>	<b>0.403</b>	<b>0.322</b>
5-year Rolling	0.295	0.230	0.427	0.354	0.587	0.481	0.822	0.656
10-year Rolling	0.290	0.221	0.437	0.354	0.601	0.472	0.828	0.692
Recursive	0.299	0.226	0.452	0.365	0.622	0.494	0.818	0.664
Random Walk	0.287	0.213	0.443	0.355	0.595	0.485	0.718	0.566

**Table 3. Average Lengths of Stable Sample Intervals Detected by LAR and LARX**  
(Unit: month)

NS factor	h = 1		h = 3		h = 6		h = 12	
	LAR	LARX	LAR	LARX	LAR	LARX	LAR	LARX
NS1	73	43	29	31	33	33	37	40
NS2	34	31	25	28	28	30	31	31
NS3	28	28	26	27	27	28	31	29

**Figure 1. Time Plot of the NS Level Factor and its Sample Autocorrelation Function**

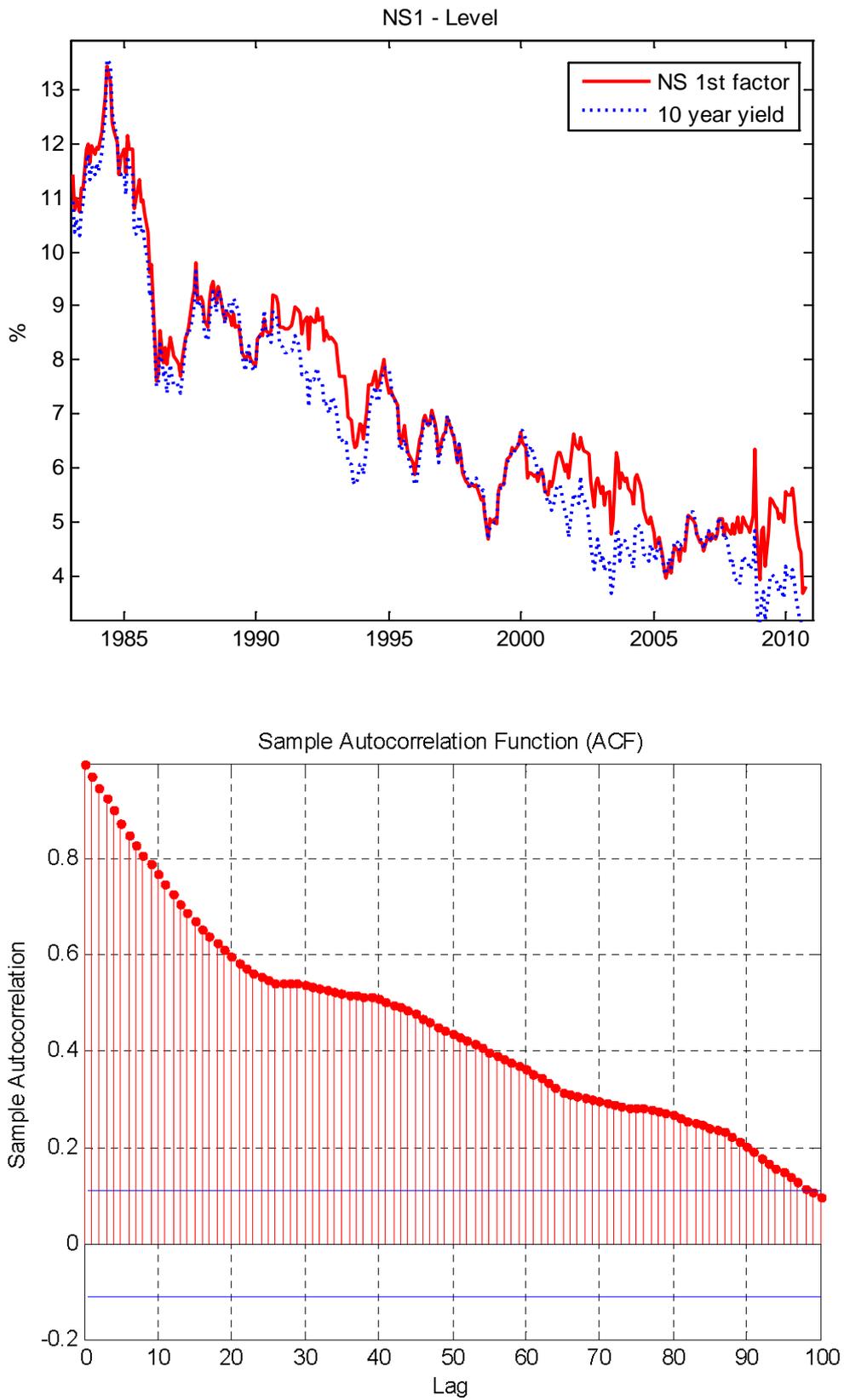
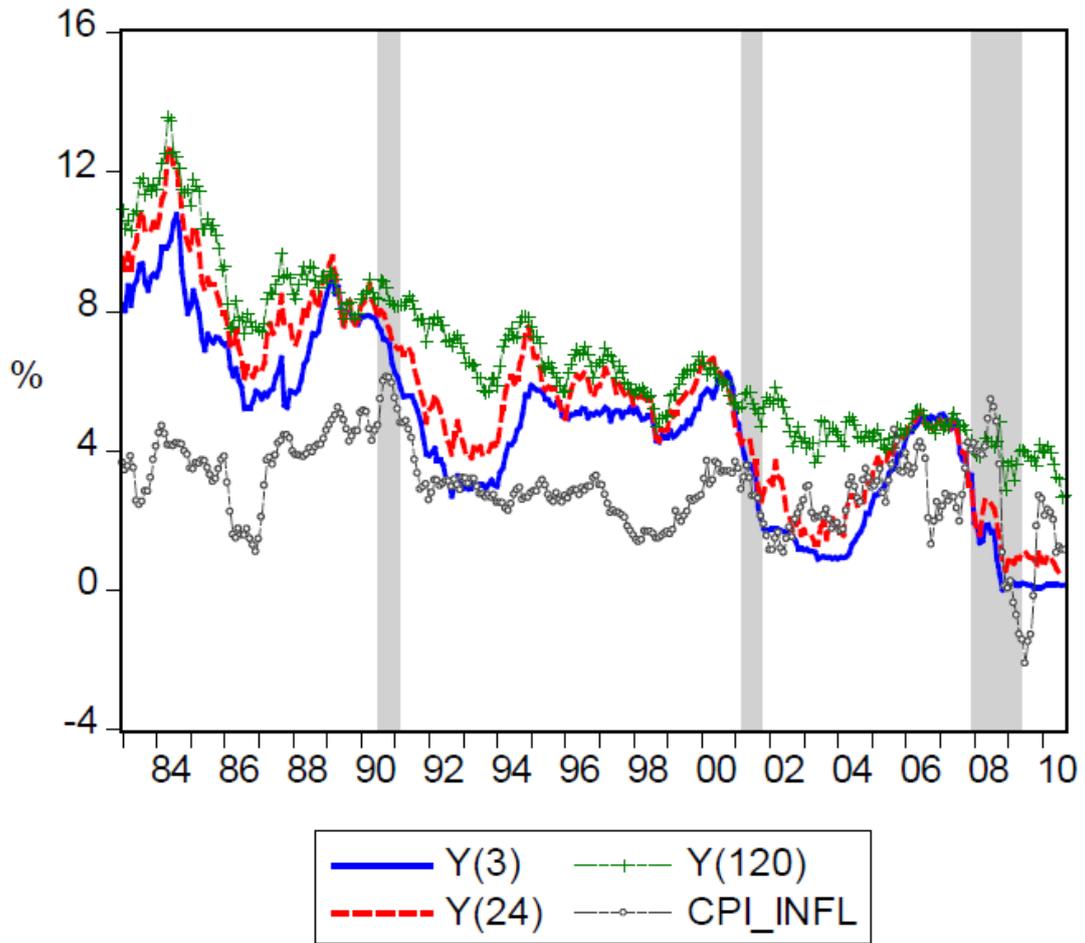


Figure 2. Three Yields and CPI Inflation in the US (1983.1-2010.9)



**Figure 3. US Yield Curve (1983.1-2010.9)**

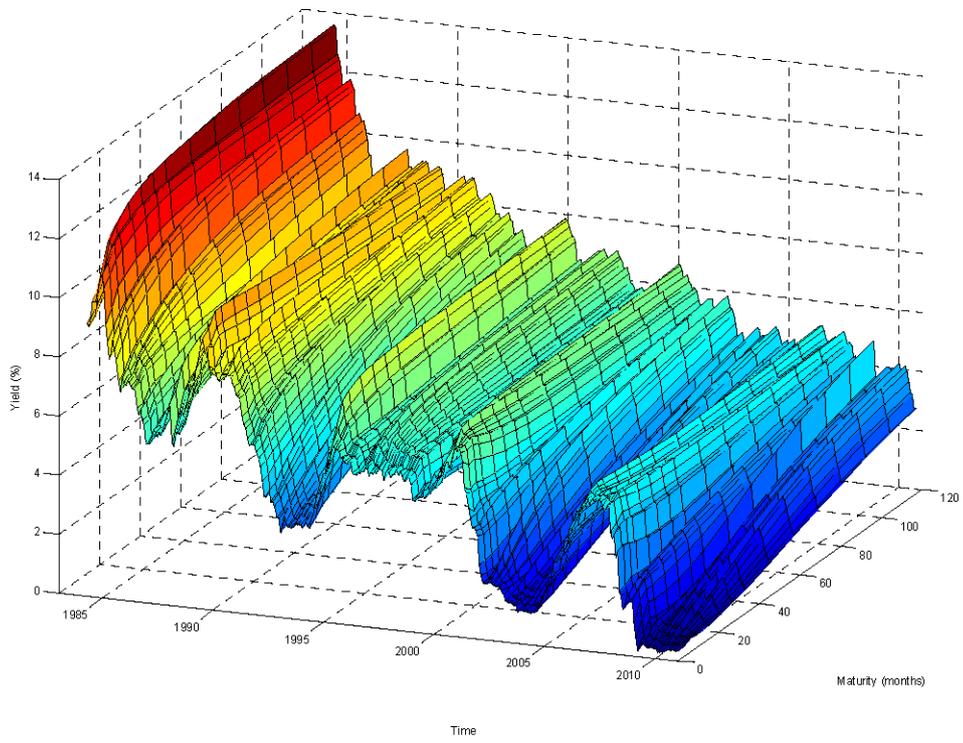
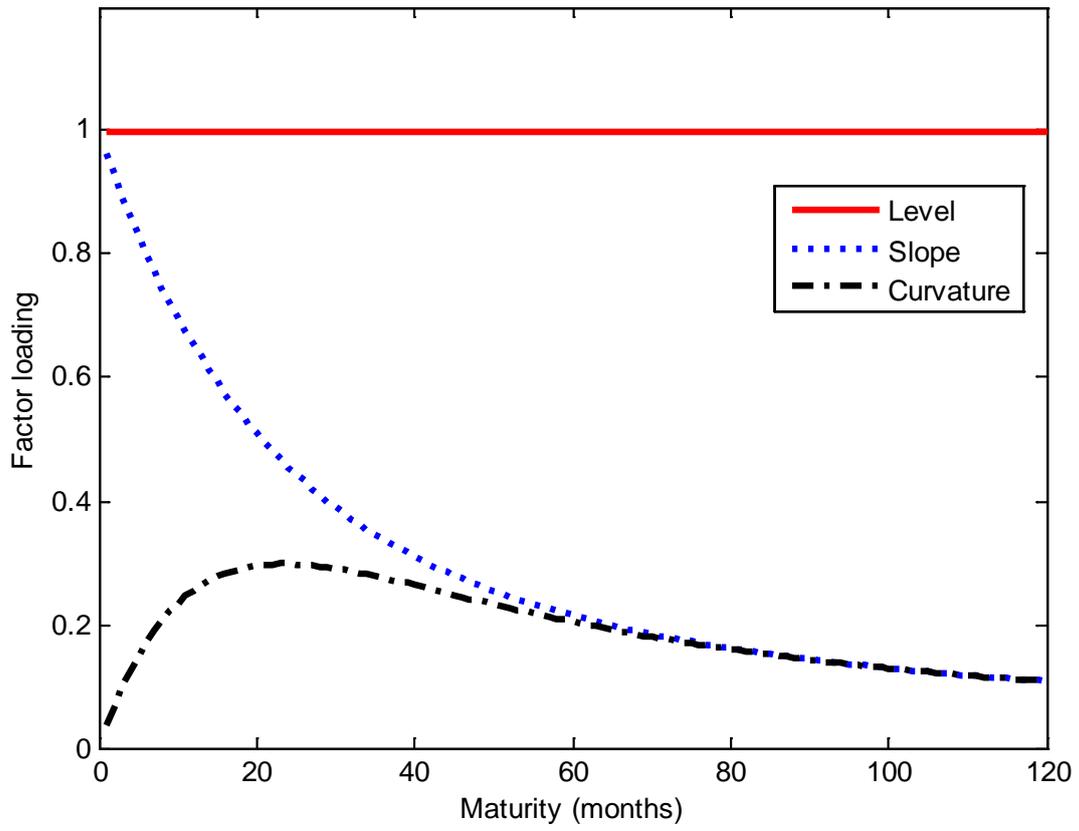


Figure 4. Nelson-Siegel Factor Loadings



**Figure 5. Time Evolution of Nelson-Siegel Factors Extracted from US Yield Curve (1983.1-2010.9)**

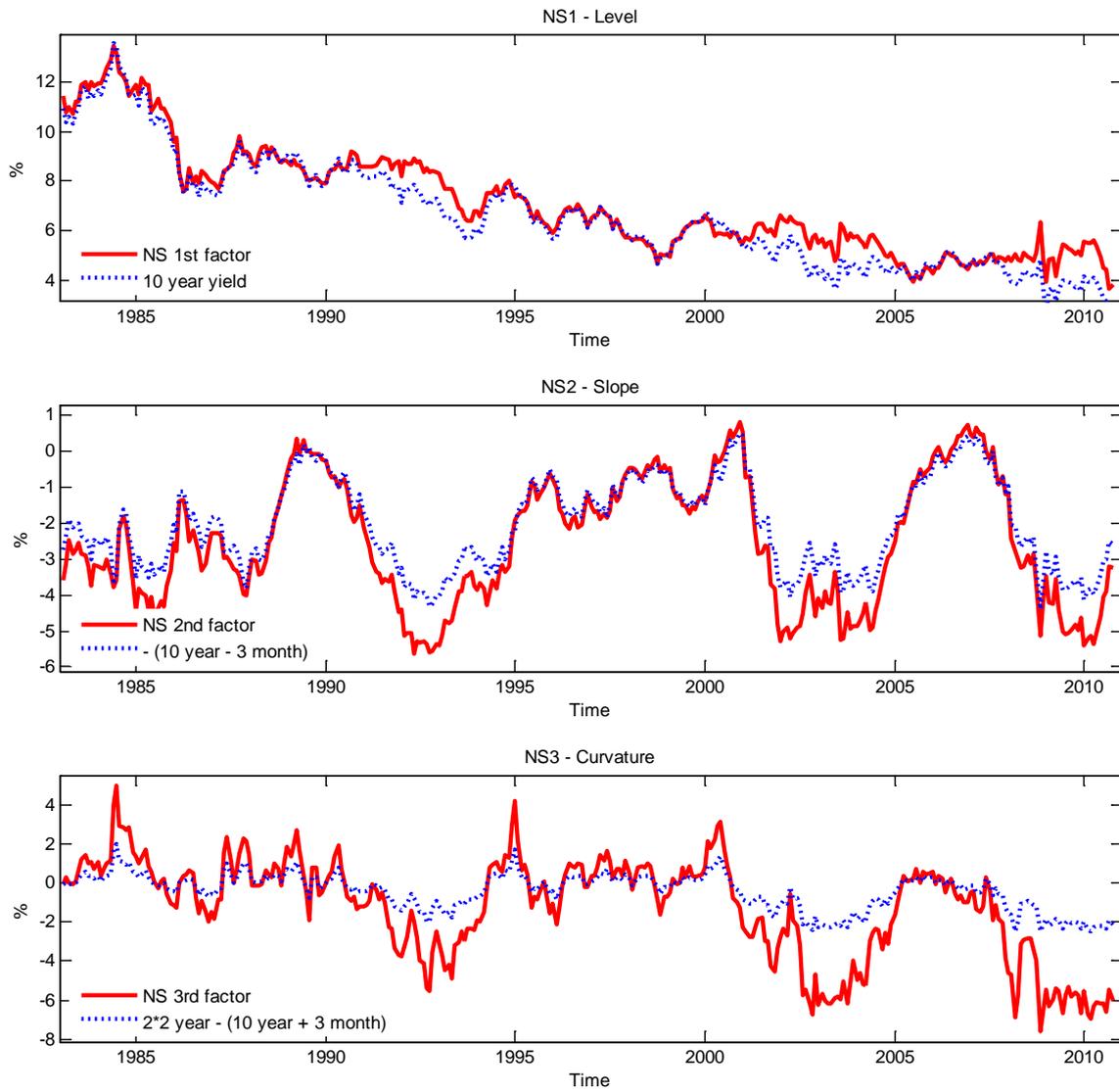
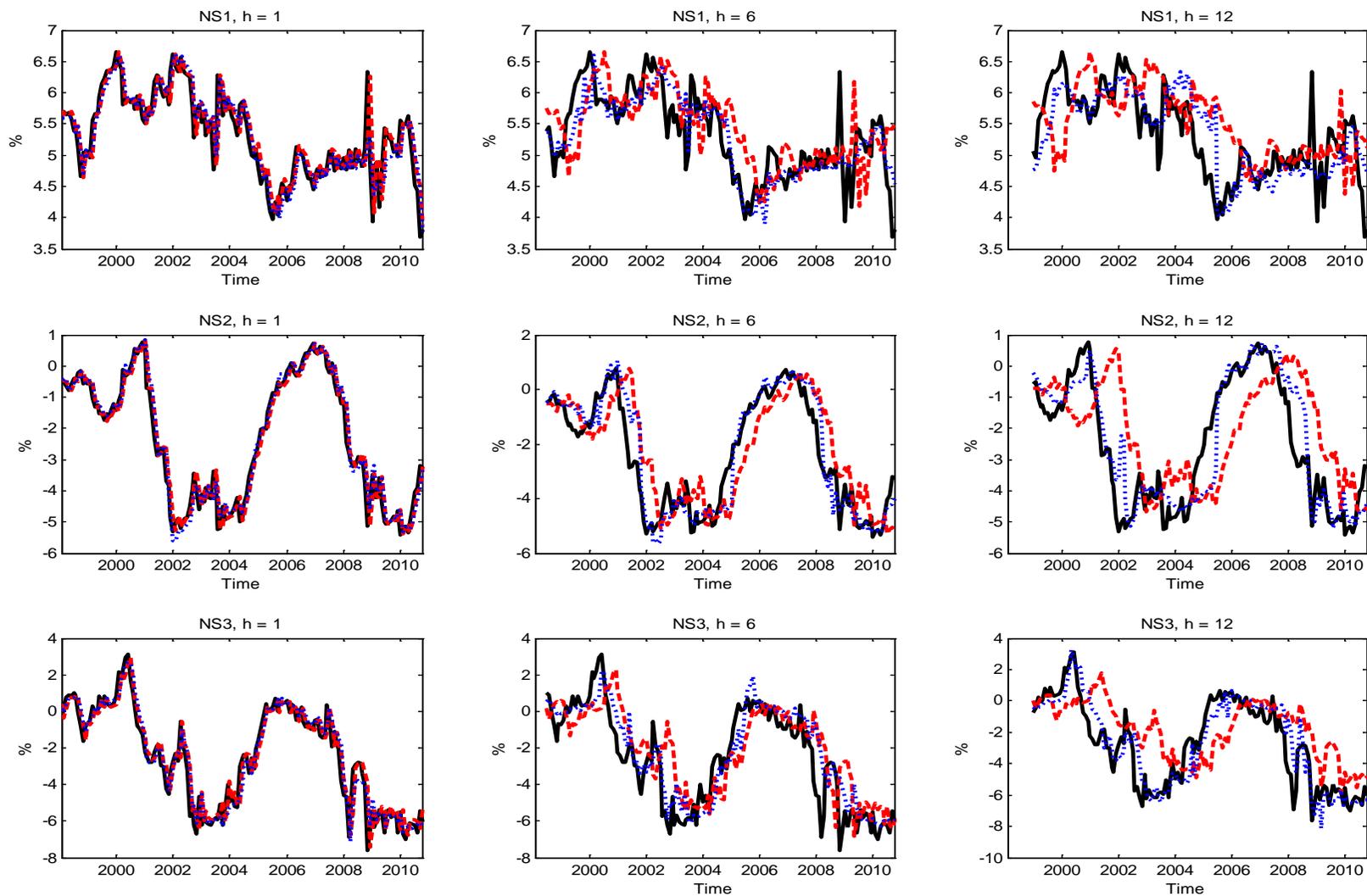
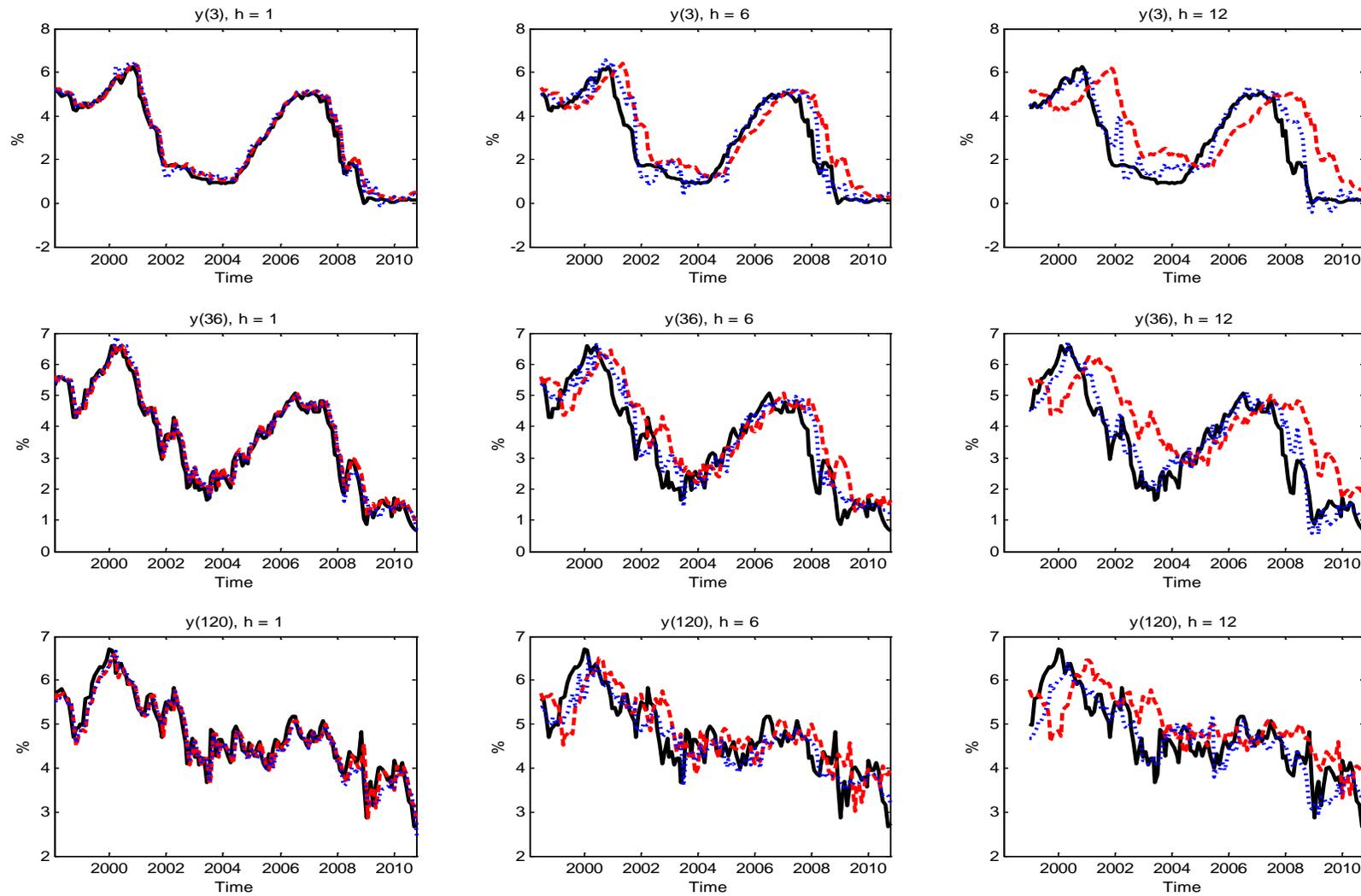


Figure 6. Plots of Factor Forecasts with LARX and 10-year Rolling



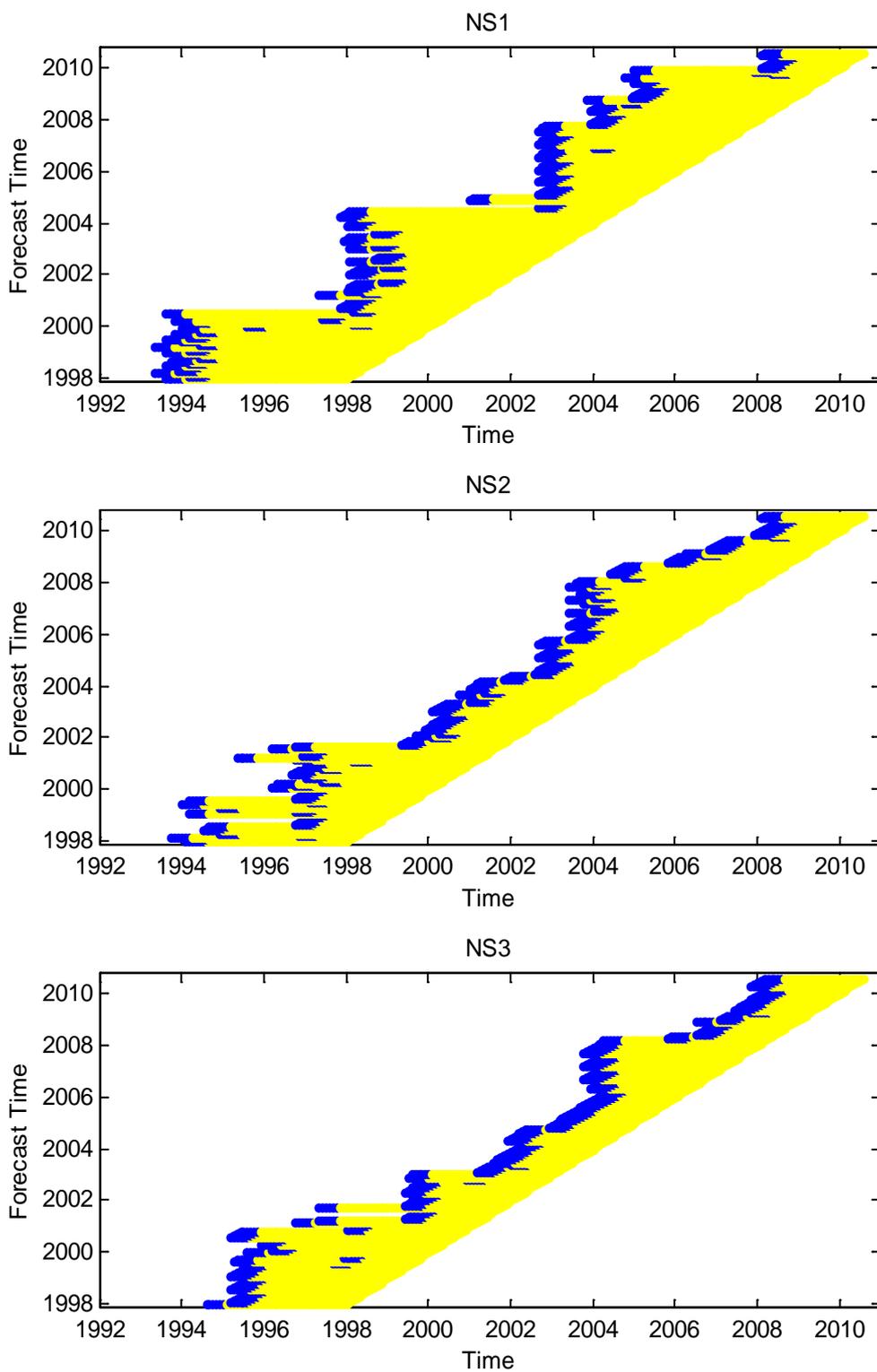
Note: Factors extracted from real yields are displayed by solid lines; LARX forecasts are displayed by dotted lines, and the 10-year rolling forecasts dashed lines.

Figure 7. Plots of Yield Forecasts with LARX and 10-year Rolling



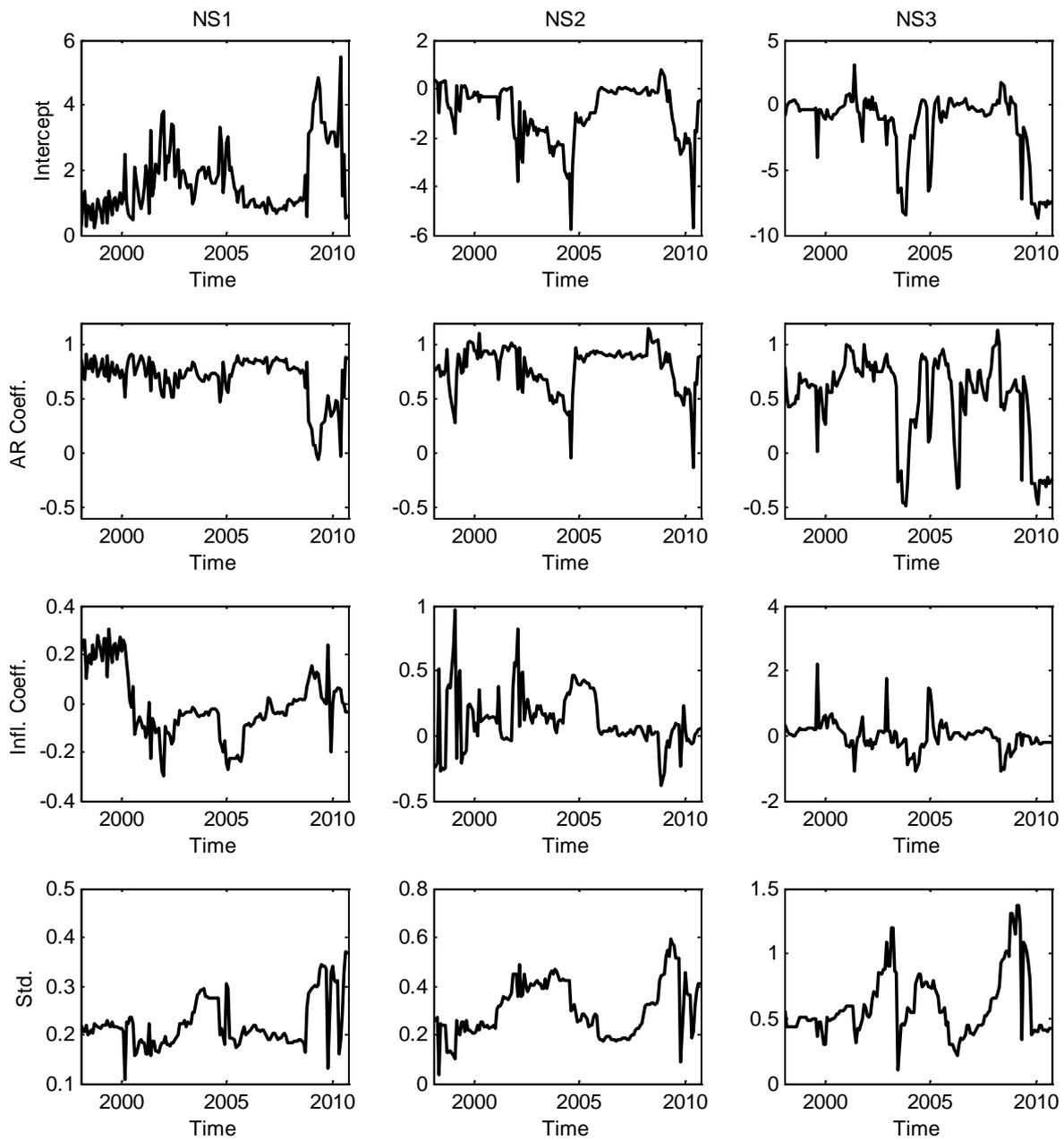
Note: Factors extracted from real yields are displayed by solid lines; LARX forecasts are displayed by dotted lines, and the 10-year rolling forecasts dashed lines.

**Figure 8. Breaks of Factors Detected in the LARX(1) Model**



Notes: The vertical axis marks the time when the estimation is made. The selected stable sample interval is marked horizontally as a light yellow line. The dark blue line marks the interval during which the most recent break is detected.

**Figure 9. Parameter Evolvement in the LARX Model**



Notes: In each column we plot the time evolvement of the estimated coefficients for the ARX(1) model of each factor. The first row displays the constant intercept; the second row is the coefficient of the lagged factor; the third row is the coefficient on inflation; the bottom row is the standard deviation of the error term.