

Adaptive Dynamic Nelson-Siegel Term Structure Model with Applications

Ying Chen^a Linlin Niu^{b,c*}

^aDepartment of Statistics & Applied Probability, National University of Singapore

^bWang Yanan Institute for Studies in Economics (WISE), Xiamen University

^cMOE Key Laboratory of Econometrics, Xiamen University

Abstract

We propose an Adaptive Dynamic Nelson-Siegel (ADNS) model to adaptively detect parameter changes and forecast the yield curve. The model is simple yet flexible and can be safely applied to both stationary and nonstationary situations with different sources of parameter changes. 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 popular reduced-form and affine term structure models; compared to random walk prediction, the ADNS steadily reduces the forecast error measurements by between 20 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 A306, Economics Building, Xiamen University, Xiamen, 361005, Fujian, China. Email: llniu@xmu.edu.cn. Phone: +86-592-2182839. Fax: +86-592-2187708.

1 Introduction

Yield curve forecasting is very important, and there are different approaches which can be generally categorised into two strands: the class of no-arbitrage and equilibrium models with theoretical underpinnings, and reduced-form models based on a data-driven statistical approach. The former class is indispensable in studying risk premia and pricing derivatives; however, it is often found to forecast poorly, compared with a simple random walk model (see Duffee, 2002; Duffee, 2011). The latter class has evolved along time from univariate time series models to multivariate time series models, and to some recent advances of dynamic factor models. The modeling approach in this paper falls into the second class, and we will briefly review the literature of yield curve forecast along this line.

The univariate class includes the random walk model, the 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. This type of models 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, as the included maturities of yields in the multivariate models increase to incorporate more information, this type of models is subject to the curse of dimensionality and, hence, cannot provide a complete view of the whole curve.

Alternatively, the recent advances of factor models enables the modeling and forecasting of the yield curve as a whole, where essential factors are extracted from 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 (NS) 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 tool for forecasting the yield curve. Instead of directly forecasting the yield curve, the DNS model extracts three factors via Nelson-Siegel factor loadings. These 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 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, the factors, like many other financial variables, are very persistent, which makes the recommended AR (1) model deficient in capturing the dynamic features of the Nelson-Siegel factors. 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 Section 3.1. 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 with 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 interest rate literature, several models of the short memory view 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 short memory 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 cycle 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, 24 and 120 months. The dashed line with circles shows CPI inflation, which is an important macroeconomic factor affecting nominal interest rates. The shaded bars mark the three recessions identified by the National Bureau of Economic Research (NBER) 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 intuitively that there may exist subsamples where the interest rates are approximately stationary and that can be modeled sufficiently well with simple autoregressive processes.

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

In the volatility forecasting literature, Mercurio and Spokoiny (2004) propose an adaptive approach where the persistent volatility is modeled as a varying mean process. With the short memory view, the structural shifts are detected in a likelihood-based testing procedure, which selects large subsamples of constant volatility, but enables switching to smaller sample sizes if a structural change is detected. The approach is shown to improve information efficiency and performs better than the GARCH model in out-of-sample forecast. Extending the approach, Chen, Härdle and Pigorsch (2010) propose a local AR (LAR) model, that allows for time-varying parameters, to capture the seemingly persistent volatility. 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 and smoothed changes. In terms of out-of-sample prediction, it performs better than several alternative models including long memory models and some regime switching models. Our study employs the methodology in Chen et al. (2010) to detect stationary subsample divisions of the yield curve in real time. 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 an LAR model with exogenous variables (LARX) is adopted to fit each state factor dynamically, although the exogenous variables do not span the yield curve cross-sectionally. The rationale is that, to conduct an estimation at any point of time, we take all available past information but only consider a subsample that includes all the recent observations that can be well represented by a local stationary model with approximately constant parameters. In other words, there are not any significant parameter 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 actually a DNS model over the detected stable subsample, where the dynamics of each factor are described by an AR(1) process with exogenous variables (ARX). The adaptive procedure of

stable subsample detection is crucial to distinguishing this model from the DNS model with predetermined sample lengths. We investigate the performance of this proposed model in a real data analysis, and compare it with popular alternative models including the DNS, the random walk and several representative affine arbitrage-free term structure 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 evolution.

For a brief summary on our modeling features, we combine the Nelson-Siegel cross-section factorization with a varying-parameter autoregressive state process to model the yield curve with a forecast objective. To deal with the non-stationary characteristics of the yield data for better predictability, we consciously choose a data-driven statistical approach to detect the local homogenous sample interval with a backward testing procedure under a parsimonious and flexible model setup. Compared to the existing data-driven criteria on model selection under prefixed sample lengths, such as the AIC (Akaike, 1973) and the BIC (Schwarz et al., 1978), the adaptive approach chooses optimal sample lengths conditional on model specifications. Instead of adopting everywhere a fixed sample size or utilizing all available sample information (e.g. Diebold and Li, 2006; Mönch, 2008), the approach chooses for each particular forecast origin an optimal sample size to avoid serious estimation errors and achieve the best possible accuracy of estimation. Compared to the popular structural break test (Lee and Strazicich, 2003) or regime shift model (e.g. Ang and Bekaert, 2002; Bansal and Zhou, 2002) which considers the whole sample period, the adaptive testing procedure works backward only up to the point where the longest possible homogenous sample ends, which suits well the forecast purpose. A related assumption in our approach is that the local model of homogenous parameters will hold with high probability for the forecast horizon. We may wish to predict structural changes out of sample, as do 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, less computationally demanding and, as will be shown in the results, it is very effective in detecting non-synchronized breaks and improving predictability.

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 analyzes the forecast effectiveness of the LARX model compared with alternative rolling window techniques with Monte Carlo experiments. Section 5 reports the real data analysis and a forecast comparison with alternative models. Section 6 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 release of selected interest rates. The remaining yields with maturities of integer years are taken from publicly available 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 to emphasize the fit for mid-range yields. 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. The US Yield Curve (1983:1 to 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, our primary goal is to forecast the yield curve, so we opt for the simple and effective method of taking inflation as an exogenous variable in the state dynamics. It is also possible and beneficial to include other macroeconomic variables or factors in the state dynamics. We relegate this investigative task of defining the optimal set of variables or factors included for future work. Our focus remains using inflation as a relevant macroeconomic variable to illustrate the flexibility and 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 given a fixed shape parameter. For each single factor, an 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 globally 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 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 τ (in months) at time t . The three factors, β_{1t} , β_{2t} and β_{3t} , are denoted as level, slope and curvature, respectively. Parameter λ 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 λ . In our study, we follow Diebold and Li (2006) to set $\lambda = 0.0609$ which maximizes the curvature loading at a medium maturity of 30 months. Under a fixed λ , 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.

Factor loadings on the yield curve are displayed in Figure 4. The loadings of β_{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 β_{2t} start from 1 at the instantaneous spot rate and decay towards zero as maturities increase. The average value of β_{2t} is negative, indicating a normal upward-sloping yield curve; when β_{2t} is positive, it leads to a downward slope of the yield curve which often anticipates recessions. The slope factor, β_{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 β_{3t} display a humped shape, which peaks around the maturities of medium duration. Thus β_{3t} is named a curvature factor. A positive value of β_{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 the 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 solid lines 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 over 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 the US Yield Curve]

3.2 Local autoregressive model with exogenous variables

Now, for each one-dimensional factor β_{it} , with $i = 1, 2$ and 3 , a local model is adopted to forecast. To simplify notations, we drop the subscript i 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 and thus the local homogeneity assumption no longer holding. A sequential testing procedure 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 good out-of-sample forecasting ability. Moreover, to increase modeling flexibility, we allow relevant exogenous variables in addition to the lagged autoregressive component. The proposed model is a local AR(1) model with exogenous variables, or a LARX(1) model, for each Nelson-Siegel factor.

3.2.1 The LARX(1) model and estimator

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

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

where $\theta_t = (\theta_{0t}, \theta_{1t}, \theta_{2t}, \sigma_t)^{\top}$. The innovation μ_t has a mean of zero and a variance of σ_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) and constant parameter θ_s . Then the local maximum likelihood estimator $\tilde{\theta}_s$ is defined as:

$$\begin{aligned}\tilde{\theta}_s &= \arg \max_{\theta_s \in \Theta} L(\beta; I_s, \theta_s, X) \\ &= \arg \max_{\theta_s \in \Theta} \left\{ -m_s \log \sigma_s - \frac{1}{2\sigma_s^2} \sum_{t=s-m_s+1}^s (\beta_t - \theta_{0s} - \theta_{1s}\beta_{t-1} - \theta_{2s}X_{t-1})^2 \right\}\end{aligned}$$

where Θ is the parameter space and $L(\beta; I_s, \theta_s, X)$ is the local log-likelihood function.

3.2.2 The testing procedure for homogeneous intervals

In practice, the stable subsample is unknown and the number of possible candidates is large, e.g. as many subsamples as there are past 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 with discrete increment of M periods ($M > 1$) between any two adjacent subsamples. As a consequence, there are K_s candidate subsamples for any particular time point s , or

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

which starts from the shortest subsample, $I_s^{(1)}$, where the ARX(1) model with constant parameters should provide a reasonable fit and the assumption of the local homogeneity is accepted by default.

We use $\hat{\theta}_s^{(k)}$ to denote the adaptive estimator in the k -th subsample. For the first subsample $I_s^{(1)}$, we have $\hat{\theta}_s^{(1)} = \tilde{\theta}_s^{(1)}$, where the maximum likelihood (ML) estimator is accepted under the local homogeneity assumption. The selection procedure then iteratively extends the subsample with the next interval of M more periods and sequentially tests for possible structural changes in the next longer subsample. This is performed via a kind of likelihood ratio test. The test statistic is defined in each following subsample $I_s^{(k)}$ as

$$T_s^{(k)} = |L(I_s^{(k)}, \tilde{\theta}_s^{(k)}) - L(I_s^{(k)}, \hat{\theta}_s^{(k-1)})|^{1/2}, \quad k = 2, \dots, K \quad (3)$$

where $L(I_s^{(k)}, \tilde{\theta}_s^{(k)}) = \max_{\theta_s \in \Theta} L(\beta; I_s^{(k)}, \theta_s, X)$ denotes the fitted likelihood under *hypothetical* homogeneity and $L(I_s^{(k)}, \hat{\theta}_s^{(k-1)}) = L(\beta; I_s^{(k)}, \hat{\theta}_s^{(k-1)}, X)$ refers to the likelihood in the current testing subsample, with the estimate under the previously *accepted* local homogeneity. The test statistic measures the difference between these two estimates. A set of critical values ζ_1, \dots, ζ_K is used to measure the significance level. If the difference is significant,

i.e. $T_s^{(k)} > \zeta_k$, it indicates that the model changes more than what would be expected due to the sampling randomness. If this is the case, then the procedure terminates and the latest accepted subsample $I_s^{(k-1)}$ is selected, such that $\hat{\theta}_s^{(k)} = \hat{\theta}_s^{(k-1)} = \tilde{\theta}_s^{(k-1)}$. Moreover, we have $\hat{\theta}_s^{(\ell)} = \tilde{\theta}_s^{(k-1)}$, for $\ell \geq k$, which means that the adaptive estimator for longer subsample is the ML estimate over the identified longest subsample of local homogeneity. Otherwise for $T_s^{(k)} \leq \zeta_k$, we accept the current subsample $I_s^{(k)}$ as being homogeneous and update the adaptive estimator $\hat{\theta}_s^{(k)} = \tilde{\theta}_s^{(k)}$. We continue the procedure until a change is found or the longest subsample, $I_s^{(K)}$, is reached under local homogeneity.

The critical values are crucial in the testing procedure. However as the sampling distribution of the test statistic is unknown, even asymptotically, we therefore calibrate the critical values via Monte Carlo experiments.

3.2.3 Critical value calibration

In the Monte Carlo simulation for critical value calibration, we mimic a situation of ideal time homogeneity in the simulated samples. In this environment, the set of critical values should be able to provide a prescribed performance of the testing procedure under the null hypothesis of time homogeneity.

In particular, we generate a globally homogeneous ARX(1) time series, where the modeling parameters are constant:

$$\beta_t = \theta_0^* + \theta_1^* \beta_{t-1} + \theta_2^* X_{t-1} + \mu_t, \quad \mu_t \sim N(0, \sigma^{*2})$$

where $\theta^* = (\theta_0^*, \theta_1^*, \theta_2^*, \sigma^*)$ for $t = 1, \dots, T$. In this case, time homogeneity is ensured everywhere, and hence the ML estimate $\tilde{\theta}_t^{(k)}$ in every subsample $I_t^{(k)}$, $k = 1, \dots, K$, is optimal. The estimation error can be measured by the fitted log-likelihood ratio:

$$R_k = E_{\theta^*} \left| L \left(I_t^{(k)}, \tilde{\theta}_t^{(k)} \right) - L \left(I_t^{(k)}, \theta^* \right) \right|^{1/2}, \quad (4)$$

where R_k can be computed numerically, with the knowledge of θ^* .

Suppose a set of critical values is given, then one obtains the adaptive estimator $\hat{\theta}_t^{(k)}$ as described before. The temporal difference between the ML estimator and the adaptive estimator, denoted as $D_s^{(k)}$, can be measured by the log likelihood ratio as well:

$$D_t^{(k)} = \left| L \left(I_t^{(k)}, \tilde{\theta}_t^{(k)} \right) - L \left(I_t^{(k)}, \hat{\theta}_t^{(k)} \right) \right|^{1/2}.$$

Depending on the critical values with which we compare the test statistic $T_s^{(k)}$, we obtain two outcomes:

- If $T_t^{(k)} \leq \zeta_k$, we accept $\hat{\theta}_t^{(k)} = \tilde{\theta}_t^{(k)}$, and $D_t^{(k)} = 0$;
- If $T_t^{(k)} > \zeta_k$, we set $\hat{\theta}_s^{(k)} = \hat{\theta}_s^{(k-1)} = \tilde{\theta}_s^{(k-1)}$, and $D_t^{(k)} = \left| L \left(I_t^{(k)}, \tilde{\theta}_t^{(k)} \right) - L \left(I_t^{(k)}, \hat{\theta}_t^{(k-1)} \right) \right|^{1/2}$.

The idea is to find the set of critical values that leads to a performance as good as the true underlying characteristics under the null of time homogeneity. In other words, it requires that the stochastic distance $D_t^{(k)}$ is bounded by the ideal estimation error R_k in Equation (4):

$$E_{\theta^*} \left(D_t^{(k)} \right) = E_{\theta^*} \left| L \left(I_t^{(k)}, \tilde{\theta}_t^{(k)} \right) - L \left(I_t^{(k)}, \hat{\theta}_t^{(k)} \right) \right|^{1/2} \leq R_k. \quad (5)$$

With this inequality as the risk bound, the critical values, the only unknown parameters in Equation (5), can be computed. With too large a critical value, there is a higher probability of accepting subsamples everywhere, and hence obtaining a lower value of $D_t^{(k)}$, and it becomes easier to accept the fitted model over longer subsamples. Though the bound, Equation (5), is definitely fulfilled under the simulated time homogeneity, it is insensitive to changing parameters. On the contrary, too small a critical value implies a more stringent test that unnecessarily favors shorter subsamples, discarding useful past observations and resulting in higher parameter uncertainty. Thus the optimal critical values are the minimum values to just accept the risk bound at each subsample. It appears that such a choice is not only optimal in this artificially homogeneous setting, but also leads to nearly optimal forecast performance in a general case with parameter changes, as will be demonstrated in Section 4 with a simulation study and in Section 5 with a real data forecast and analysis of the yield curve.

The computation of critical values relies on a set of hyperparameters (θ^*, K, M) . Ideally, θ^* should be close to the true parameter θ_s^* underlying the real data series at each point of time, which is actually the target of our estimation. In the following numerical analysis, we consider θ^* as the ML estimate using information from the available real sample of 1983:1 to 1997:12 before the forecast exercise starts. We use θ^* to simulate N series of data, each of length $T = 120$, and calibrate the set of critical values as described above. The same set of calibrated critical values are adopted for every time point throughout the real-time estimation and forecast. In our procedure, at each point of time, we consider $K = 20$ subsamples for the test procedure, with the increment of $M = 6$ months between any adjacent subsamples, i.e., 120 months (10 years) is the maximal subsample size. We find that the technique is robust to the selection of the hyperparameters (θ^*, K, M) , as will be illustrated in Section 4.2. There is no significant difference in forecast accuracy for possible misspecifications in θ^* or for different sets of subintervals determined by K and M .

3.2.4 Real-time estimates

Following the test and estimation procedure described above, we estimate the LARX(1) model for the factor dynamics over time from 1997:12 onwards, and plot the evolution of parameter estimates in Figure 6. For each of the three factors, the ARX process involves four parameters as in Equation (2): the intercept, θ_{0t} ; the autoregressive coefficient, θ_{1t} ; the coefficient of inflation, θ_{2t} ; and the standard deviation of the error term, σ_t . We report in each column the evolution of the four parameters for each factor.

[Figure 6. Parameter Evolution in the LARX Model]

Figure 6 shows that the coefficients of the ARX(1) model of each Nelson-Siegel factor vary wildly over time, which justifies the application of an adaptive approach. Within the optimal interval detected at each point in 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.

Examining the patterns of changes, parameters of slope and curvature factors are, in general, more volatile than those of the level factor. Consequently, the selected subsamples of slope and curvature are also relatively shorter than those for level. The average lengths of selected subsamples for level, slope and curvature are 36, 29 and 26, respectively. As for the timing of big shifts, we find that the outbreak of the recent financial crisis is a common shifting time for all three factors. In particular, the autoregressive coefficients sharply decline and the standard deviations of innovations surge rapidly. However, there are also individual shifts in one factor which do not present in the others. For example, the autoregressive coefficients for slope and curvature fall during 2004 and sharply revert in 2005, but the autoregressive coefficient for level is relatively stable during this period; the CPI inflation coefficient on the level factor became negative after 2000, but this shifting pattern with inflation is not obvious for either slope or curvature.

The observations that the timing of parameter changes do not frequently align with each other support our choice of an ARX process for each single factor instead of a joint VAR or VAR with exogenous variables (VARX) process for all factors. In a VAR or VARX setting, unless parameters have common shifting patterns, the test of joint homogeneity in the adaptive procedure tends to select a longer subsample under homogeneity than in the ARX setting, because the testing power decreases with a higher dimension of parameters. Since the dynamic interaction of these three factors is limited that off-diagonal parameters in the autoregressive matrix are often insignificant so as to contribute to lower the forecast errors, the longer selected sample and slower detection of instability will deteriorate the forecast performance. We will further illustrate with a real data application in Section 5. The

bottom line here is that the adaptive estimation procedure is designed to detect parameter changes within each sample; to extract the maximum benefit for forecast accuracy, the underlying model needs to be parsimonious per se to achieve higher information efficiency.

3.3 Modeling, forecasting factors and the yield curve

After the extraction of factors and estimation of the local state dynamics over the selected subsample, the model can be used for prediction.

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

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon_t(\tau), \quad \epsilon_t(\tau) \sim (0, \sigma_{\epsilon, \tau}^2) \quad (6)$$

$$\beta_{it} = \theta_{0t}^{(i)} + \theta_{1t}^{(i)} \beta_{it-1} + \theta_{2t}^{(i)} X_{t-1} + \mu_t^{(i)}, \quad \mu_t^{(i)} \sim (0, \sigma_{it}^2), \quad i = 1, 2, 3. \quad (7)$$

At each point of time, once the most recent optimal subsample is selected for each factor, the resulting model amounts to a DNS model with an exogenous factor but possibly different subsample length for the individual ARX processes. However, we will show that, with the adaptive selection of the subsample length, the forecast accuracy is substantially improved over the DNS with prescribed sample length such as those with rolling or recursive windows.

We estimate two specifications of the ADNS model. One is denoted as LAR and has no exogenous factors, 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 or to the DNS model with a VAR(1) joint process for all three factors, and shows the stark difference in forecasts resulting from the adaptive feature. Compared with the LAR model, 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:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \hat{\beta}_{3,t+h/t} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (8)$$

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

$$\hat{\beta}_{i,t+h/t} = \hat{\theta}_{0t} + \hat{\theta}_{1t} \hat{\beta}_{i,t}, \quad i = 1, 2, 3, \quad (9)$$

and, for the LARX specification, by

$$\hat{\beta}_{i,t+h/t} = \hat{\theta}_{0t} + \hat{\theta}_{1t} \hat{\beta}_{i,t} + \hat{\theta}_{2t} X_t, \quad i = 1, 2, 3. \quad (10)$$

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, or 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 multi-step ahead forecast. In other words, we estimate the LAR or LARX model for each specific forecast horizon. As shown in Marcellino, Stock and Watson (2006), the direct forecast is more robust than the iterated forecast under model misspecification, which is the general situation for parsimonious models.

4 Monte Carlo study on the LARX model

Questions remain on whether the adaptive procedure performs well compared to alternative prescribed subsample selection methods such as rolling window estimation or recursive estimation, and whether the performance of the adaptive procedure is robust with respect to our choice of hyperparameters (θ^*, K, M) . Before the forecast exercise with real data, we address these issues first by performing a simulation study under known data generating processes.

We consider a series of simulation studies to investigate the performance of the LARX model. We first compare the forecast accuracy using our adaptive technique, as described in Section 3.2, to the alternatives based on the rolling window technique of various prefixed sample sizes. Since the forecast accuracy of the adaptive model depends on the critical values that further rely on the hyperparameters (θ^*, K, M) used in the Monte Carlo experiments, we also analyze the robustness of the choice of the hyperparameters, and the impact of the hyperparameter misspecifications on the forecast performance.

4.1 Forecast comparison with alternative methods

The adaptive LARX model is proposed for realistic situations with various sources of changes. Its performance should be both stable in a nearly time homogeneous situation and sensitive to regime shifts. In the simulation study, we consider two baseline scenarios, one with a set of globally constant parameters, i.e., a homogeneous scenario, denoted as HOM, and the other with time-varying parameters, specifically a regime-switching scenario, denoted as RS.

The parameters for our simulation are based on the characteristics of adaptive estimates from the first NS yield factor – the level factor, β_{1t} , as shown in Section 3.2. Empirically, the level factor is very close to the first principal component extracted from a data-driven approach – the principal component analysis, which is able to explain more than 75% variations of the yield data. Without loss of generality, we set the intercept of the simulated ARX model to zero, and simulate a series of mean zero i.i.d data as the exogenous variable. The

default parameter values are close to the averages of the adaptively estimated coefficients of the level factor, $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) = (0.698, -0.007, 0.224)$, where $\bar{\theta}_1$ is the AR coefficient, $\bar{\theta}_2$ is the ARX coefficient and $\bar{\sigma}$ is the volatility of innovations.

In the HOM scenario, the ARX coefficients are set to the default values and kept constant throughout the whole sample. In the RS scenario, we design four experiments and denote them as:

RS- A , where A denotes the autoregressive coefficient, θ_1 ;

RS- X , where X denotes the coefficient of the exogenous variable, θ_2 ;

RS- V , where V denotes the standard deviation of innovations, σ ;

RS- AV , where AV denotes both θ_1 and σ .

In the first three experiments of the RS scenarios, only the labeled parameters shift among four phases,

$$[\text{mean, low, mean, high}],$$

where the values of each phase are calibrated using the parameter estimates for the level factor, as shown in the first column of Figure 6. The fourth experiment of the RS scenario, shifting two parameters together, is motivated by the empirical observation that the autoregressive coefficient is negatively correlated with the volatility of innovations, as is evident in Figure 6 when comparing the second and fourth rows.

For each scenario and experiment, we simulate 1,000 series of 600 observations, given the time-dependent parameter sets as described in Table 1. The exogenous variable X_t is generated under the normal distribution according to characteristics of the demeaned CPI inflation, and is used in all simulations. The simplicity of the exogenous variable generation helps concentrate on the forecast performance of the adaptive technique. In the RS scenarios, each of the four phases lasts for 150 periods.

[Table 1. Parameters Used in the Simulated Scenarios]

For each of the simulated time series, from $t = 150$ to $t = 599$, we iteratively select the subsample, estimate the parameters, and calculate the one-step ahead forecast for $t = 151$ to 600. During the test procedure, we set $K = 20$ and $M = 6$ as the default choice, i.e., the interval set is given by $\mathbf{I}_t = \{I_t^k\}_{k=1}^{20} = \{6 \text{ months}, 1 \text{ year}, 1.5 \text{ years}, \dots, 10 \text{ years}\}$ with the longest subinterval 10 years. To calculate the critical values we use the default parameter $\theta^* = \bar{\theta} = (0.698, -0.007, 0.224)$ to simulate Monte Carlo experiments.

For forecast comparison, we employ rolling window strategies for the ARX model to make one-step ahead forecasts for the same period of $t = 151$ to 600. We allow twenty alternative

lengths of rolling windows, i.e., from I^1 (6 months) to I^K (10 years), which corresponds to our choice of subsample sets in the adaptive procedure. We compute the forecast root mean squared error (RMSE) and the mean absolute error (MAE) as measures of forecast accuracy.

Table 2 presents the RMSE and MAE of the adaptive technique and the alternatives. For ease of exposition we do not report all forecasting results of the rolling windows. Instead, for each scenario and experiment, we focus only on those rolling window choices that yield the best forecast accuracy (with a minimum RMSE or MAE, denoted as k^*) and the worst accuracy (with a maximum RMSE or MAE, denoted as $k!$). For the best and worst rolling window choices, the related index, k , is indicated in the parentheses, which corresponds to a window size of $k \times M$. The number of times that the LARX is superior to the 20 alternative window choices is highlighted in the “No. of Win” column.

[Table 2. Forecast Comparison of the Adaptive and Rolling Estimation in Simulated Scenarios]

The numerical results reveal that an adaptive approach with varying local window sizes introduces more flexibility into the procedure, leading to comparable performance as the optimal sample under the HOM scenario and generally better performance under regime shifts. More specifically, in the HOM scenario, the adaptive technique, though with a “wrong” assumption of time-varying coefficients, still provides reasonable accuracy. Under homogeneity with constant parameters, the optimal sample of the ARX must be the longest subsample I^K for which $k^* = K = 20$, as confirmed by the best rolling window selected for both RMSE and MAE. The adaptive technique, though not optimal, still results in similar accuracy to the best rolling window, with values of 0.252 versus 0.240 for RMSE and 0.190 versus 0.184 for MAE, respectively, and outperforms a few alternatives. In the regime-switching scenarios of RS-*A* with a time-varying autoregressive parameter or RS-*X* with a time-varying coefficient of the exogenous variable, our technique is superior to the majority of the 20 alternative rolling window estimation, winning 18 and 11 of the times for RMSE and 20 and 13 of the times for MAE, respectively. The relative performance in terms of RMSE is slightly weaker, as this measure penalizes large errors. For RS-*V* with time-varying volatility, our technique only outperforms 3 alternative rolling windows. This is to be expected as varying volatility leads to a changing signal-to-noise ratio that interrupts the detection of change from noise. In the more realistic RS-*AV* scenario where the shifts of the autoregressive parameter and volatility negatively correlate, both the adaptive procedure and the rolling window produce higher prediction errors. Nevertheless, due to its quick detection of changes in the autoregressive parameter, the adaptive technique outperforms 14 of 20 and 16 of 20 rolling window alternatives for RMSE and MAE, respectively.

The direct comparison of the LARX forecasts with those based on a constant window size confirms that the selection of the local subsample is, indeed, important. The advantage of the adaptive technique is that an optimal subsample is reasonably selected at each time point. To illustrate the dynamic process, we take the RS-*A* scenario to plot the average lengths of selected subsamples along the vertical axis in Figure 7, and the time of estimation on the horizontal axis. The estimation starts from $t = 150$ and the first forecast is made for $t = 151$. Parameter shifts happen at $t = 151, 301$ and 451 , as highlighted by the vertical dashed lines.

[Figure 7. Average Lengths of Selected Subsamples]

The figure shows a typical three-stage behavior of the subsample selection between two breaks of the varying parameter.

The first stage is a short period of inertia before the change is detected, and longer subsamples are chosen. This is because, immediately after a change occurs, the change point is contained in I^1 , for which we accept homogeneity by assumption. It is difficult for the test of homogeneity between I^2 and I^1 to reject the null due to the lack of sufficient information. As the interval is extended further backward, it becomes more difficult to reject the null because the data after break is less important in the likelihood, and the test statistic is barely significant. As a result, the selected subsample tends to be even longer than that before the break. But in these circumstances, the excessive length of subsample may contribute to better inference on those unchanged parameters to increase information efficiency.

The second stage is about 6 to 12 months after the change point, where the break is discovered and the subsample length is correctly selected and extended over time. It is noteworthy how quickly the procedure detects the breaks once sufficient information is available, which is the key to accurate prediction with parameter changes. Due to the interval between the choice of subsamples, we observe a stepwise increase in the average lengths of the selected subsamples after parameter change.

The third stage occurs when the interval between two changes is sufficiently long, as is the case in our experiment where the stable interval is 150 periods (12.5 years). During this stage the selected subsample tends to stabilize around 4.5 years instead of extending further backward along the homogeneous interval. This happens for two reasons: 1) we impose the maximal subsample of $K = 20$, with an increment step of $M = 6$, which amounts to 120 months or 10 years; 2) the sequential test has the probability of making a type I error to falsely reject the null of homogeneity at any k . Although this downward bias presents in a long, homogeneous interval in our RS-*A* scenario, the adaptive procedure still produces

more desirable forecast compared to various rolling window options. In a real setting with frequent changes in parameters, the adaptive procedure is advantageous for its capacity to quickly detecting changes as with the second-stage feature.

To summarize, the key to the success of the LARX model is the balance struck between the detection power of the instability and the information efficiency of the selection of the longest homogeneous sample.

To simplify our illustration, we have only studied the forecast ability with abrupt breaks using the LARX model. Chen et al. (2010) also show that, in an adaptive LAR setting, the adaptive method is effective to forecast with smooth parameter changes.

4.2 Robustness with respect to misspecifications and hyperparameters

Local interval selection is crucial to the success of the adaptive technique, which relies on the calibration of critical values via Monte Carlo simulations based on the used hyperparameters. Thus, the computation of critical values may be subject to possible parameter misspecifications and improper choice of hyperparameters. In the RS-*A* scenario with a time varying autoregressive coefficient θ_{1t} , we investigate the robustness of the forecast accuracy with respect to possible misspecifications and alternative choices of hyperparameters.

Above all, the critical values depend on the hypothetical ARX parameters θ^* that are used to generate the Monte Carlo data series. Instead of using the default parameter values chosen for the previous study, i.e., $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) = (0.698, -0.007, 0.224)$, we compute critical values under two sets of misspecified hypothetical parameters of 20% deviation, $0.8\bar{\theta}$ and $1.2\bar{\theta}$, and denote the scenario as mis08 and mis12 respectively. In other words, we use the wrong parameter set, $\theta^* = 0.8\bar{\theta}$ or $\theta^* = 1.2\bar{\theta}$, to generate the Monte Carlo experiments and calibrate the critical values, though the series actually follows the ARX model with $\bar{\theta}$. Using the critical values thus obtained, we perform subsample selection, parameter estimation and forecast.

Moreover, our technique involves the selection of subintervals $\{I^k\}_{k=1}^K$, which are defined by the total number of subsamples K and the subsample distance of M periods. Instead of using the default values of $K = 20$ and $M = 6$, we consider four alternative subsample sets of (K, M) : $(10, 6)$, $(30, 6)$, $(20, 3)$ and $(20, 12)$. That is: given $M = 6$, taking fewer and more subsamples; given $K = 20$ and restrict the first interval to be 6 months, taking more refined or sparse steps for the distance between subsequent subsamples. With these alternative sets of subsamples, the rolling window forecast in comparison also has corresponding choices of K window lengths and M step distances.

Table 3 presents the forecast accuracy under misspecified parameters for critical value calibration and alternative hyperparameters. The results confirm the robustness of the adaptive technique. In the misspecified scenarios, the LARX model, though using critical

values computed with imprecise hypothetical parameters with $\pm 20\%$ deviation, still outperforms by 17 out of 20 and 19 out of 20 alternative models in terms of RMSE, and all 20 alternatives in terms of MAE. Under the alternative sets of hyperparameters, for a fixed interval M but a lower or higher K , the adaptive technique outperforms by 7 out of 10 and 28 out of 30 alternative models in terms of RMSE; while for fixed K but shorter or longer steps M , it outperforms by 14 out of 20 and 20 out of 20 alternative rolling windows in terms of RMSE. Certainly for MAE, the adaptive method dominates all alternative rolling window estimations without exception.

[Table 3. Impact of Hyperparameters on Forecast Accuracy]

This analysis suggests that, with an adaptive and data-driven computation of the critical values, the forecast accuracy is, in general, robust with respect to possible parameter misspecifications and to hyperparameters selection.

5 Empirical applications of the ADNS model

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

5.1 Out-of-sample forecast performance of the ADNS model

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 does the adaptive feature benefit the ADNS model when forecasting yields? We perform an out-of-sample forecast comparison to answer this question.

5.1.1 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 from 1997:12, and predict 1-, 3-, 6- and 12-month ahead forecasts for 1998:1, 1998:3, 1998:6 and 1998:12, respectively. We move forward one period at a time to re-do the estimation and forecast until reaching the end of the sample. The 1-month ahead forecast is made for 1998:1 to 2010:9, a total of 153 months; the 3-,

6- and 12-month ahead forecasts are made for 1998:3 to 2010:9 for 151 months, 1998:6 to 2010:9 for 148 months, and 1998:12 to 2010:9 for 142 months, respectively.

At each estimation point, we use data up to that point. We first extract the three NS factors using OLS for the available sample. For each factor, we assume an LAR(1) or LARX(1) process for specific forecast horizons, and test the stability backward using our method as described in Section 3. To reduce computational burden, we use a universal set of subsamples with $K = 20$ and $M = 6$ for each time point, i.e., starting with the shortest subsample covering the past 6 months, with the interval between two adjacent subsamples equal to 6 months, and ending with the longest subsample of 120 months. Once an optimal subsample is selected against the next longer subsample, the change is understood to have taken place sometime in the 6 months 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 homogeneous 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 of 1-, 3-, 6- and 12-month ahead as described by Equation (9) for LAR and by Equation (10) for LARX, respectively. The predicted NS factors at the specific horizon will then be used to form a yield forecast as in Equation (8).

5.1.2 Alternative models for comparison

For comparison, we choose the DNS model, the random walk model and several representative affine arbitrage-free term structure models. The DNS model is a direct and effective comparison to our ADNS model. The random walk model serves as a very natural and popular benchmark for financial time series with non-stationary characteristics. The affine models further the range of comparison among the popular term structure models in the literature.

1) DNS model

We estimate the DNS model with two specifications for the state dynamics: an AR(1) for each factor, and a VAR(1) for all factors jointly. Although Diebold and Li (2006) find that the AR(1) setting for each factor provides a better forecast than the VAR(1), Favero, Niu and Sala (2012) find that the VAR(1) outperforms the AR(1) once the forecast sample extends beyond 2000, which, in general, is the case in our forecast comparison. We provide forecast results using both specifications.

We forecast the NS factors and yield curve with the DNS model using three strategies of sample length selection: a 5-year (60-month) rolling estimation, a 10-year (120-month)

rolling estimation and a recursive estimation. With a rolling estimation of window size n , we fix the sample length n at each point of time. 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.

The 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 the 15-year rolling window used in Favero et al. (2012). Note that a predetermined sample length is unable to simultaneously serve both purposes: to ensure stable estimation under stationarity and have a quick reaction to structural changes.

The forecast procedure of factors and yields is similar to the ADNS model with the selected sample length, as shown in Equations (8) and (9). 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 forecast by their current values, or

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

3) Affine arbitrage-free term structure models

The arbitrage-free assumption is important in an efficient financial market, and no-arbitrage restricted term structure models provide tractability and consistency in the pricing of bonds and related derivatives. For a comprehensive comparison of yield curve forecasts, we consider three representative models in this category: the affine arbitrage-free Nelson-Siegel (AFNS) model, and the three-factor affine arbitrage free term structure models, $A_0(3)$ and $A_1(3)$.

The AFNS model bridges the gap between the traditional NS model with no-arbitrage restrictions, proposed by Christensen, Diebold and Rudebusch (2011) in a continuous-time setup. Niu and Zeng (2012) characterize the discrete-time AFNS model, which is the closest no-arbitrage counterpart of the DNS model, and propose a quick and robust estimation procedure of reduced-dimension optimization with embedded linear regressions. We estimate the discrete-time AFNS for forecast comparison.

Among the traditional three-factor affine arbitrage-free term structure models as classified by Dai and Singleton (2000), we choose $A_0(3)$ and $A_1(3)$, which are relatively flexible and

perform well among the $A_n(3)$ class of models for yield level prediction (Duffee, 2002). For efficient and reliable estimation of the models, we use the estimation method of the closed-form maximum likelihood expansion developed by Aït-Sahalia and Kimmel (2010), which is a generalization of the method of Aït-Sahalia (2002) and Aït-Sahalia (2008), in the application to continuous-time affine arbitrage-free term structure models. For the quick inversion of latent factors in the estimation, we follow the Chen-Scott method (Chen and Scott, 1993) to assume three yields (3 months, 2 years and 7 years) measured without errors.

Due to the no-arbitrage restrictions underlying these models, the yield multi-step ahead forecast has to be iteratively made by the one-step ahead forecast.

5.1.3 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 forecast of NS factors in the ADNS and DNS models, and for yields in all compared models.

5.1.4 Forecast results

Table 4 reports the forecast RMSE and MAE of the three NS 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 AR(1) and VAR(1) specifications with three different sample length strategies: a 5-year and a 10-year rolling window with fixed lengths and the recursive forecast with expanding window length. In each column, the number in bold-face indicates the best forecast for each horizon and forecast measure, and the underlined number indicates the second best.

[Table 4. Forecast Comparison of Factors]

Table 4 shows that, without exception, the LAR and LARX specifications of the ADNS model are always better than the DNS specifications with predetermined sample lengths. The advantage grows as the forecast horizon increases. At the 12-month forecast horizon, the RMSE and MAE of the ADNS are, in general, 40 to 50 percent less than those produced by the DNS models. The LARX model outperforms the LAR model, indicating that the inflation factor, in addition to their lagged information, helps to predict the yield factors in the ADNS model. Among the DNS models, the VAR(1) recursive forecast tends to perform better for factors NS1 and NS2, while the AR(1) forecast with a 120-month rolling window tends to perform better for NS3.

To visualize the forecast differences, we plot the factor forecasts from the LARX of the ADNS model and the recursive forecast of the VAR of the DNS model together with the

real data in Figure 8. Each row displays the forecast of 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 recursive forecasts of the DNS-VAR with dashed lines. Although it is difficult to distinguish between 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 dotted lines of the ADNS forecasts tend to trace the factor dynamics much more closely than do the DNS forecasts. What is dramatic is its capacity to predict 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 recursive window, however, not only reacts with a persistent lag, but also wrongly predicts the change in direction for a persistent period.

[Figure 8. Plots of NS Factors with ADNS-LARX and DNS-VAR Recursive Forecasts]

Table 5 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 forecasts from the random walk model and the three affine arbitrage-free term structure models.

[Table 5. 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 models and the best result among the DNS models, the reduction in the RMSE and MAE of the ADNS model reaches about 30 percent for the 3-month ahead forecast, and almost 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 3- to 12-month yields at the 1-month ahead forecast. The outperformance begins to be obvious from the 3-month ahead forecast for all yields. Also, the reduction in RMSE and MAE when compared with the random walk is fairly substantial, ranging between 20 and 60 percent, in the multi-step ahead forecasts. Compared to the affine models, only the AFNS model tends to do better for the 3- to 12-month yields at the 1-month ahead forecast; in all other cases the ADNS model shows a substantial advantage.

In Figure 9 we plot the yield forecast from the ADNS-LARX and the recursive forecast of the DNS-VAR together with the realised yields. 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. Realised yields are displayed by solid lines; ADNS-LARX forecasts are displayed by dotted lines, and DNS-VAR recursive forecasts by dashed lines. As the forecast horizon lengthens, it is clear that the dotted lines of the ADNS forecasts tend to follow the yield dynamics much more closely than do the DNS forecasts. The excellence of the LARX in predicting dramatic changes and large swings in factors now persists 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-month or even 12-month ahead.

[Figure 9. Plots of Yields with ADNS-LARX and DNS-VAR Recursive Forecasts]

5.1.5 Stability of the LARX forecast performance

The LARX demonstrates visually excellent forecast precision for the recent financial crisis which features abrupt parameter changes. Is the superior performance restricted only to this special period or is it a general property across the whole forecast path? To answer this question, we split the forecast sample into two periods, 1998:1 – 2007:11 and 2007:12 – 2010:9, taking the NBER business cycle peak, 2007:12, as the breaking point. We take the random walk model as a benchmark for comparison, and compute the ratios of RMSE and MAE from the LARX model with respect to the random walk forecast for the whole sample and also the subsamples. A ratio smaller than one indicates the better performance of the LARX model. Results are presented in Table 6, where we show the ratios of RMSE and MAE for five selected yields at specific forecast horizons, and also report the average ratios across all 15 yields in the bottom row.

[Table 6. Subsample Performance of LARX versus Random Walk]

It is clear from Table 6 that the forecast performance of the LARX model is very stable across different sample periods. Although for the 1-month ahead forecast it is not always superior to the random walk model, it reduces the ratios of RMSE and MAE for the 3-, 6- and 12-month forecast by about 20%, 40% and 60%, respectively.

5.1.6 Lengths of stable subsamples

Table 7 summarizes the average lengths detected for sample stability along the forecast for each factor and each horizon. For $h = 1$, it is a standard AR(1)/ARX(1) model with the factor's first lag as the regressor, and for $h > 1$ with its h -th lag as the regressor. Compared to the conventional sample size used in the literature for yield curve forecasts at the monthly frequency, the average length of the stable subsample is surprisingly low, with a maximum of 41 months (about 3.5 years) for $h = 12$ in the LARX of NS1. Overall, the average lengths of the stable subsample is around 30 months, or 2.5 years, which is much shorter than the conventionally used 5-, 10-year rolling or recursive samples.

[Table 7. Average Lengths of Stable Sample Intervals Detected by LAR and LARX]

5.2 Sources of the predictability of the ADNS model

From the previous Monte Carlo simulations of the known data generating process of a univariate series, we understand that the advantage of the LARX model is the combination of detection power on the instability and the information efficiency on the selected homogeneous sample. For the success of the LARX model to continue with the multivariate factor ADNS model for yield prediction, parsimony and flexibility are also crucial.

Although in the DNS model, the VAR(1) specification, in general, leads to better forecasts than the AR(1) specification in our forecast periods, we have refrained from using a VAR setting for the adaptive model out of a concern for parsimony. The reason is, the adaptive procedure per se is a tool to detecting breaks within a sample which tests the joint hypothesis of parameter homogeneity; given sample information, the likelihood test on a large set of parameters tends to have lower power to reject the null when it is false, hence the selected sample tends to be much longer which sacrifices the quick detection power for simpler models with less parameters. To illustrate, we take the LAR forecast as the benchmark, and run a local VAR (LVAR) model with a similar procedure. We estimate the LVAR(1) with one lag and iteratively produce multistep ahead forecast. The average selected sample length is 78 months, more than double the average sample length of the LAR(1) as reported in Table 7. The forecast of the NS three factors using this LVAR specification is reported in Table 8 together with the LAR specification. It is evident that the forecast accuracy deteriorates substantially, due to the slower detection of breaks. Choosing the LAR specification separately for each factor also enables us to detect instability for each factor process with higher flexibility. Hence, parsimony is the key to quick and accurate break detection.

Conditional on the parsimony and flexibility of the LAR or LARX model, the detection power of changes of this adaptive method contributes greatly to unbiased estimation and

forecasts. It is not that choosing shorter interval on average will work. To see this point, we add a DNS 30-month rolling forecast, both the AR(1) and VAR(1) specifications, to Table 8. This is because 30 months is the average length of the homogeneous intervals for the LAR and LARX, as reported in Table 7. The results show clearly that the 30-month rolling forecast is much worse than the LAR forecast.

[Table 8. Sources of the Prediction Power of the ADNS Model]

To summarize, the key to the success of the ADNS model for yield curve forecast is a combination of the following:

- 1) parsimony and flexibility of the univariate dynamic modeling of the NS yield factor, and
- 2) the superior detection power of instability and information efficiency in the homogeneous subsample selection.

5.3 Detecting breaks and monitoring parameter evolution

We have shown that the ADNS model predicts yield dynamics remarkably better than the DNS model, benefiting from its adaptive feature. There must be useful information hidden in the detected breaks and parameter evolvments 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 10 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. 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 solid line along the horizontal axis of time. At the end of the line, a dark 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 10. Breaks of Factors Detected in the ADNS-LARX Model]

Observing Figure 10, we find that the stable subsamples of the level factor are relatively longer than those of slope and curvature. For level, the intervals of 1994-1998, 1998-2003

and 2005-2008 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. The initial breaks are led by the level factor around 1994, with breaks in slope and curvature following in 1995. There seems to be a common break in all three factors just at the beginning of the financial crisis and recession in 2008. In between, there seem to be sequential breaks without clear patterns.

Those major breaks in level align with some big events and debated policy change. For example, the breaks around 1998 and 2008 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 “interest 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 from a richer data set of the whole yield curve. The policy change is criticized as having accommodated the financial bubble which eventually ended 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 evolution of parameters is helpful to understand the origins of the breaks. When we look into the parameter changes as presented earlier in Figure 6, we observe that the CPI inflation coefficient on the level factor became 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 of changing inflation coefficient is very revealing about an important break in the relationship between interest rates and inflation. Figure 6 is also very informative on the risks inherent in interest rate dynamics. The variance parameters of innovations of all three factors build up during two periods: gradually increasing during 2002 to 2004 under loose monetary policy and rapidly surging during the recent financial crisis. In the latter period, the decline in the autoregressive coefficients of all three processes implies an even higher uncertainty in yield prediction.

6 Conclusion

By optimally selecting the sample period over which parameters are approximately constant for the factor dynamics, the ADNS model we propose produces superior forecasts of the yield curve compared with the DNS model and affine arbitrage-free term structure models

with predetermined sample lengths and the random walk model. The ADNS model not only substantially improves forecast precision by reducing forecast errors by between 20 to 60 percent relative to the random walk, 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, combined with parsimony. 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 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 non-stationary, with different sources of changes including heteroscedasticity, structure breaks, and regime switching.

Acknowledgement

Linlin Niu acknowledges the support of the Natural Science Foundation of China (Grant No. 70903053 and Grant No. 71273007).

References

- Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely sampled diffusions: a closed-form approximation approach, *Econometrica* **70**: 223–262.
- Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions, *Annals of Statistics* **36**: 906–937.
- Aït-Sahalia, Y. and Kimmel, R. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions, *Journal of Financial Economics* **98**: 113–144.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second international symposium on information theory*, Akademinai Kiado, pp. 267–281.
- Ang, A. and Bekaert, G. (2002). Regime switches in interest rates, *Journal of Business and Economic Statistics* **20**: 163–182.
- 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.

- 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.
- Bansal, R. and Zhou, H. (2002). Term structure of interest rates with regime shifts, *Journal of Finance* **57**(5): 1997–2043.
- Chen, R. and Scott, L. (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* **3**: 14–31.
- Chen, Y., Härdle, W. and Pigorsch, U. (2010). Localized realized volatility modelling, *Journal of the American Statistical Association* **105**: 1376–1393.
- Christensen, J., Diebold, F. and Rudebusch, G. (2011). The affine arbitrage-free class of nelson-siegel term structure models, *Journal of Econometrics* **164**: 4–20.
- Dai, Q. and Singleton, K. (2000). Specification analysis of affine term structure models, *Journal of Econometrics* **55**: 1943–1978.
- 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, *Economics working paper archive*, 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 restriction vs. large information set, *Journal of Forecasting* **31**: 124–156.
- 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.

- 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.
- 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.
- Lee, J. and Strazicich, M. C. (2003). Minimum lagrange multiplier unit root test with two structural breaks, *Review of Economics and Statistics* **85**(4): 1082–1089.
- Marcellino, M., Stock, J. and Watson, M. (2006). A comparison of direct and iterated AR methods for forecasting macroeconomic series h-steps ahead, *Working paper*, CEPR. No. 4976.
- Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time inhomogeneous volatility models, *The Annals of Statistics* **32**: 577–602.
- Mishkin, F. S. (1990). The information in the longer maturity term structure about future inflation, *The Quarterly Journal of Economics* **105**: 815–828.
- Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach, *Journal of Econometrics* **146**(1): 26–43.
- Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curve, *Journal of Business* **60**: 473–489.
- Niu, L. and Zeng, G. (2012). The discrete-time framework of arbitrage-free Nelson-Siegel class of term structure models, *Technical report*, Xiamen University. Working paper.
- 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.
- Schwarz, G. et al. (1978). Estimating the dimension of a model, *The annals of statistics* **6**(2): 461–464.
- 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*, The MIT Press, Cambridge, MA, pp. 13–22.

Table 1. Parameters Used in the Simulated Scenarios

Scenarios	Parameter Values				
HOM	Constant parameters	$(\theta_{1t}, \theta_{2t}, \sigma) = (0.698, -0.007, 0.224)$ for $t = 1, \dots, 600$.			
RS	Time-varying parameters	Phase 1 $t = 1, \dots, 150.$	Phase 2 $t = 151, \dots, 300$	Phase 3 $t = 301, \dots, 450.$	Phase 4 $t = 451, \dots, 600.$
RS-A	θ_{1t}	0.698	-0.067	0.698	0.917
RS-X	θ_{2t}	-0.007	-0.297	-0.007	0.305
RS-V	σ	0.224	0.110	0.224	0.369
RS-AV	(θ_{1t}, σ)	(0.698, 0.224)	(0.917, 0.110)	(0.698, 0.224)	(-0.067, 0.369)

Notes: Scenario HOM refers to the ARX process with constant parameters. In each of the regime-switching scenarios initiated with RS, only one or two of the parameters listed in the second column varies, while the remaining parameters are fixed to the default value as in the constant parameter setting.

Table 2. Forecast Comparison of the Adaptive and Rolling Estimation in Simulated Scenarios

Scenarios	RMSE				MAE			
	Rolling		Adaptive	No. of Win	Rolling		Adaptive	No. of Win
	Best (k^*)	Worst ($k!$)			Best (k^*)	Worst ($k!$)		
HOM	0.240 (20)	0.305 (1)	0.252	3/20	0.184 (20)	0.230 (1)	0.190	5/20
RS-A	0.274 (3)	0.372 (20)	0.277	18/20	0.207 (3)	0.272 (20)	0.205	20/20
RS-X	0.246 (4)	0.294 (1)	0.255	11/20	0.194 (5)	0.226 (1)	0.198	13/20
RS-V	0.270 (20)	0.343 (1)	0.283	3/20	0.192 (20)	0.240 (1)	0.200	4/20
RS-AV	0.303 (4)	0.360 (20)	0.314	14/20	0.212 (4)	0.251 (20)	0.215	16/20

Notes:

- 1) The simulated scenarios are described in Table 1. The rolling window adopts one of the predetermined window lengths of $k \cdot M$, where $k = 1, \dots, 20$, and $M = 6$, throughout the whole sample. The adaptive technique adopts a selected time-varying window length among the choices of $k \cdot M$, where $k = 1, \dots, 20$, and $M = 6$, at each point of time.
- 2) For the performance of the rolling windows, only the best and worst results with the related window choices, denoted by k^* and $k!$, respectively, are reported. We also report the number of wins of the adaptive technique compared to the 20 rolling estimation alternatives.

Table 3. Impact of Hyperparameters on Forecast Accuracy

Scenarios	RMSE				MAE			
	Rolling		Adaptive	No. of Win	Rolling		Adaptive	No. of Win
	Best (k^*)	Worst ($k!$)			Best (k^*)	Worst ($k!$)		
mis08	0.274 (3)	0.371 (20)	0.281	17/20	0.207 (3)	0.272 (20)	0.206	20/20
mis12	0.274 (3)	0.371 (20)	0.275	19/20	0.207 (3)	0.272 (20)	0.204	20/20
$K = 10$	0.273 (3)	0.312 (10)	0.276	7/10	0.207 (3)	0.232 (1)	0.205	10/10
$K = 30$	0.263 (3)	0.408 (30)	0.265	28/30	0.202 (4)	0.308 (30)	0.199	30/30
$M = 3$	0.271 (5)	0.309 (20)	0.277	14/20	0.205 (5)	0.231 (1)	0.205	20/20
$M = 12$	0.269 (2)	0.411 (20)	0.264	20/20	0.204 (2)	0.319 (20)	0.200	20/20

Notes:

- 1) The simulated scenarios are based on the RS-A scenario as described in Table 1, but with misspecification in the simulated series to calculate critical values, or with alternative sets of hyperparameters, K and M . The rolling window adopts one of the predetermined window lengths of $k \cdot M$ ($k \leq K$) throughout the whole sample. The adaptive technique adopts a selected time-varying window length among the choices of $k \cdot M$ ($k \leq K$).
- 2) For the first two cases, $K = 20$ and $M = 6$; but to simulate the training sample for calculating the critical values, θ_{1t} is set to 0.8 or 1.2 times of the mean value in the true sample, 0.698. For the third and fourth cases, the other conditions are the same as in the default setting, except that K is set to 10 or 30. For the fifth and sixth cases, the other conditions are the same as in the default setting, except that M is set to 3 or 12.
- 3) For the performance of the rolling windows, only the best and worst results with the related window choices, denoted by k^* and $k!$, respectively, are reported. We also report the number of wins of the adaptive technique compared to the K rolling estimation alternatives.

Table 4. Forecast Comparison of Factors

NS1		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.271	0.193	0.349	0.256	0.389	0.289	0.479	0.358
	LAR	<u>0.279</u>	<u>0.199</u>	<u>0.365</u>	<u>0.269</u>	<u>0.428</u>	<u>0.334</u>	<u>0.479</u>	<u>0.367</u>
DNS	AR Rolling 60m	0.333	0.239	0.462	0.356	0.623	0.496	0.738	0.540
	AR Rolling 120m	0.330	0.232	0.463	0.350	0.617	0.495	0.780	0.622
	AR Recursive	0.331	0.232	0.461	0.346	0.610	0.490	0.729	0.559
	VAR Rolling 60m	0.345	0.253	0.526	0.415	0.740	0.621	0.780	0.645
	VAR Rolling 120m	0.336	0.238	0.481	0.379	0.661	0.536	0.846	0.732
	VAR Recursive	0.332	0.232	0.455	0.341	0.582	0.475	0.669	0.557

NS2		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.366	0.269	0.523	0.409	0.627	0.435	0.850	0.595
	LAR	<u>0.371</u>	<u>0.278</u>	<u>0.569</u>	<u>0.426</u>	<u>0.715</u>	<u>0.540</u>	<u>0.998</u>	<u>0.674</u>
DNS	AR Rolling 60m	0.453	0.343	0.893	0.714	1.570	1.922	3.026	2.568
	AR Rolling 120m	0.434	0.321	0.782	0.599	1.234	1.025	1.889	1.619
	AR Recursive	0.431	0.320	0.772	0.600	1.203	1.013	1.853	1.601
	VAR Rolling 60m	0.434	0.331	0.762	0.594	1.259	1.011	2.563	1.878
	VAR Rolling 120m	0.421	0.319	0.704	0.546	1.071	0.859	1.832	1.520
	VAR Recursive	0.415	0.311	0.679	0.530	1.013	0.825	1.555	1.300

NS3		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.712	0.511	0.980	0.726	1.130	0.815	0.992	0.770
	LAR	<u>0.768</u>	<u>0.544</u>	<u>1.104</u>	<u>0.810</u>	<u>1.260</u>	<u>0.916</u>	<u>1.259</u>	<u>0.952</u>
DNS	AR Rolling 60m	0.889	0.665	1.497	1.125	2.171	1.693	3.204	2.683
	AR Rolling 120m	0.863	0.644	1.394	1.039	1.913	1.496	2.529	2.085
	AR Recursive	0.868	0.647	1.442	1.075	1.992	1.537	2.660	2.206
	VAR Rolling 60m	0.913	0.682	1.658	1.199	2.679	2.020	5.087	3.521
	VAR Rolling 120m	0.878	0.652	1.537	1.104	2.373	1.838	3.797	3.080
	VAR Recursive	0.861	0.648	1.444	1.099	2.004	1.590	2.761	2.298

Notes:

- 1) h denotes the forecast horizon. 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 to 2010:9 with 151 months. The 6-month ahead forecast period is 1998:6 to 2010:9 with 148 months and the 12-month ahead forecast period is 1998:12 to 2010:9 with 142 months.
- 2) For each column, the best forecast, i.e., the smallest RMSE or MAE, is marked in bold-face; the second best is underlined. When the first and second best are the same, they are both marked in bold-face, and no second best is further indicated. When the second and third best are the same, they are both underlined.

Table 5. Forecast Comparison of Yields

y (3)		h = 1		h = 3		h = 6		h = 12	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.312	0.215	0.421	<u>0.312</u>	0.550	0.379	0.638	0.426
	LAR	0.309	0.213	<u>0.440</u>	0.311	<u>0.585</u>	<u>0.420</u>	<u>0.749</u>	<u>0.501</u>
DNS	AR Rolling 60m	0.346	0.237	0.722	0.529	1.345	1.066	2.871	2.339
	AR Rolling 120m	0.326	0.219	0.637	0.447	1.077	0.860	1.814	1.568
	AR Recursive	0.318	0.208	0.626	0.460	1.048	0.878	1.778	1.577
	VAR Rolling 60m	0.272	0.174	0.596	0.429	1.103	0.854	2.853	1.864
	VAR Rolling 120m	0.265	0.163	0.529	0.344	0.956	0.669	2.037	1.585
	VAR Recursive	0.266	0.155	0.515	0.320	0.890	0.632	1.617	1.352
Random Walk		0.254	0.153	0.547	0.363	0.927	0.673	1.650	1.306
Affine	AFNS Rolling 60m	<u>0.251</u>	0.161	0.550	0.399	0.993	0.777	1.832	1.515
	AFNS Rolling 120m	0.242	<u>0.145</u>	0.494	0.326	0.857	0.620	1.574	1.295
	AFNS Recursive	<u>0.251</u>	0.141	0.492	0.321	0.835	0.608	1.512	1.231
	A ₀ (3) Rolling 60m	0.269	0.173	1.281	0.888	2.249	1.743	3.169	2.565
	A ₀ (3) Rolling 120m	0.278	0.182	1.302	0.950	2.087	1.709	2.604	2.262
	A ₀ (3) Recursive	0.257	0.154	1.188	0.791	1.959	1.545	2.575	2.139
	A ₁ (3) Rolling 60m	0.292	0.189	1.440	1.074	2.487	2.037	3.357	2.760
	A ₁ (3) Rolling 120m	0.324	0.232	1.587	1.194	2.525	2.011	3.308	2.660
A ₁ (3) Recursive	0.319	0.221	1.520	1.080	2.493	1.987	3.267	2.689	

y (12)		h = 1		h = 3		h = 6		h = 12	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.251	0.185	0.381	0.284	0.503	0.360	0.574	0.430
	LAR	0.247	0.184	<u>0.410</u>	<u>0.296</u>	<u>0.542</u>	<u>0.406</u>	<u>0.666</u>	<u>0.514</u>
DNS	AR Rolling 60m	0.280	0.216	0.675	0.508	1.264	1.020	2.608	2.150
	AR Rolling 120m	0.260	0.197	0.601	0.458	1.024	0.833	1.722	1.511
	AR Recursive	0.269	0.206	0.622	0.498	1.042	0.884	1.732	1.544
	VAR Rolling 60m	0.272	0.214	0.645	0.493	1.198	0.917	3.012	1.925
	VAR Rolling 120m	0.255	0.194	0.579	0.409	1.070	0.767	2.182	1.696
	VAR Recursive	0.258	0.197	0.561	0.415	0.951	0.732	1.660	1.413
Random Walk		0.251	0.176	0.561	0.401	0.927	0.699	1.582	1.284
Affine	AFNS Rolling 60m	0.257	0.193	0.609	0.459	1.089	0.847	1.897	1.574
	AFNS Rolling 120m	0.248	<u>0.181</u>	0.564	0.404	0.980	0.730	1.685	1.378
	AFNS Recursive	0.247	0.183	0.545	0.403	0.915	0.707	1.556	1.286
	A ₀ (3) Rolling 60m	0.288	0.212	1.514	1.104	2.696	2.104	3.840	3.043
	A ₀ (3) Rolling 120m	0.266	0.199	1.284	0.923	1.950	1.598	2.411	2.088
	A ₀ (3) Recursive	0.255	0.186	1.262	0.879	1.991	1.613	2.590	2.183
	A ₁ (3) Rolling 60m	0.260	0.186	1.352	0.967	2.202	1.786	2.943	2.430
	A ₁ (3) Rolling 120m	0.289	0.228	1.429	1.114	2.144	1.801	2.626	2.296
A ₁ (3) Recursive	0.271	0.204	1.375	0.978	2.178	1.771	2.841	2.366	

y (36)		h = 1		h = 3		h = 6		h = 12	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.282	0.214	0.426	0.328	0.510	0.392	0.466	0.360
	LAR	<u>0.294</u>	<u>0.223</u>	<u>0.465</u>	<u>0.347</u>	<u>0.534</u>	<u>0.404</u>	<u>0.526</u>	<u>0.414</u>
DNS	AR Rolling 60m	0.330	0.256	0.640	0.497	1.045	0.802	1.954	1.559
	AR Rolling 120m	0.326	0.253	0.627	0.478	0.955	0.754	1.478	1.234
	AR Recursive	0.330	0.256	0.650	0.507	0.981	0.781	1.486	1.260
	VAR Rolling 60m	0.328	0.253	0.702	0.519	1.202	0.855	2.676	1.720
	VAR Rolling 120m	0.323	0.250	0.661	0.482	1.111	0.813	2.038	1.582
	VAR Recursive	0.317	0.243	0.619	0.473	0.940	0.750	1.499	1.258
Random Walk		0.300	0.228	0.573	0.448	0.845	0.679	1.237	1.026
Affine	AFNS Rolling 60m	0.325	0.253	0.642	0.478	0.999	0.728	1.534	1.209
	AFNS Rolling 120m	0.319	0.245	0.627	0.461	0.978	0.720	1.498	1.197
	AFNS Recursive	0.317	0.242	0.597	0.459	0.890	0.709	1.323	1.106
	A ₀ (3) Rolling 60m	0.339	0.256	1.535	1.125	2.564	1.972	3.526	2.740
	A ₀ (3) Rolling 120m	0.315	0.240	1.200	0.839	1.639	1.275	1.959	1.658
	A ₀ (3) Recursive	0.326	0.252	1.242	0.868	1.753	1.375	2.171	1.788
	A ₁ (3) Rolling 60m	0.323	0.248	1.237	0.851	1.776	1.386	2.272	1.826
	A ₁ (3) Rolling 120m	0.340	0.261	1.295	0.956	1.764	1.428	2.062	1.722
A ₁ (3) Recursive	0.324	0.251	1.240	0.866	1.754	1.367	2.202	1.771	

y (60)		h = 1		h = 3		h = 6		h = 12	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	0.285	0.217	0.412	0.324	0.477	0.369	0.445	0.352
	LAR	<u>0.297</u>	<u>0.228</u>	<u>0.444</u>	<u>0.340</u>	<u>0.498</u>	<u>0.393</u>	<u>0.476</u>	<u>0.391</u>
DNS	AR Rolling 60m	0.335	0.259	0.575	0.453	0.873	0.673	1.487	1.166
	AR Rolling 120m	0.340	0.265	0.592	0.459	0.859	0.680	1.258	1.063
	AR Recursive	0.342	0.266	0.608	0.471	0.878	0.694	1.252	1.044
	VAR Rolling 60m	0.343	0.267	0.664	0.497	1.077	0.785	2.168	1.440
	VAR Rolling 120m	0.339	0.264	0.630	0.478	1.007	0.744	1.730	1.339
	VAR Recursive	0.328	0.254	0.583	0.453	0.843	0.671	1.272	1.058
Random Walk		0.301	0.231	0.532	0.429	0.753	0.607	1.005	0.829
Affine	AFNS Rolling 60m	0.323	0.255	0.581	0.443	0.846	0.623	1.198	0.905
	AFNS Rolling 120m	0.313	0.238	0.575	0.435	0.865	0.648	1.259	0.979
	AFNS Recursive	0.314	0.238	0.547	0.429	0.784	0.629	1.085	0.897
	A ₀ (3) Rolling 60m	0.322	0.244	1.367	0.966	2.125	1.608	2.810	2.175
	A ₀ (3) Rolling 120m	0.310	0.236	1.115	0.734	1.444	1.042	1.650	1.311
	A ₀ (3) Recursive	0.323	0.247	1.153	0.770	1.529	1.130	1.800	1.405
	A ₁ (3) Rolling 60m	0.317	0.242	1.140	0.754	1.530	1.134	1.860	1.431
	A ₁ (3) Rolling 120m	0.326	0.245	1.175	0.811	1.528	1.149	1.732	1.351
A ₁ (3) Recursive	0.317	0.242	1.138	0.758	1.510	1.107	1.803	1.369	

y(120)		h = 1		h = 3		h = 6		h = 12	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LARX	<u>0.269</u>	<u>0.213</u>	0.352	0.275	0.388	0.302	0.433	0.338
	LAR	0.268	0.208	<u>0.362</u>	<u>0.281</u>	<u>0.404</u>	<u>0.331</u>	<u>0.422</u>	<u>0.339</u>
DNS	AR Rolling 60m	0.295	0.230	0.427	0.354	0.587	0.481	0.822	0.656
	AR Rolling 120m	0.290	0.221	0.437	0.354	0.601	0.472	0.828	0.692
	AR Recursive	0.299	0.226	0.452	0.365	0.622	0.494	0.818	0.664
	VAR Rolling 60m	0.303	0.238	0.512	0.399	0.786	0.606	1.306	0.892
	VAR Rolling 120m	0.298	0.227	0.468	0.374	0.706	0.555	1.124	0.884
	VAR Recursive	0.303	0.228	0.448	0.365	0.604	0.512	0.822	0.684
Random Walk		0.287	<u>0.213</u>	0.443	0.355	0.595	0.485	0.718	0.566
Affine	AFNS Rolling 60m	0.301	0.234	0.476	0.380	0.657	0.503	0.837	0.632
	AFNS Rolling 120m	0.303	0.233	0.489	0.392	0.705	0.552	0.970	0.769
	AFNS Recursive	0.296	0.221	0.451	0.368	0.618	0.513	0.779	0.633
	A ₀ (3) Rolling 60m	0.298	0.231	1.134	0.725	1.572	1.097	1.903	1.389
	A ₀ (3) Rolling 120m	0.304	0.233	1.026	0.619	1.290	0.789	1.378	0.889
	A ₀ (3) Recursive	0.309	0.233	1.028	0.618	1.300	0.815	1.410	0.938
	A ₁ (3) Rolling 60m	0.310	0.242	1.043	0.634	1.328	0.868	1.477	1.001
	A ₁ (3) Rolling 120m	0.300	0.231	1.031	0.623	1.304	0.820	1.407	0.907
A ₁ (3) Recursive	0.308	0.241	1.037	0.632	1.314	0.843	1.441	0.943	

Notes:

- 1) h denotes the forecast horizon. 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 to 2010:9 with 151 months. The 6-month ahead forecast period is 1998:6 to 2010:9 with 148 months and the 12-month ahead forecast period is 1998:12 to 2010:9 with 142 months.
- 2) For each column, the best forecast, i.e., the smallest RMSE or MAE, is marked in bold-face; the second best is underlined. When the first and second best are the same, they are both marked in bold-face, and no second best is further indicated. When the second and third best are the same, they are both underlined.

Table 6. Subsample Performance of LARX versus Random Walk

Whole Sample 1998:1 - 2010:9	$h = 1$		$h = 3$		$h = 6$		$h = 12$	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
y(3)	1.229	1.411	0.770	0.860	0.593	0.563	0.387	0.326
y(12)	0.999	1.052	0.678	0.708	0.543	0.514	0.363	0.335
y(36)	0.939	0.939	0.744	0.731	0.603	0.578	0.377	0.350
y(60)	0.947	0.941	0.776	0.756	0.634	0.608	0.443	0.425
y(120)	0.938	1.003	0.794	0.773	0.652	0.623	0.603	0.596
Average	0.962	0.997	0.739	0.748	0.599	0.571	0.432	0.408

Subsample 1 1998:1 - 2007:11	$h = 1$		$h = 3$		$h = 6$		$h = 12$	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
y(3)	1.189	1.316	0.804	0.838	0.635	0.588	0.366	0.324
y(12)	0.952	0.998	0.674	0.692	0.587	0.540	0.327	0.317
y(36)	0.931	0.934	0.762	0.732	0.631	0.595	0.352	0.337
y(60)	0.953	0.949	0.782	0.754	0.657	0.633	0.426	0.423
y(120)	0.952	1.002	0.785	0.761	0.707	0.662	0.556	0.570
Average	0.957	0.983	0.748	0.741	0.638	0.599	0.403	0.395

Subsample 2 2007:12 - 2010:9	$h = 1$		$h = 3$		$h = 6$		$h = 12$	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
y(3)	1.297	1.728	0.750	0.991	0.483	0.574	0.213	0.235
y(12)	1.106	1.232	0.760	0.834	0.454	0.500	0.367	0.351
y(36)	0.960	0.955	0.733	0.748	0.527	0.554	0.386	0.375
y(60)	0.932	0.919	0.780	0.775	0.578	0.546	0.475	0.438
y(120)	0.915	1.005	0.837	0.852	0.546	0.560	0.679	0.711
Average	0.974	1.044	0.760	0.809	0.507	0.529	0.440	0.428

Notes:

- 1) h denotes the forecast horizon.
- 2) The reported measures are the ratios of the RMSE and the MAE of the LARX model with respect to those of the Random Walk. When the number is less than one, the LARX model performs better.
- 3) The bottom row shows the average ratios of RMSE and MAE across the 15 maturities of yields.

Table 7. 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	37	36	30	34	32	35	40	41
NS2	29	29	25	28	26	27	26	30
NS3	26	26	26	26	26	27	30	26

Notes: h denotes the forecast horizon.

Table 8. Sources of the Prediction Power of the ADNS Model

NS1		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LAR	0.279	0.199	0.365	0.269	0.428	0.334	0.479	0.367
	LVAR	0.337	0.240	0.487	0.387	0.680	0.550	0.932	0.752
DNS	AR Rolling 30m	0.341	0.242	0.474	0.375	0.651	0.503	0.751	0.580
	VAR Rolling 30m	0.354	0.264	0.547	0.440	0.874	0.696	0.980	0.788

NS2		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LAR	0.371	0.278	0.569	0.426	0.715	0.540	0.998	0.674
	LVAR	0.416	0.309	0.717	0.549	1.178	0.934	2.294	1.816
DNS	AR Rolling 30m	0.458	0.341	0.926	0.726	1.720	1.298	4.022	3.051
	VAR Rolling 30m	0.474	0.354	0.872	0.661	1.558	1.186	3.762	2.858

NS3		$h = 1$		$h = 3$		$h = 6$		$h = 12$	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ADNS	LAR	0.768	0.544	1.104	0.810	1.260	0.916	1.259	0.952
	LVAR	0.976	0.707	1.720	1.234	2.476	1.880	3.832	2.855
DNS	AR Rolling 30m	0.905	0.664	1.692	1.223	2.417	1.902	3.901	3.189
	VAR Rolling 30m	0.946	0.702	2.084	1.470	3.062	2.275	5.138	3.735

Notes:

- 1) h denotes the forecast horizon. 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 to 2010:9 with 151 months. The 6-month ahead forecast period is 1998:6 to 2010:9 with 148 months and the 12-month ahead forecast period is 1998:12 to 2010:9 with 142 months.
- 2) For each column, the best forecast, i.e., the smallest RMSE or MAE, is marked in bold-face.
- 3) The row indicated as LVAR shows the result of an ADNS specification where the state factors are jointly modeled as a VAR(1). At each point of time, the LVAR(1) is detected and estimated, and the model is used to iteratively produce multi-step ahead forecast for the NS three factors.

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

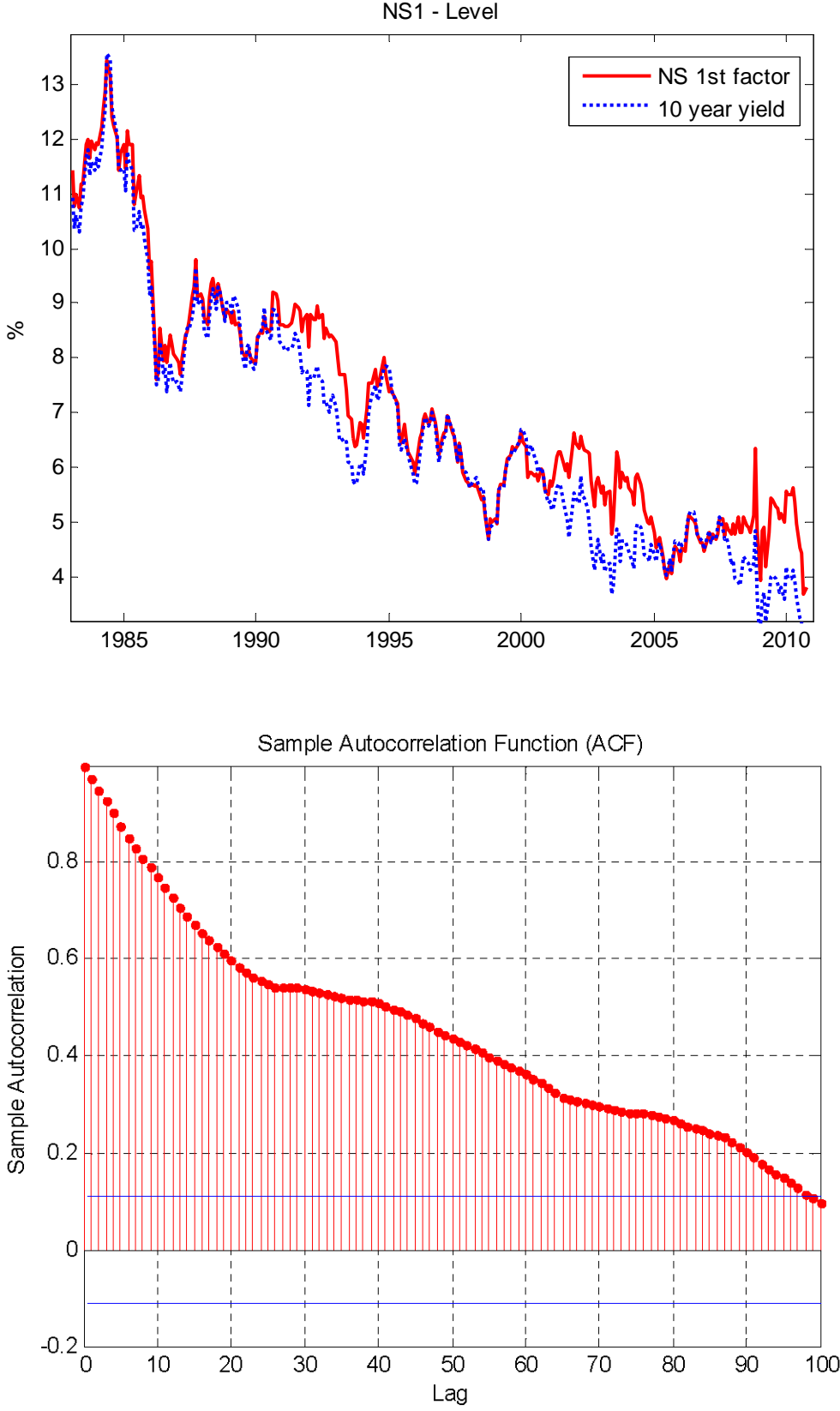


Figure 2. Three Yields and CPI Inflation in the US (1983.1-2010.9)

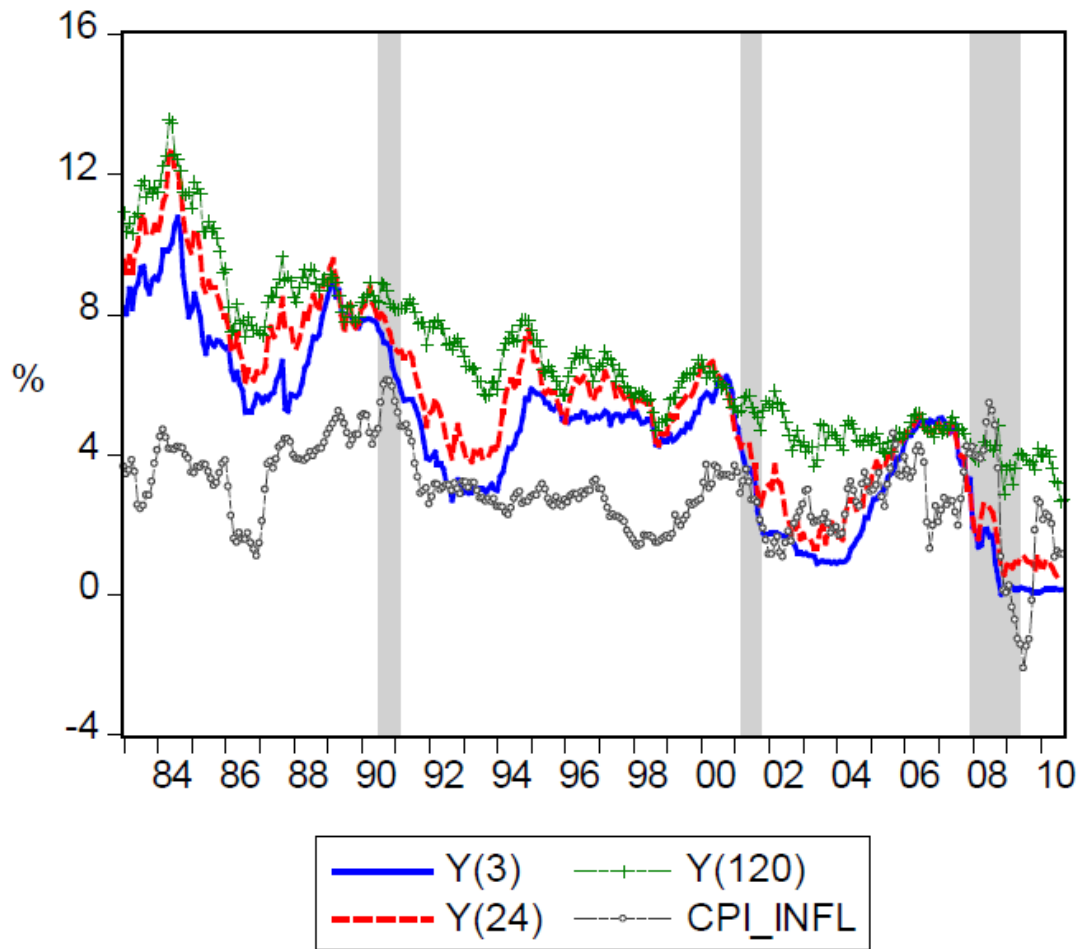


Figure 3. The US Yield Curve (1983:1 to 2010:9)

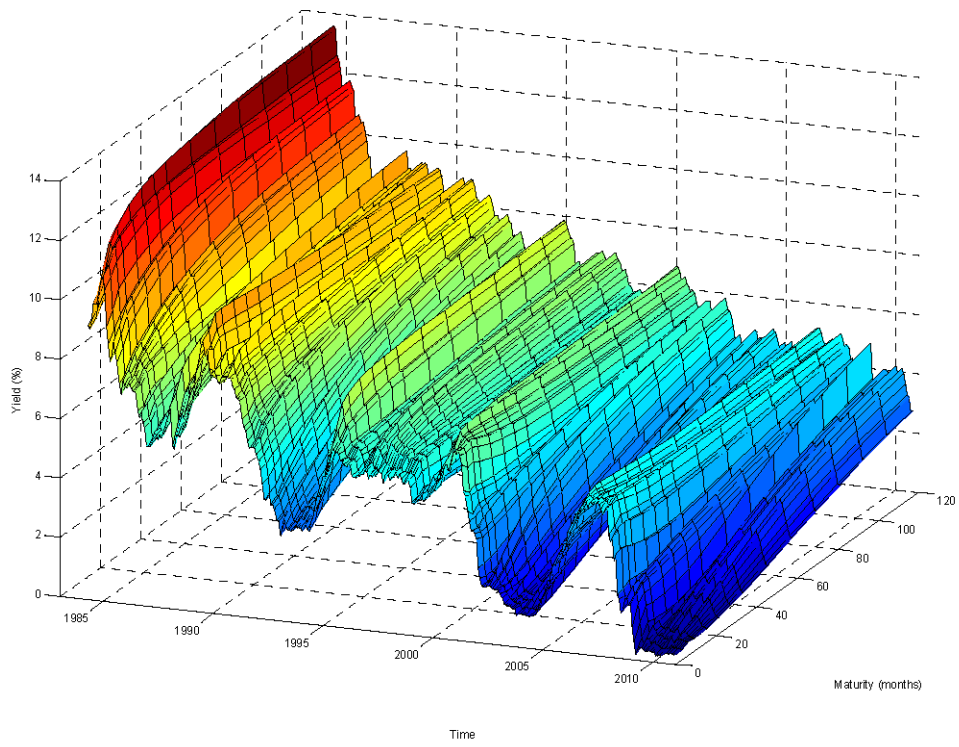


Figure 4. Nelson-Siegel Factor Loadings

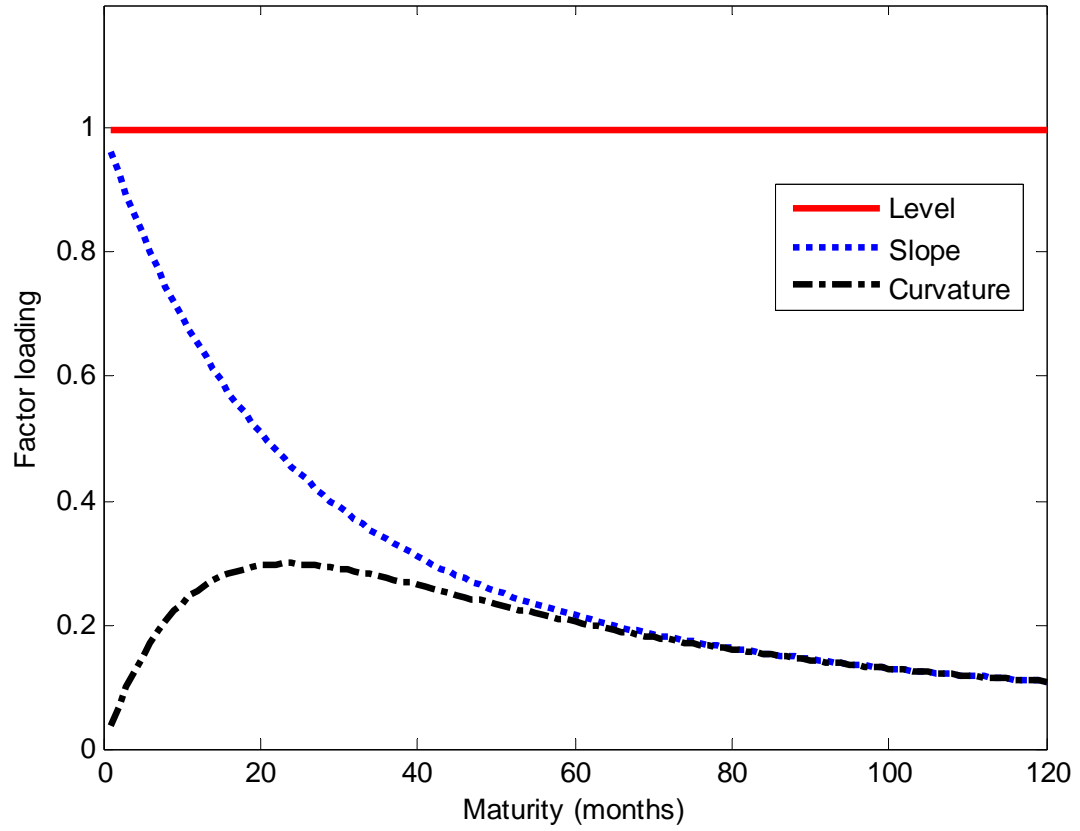


Figure 5. Time Evolution of Nelson-Siegel Factors Extracted from US Yield Curve (1983:1 to 2010:9)

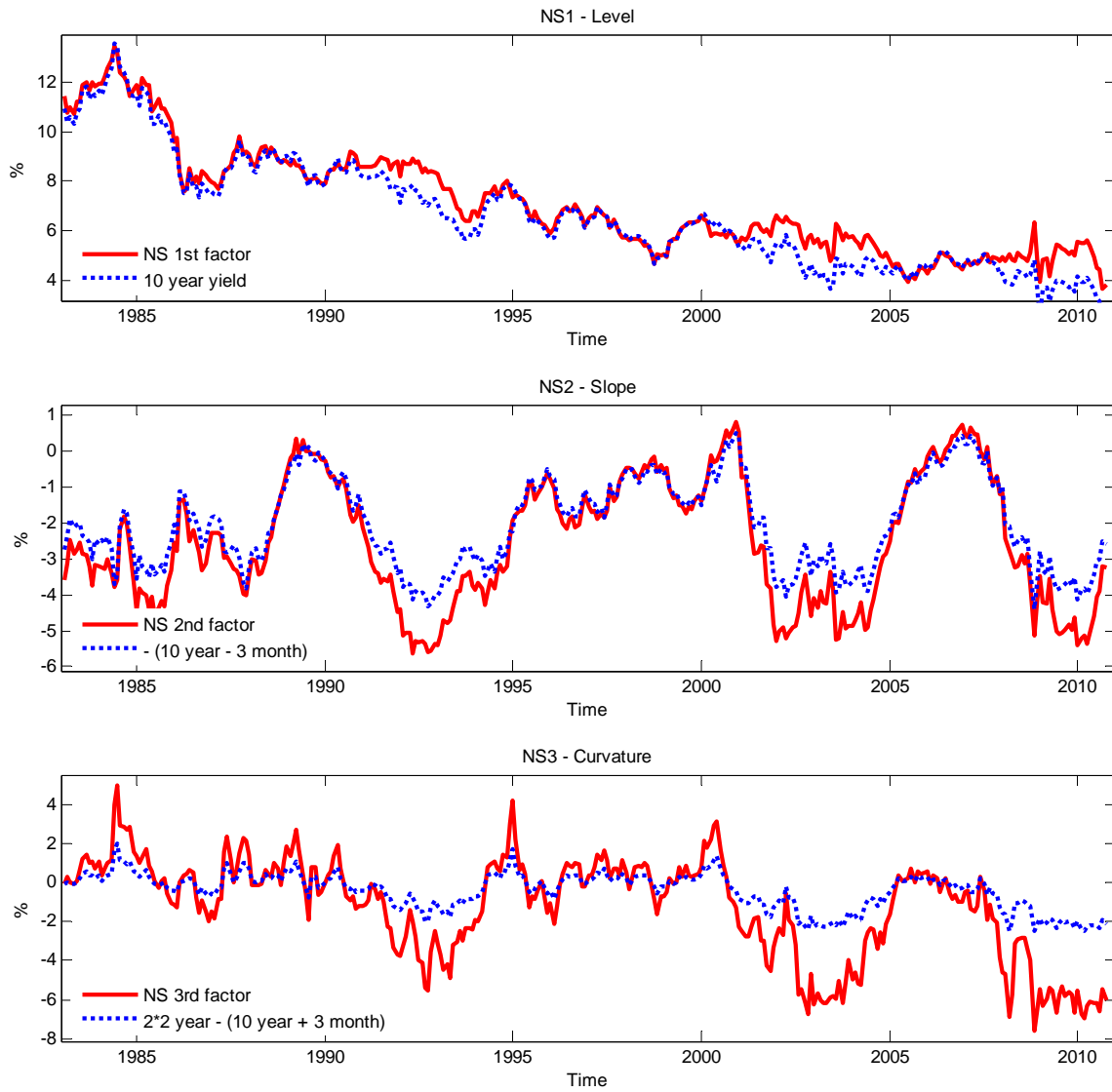
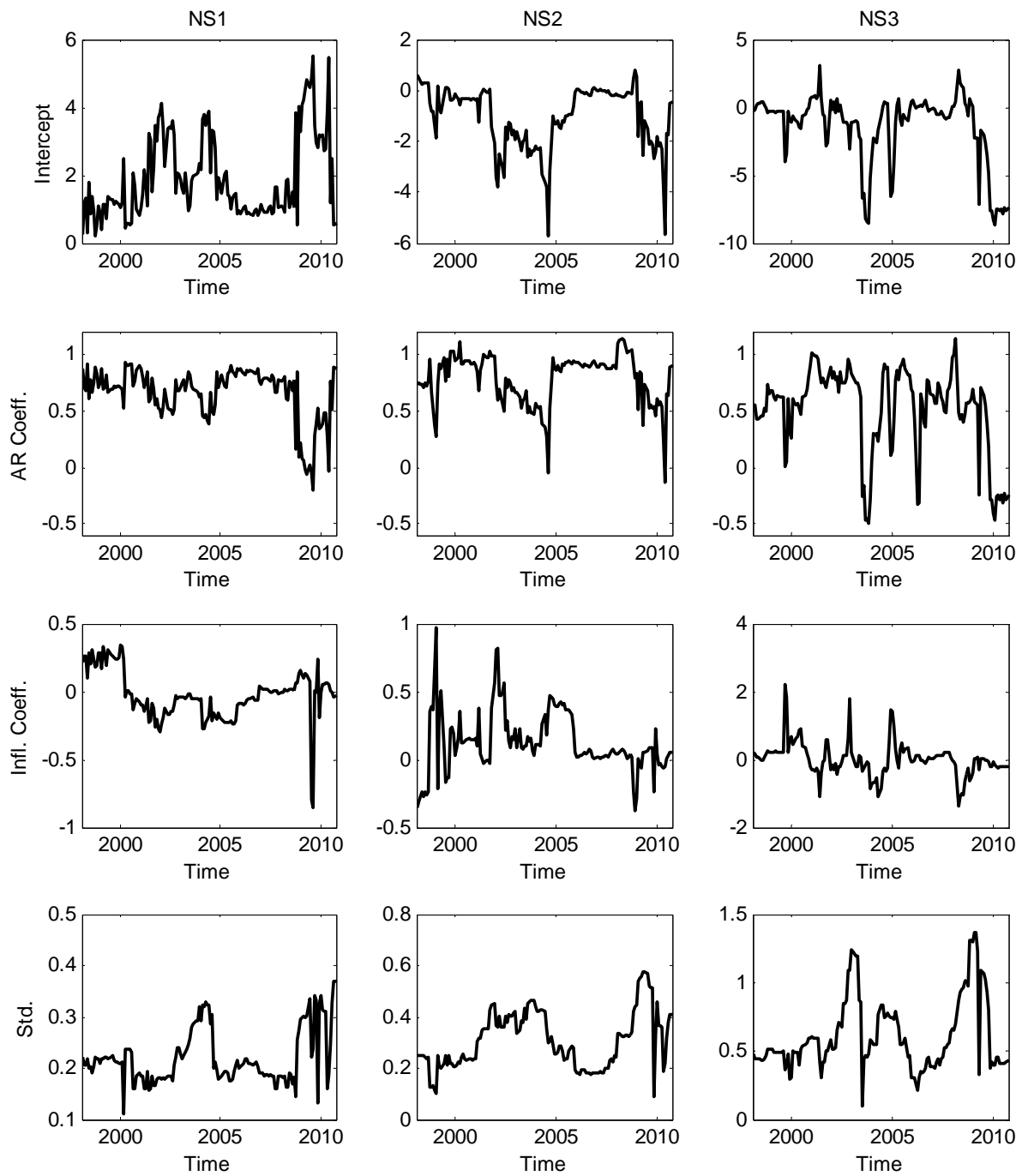
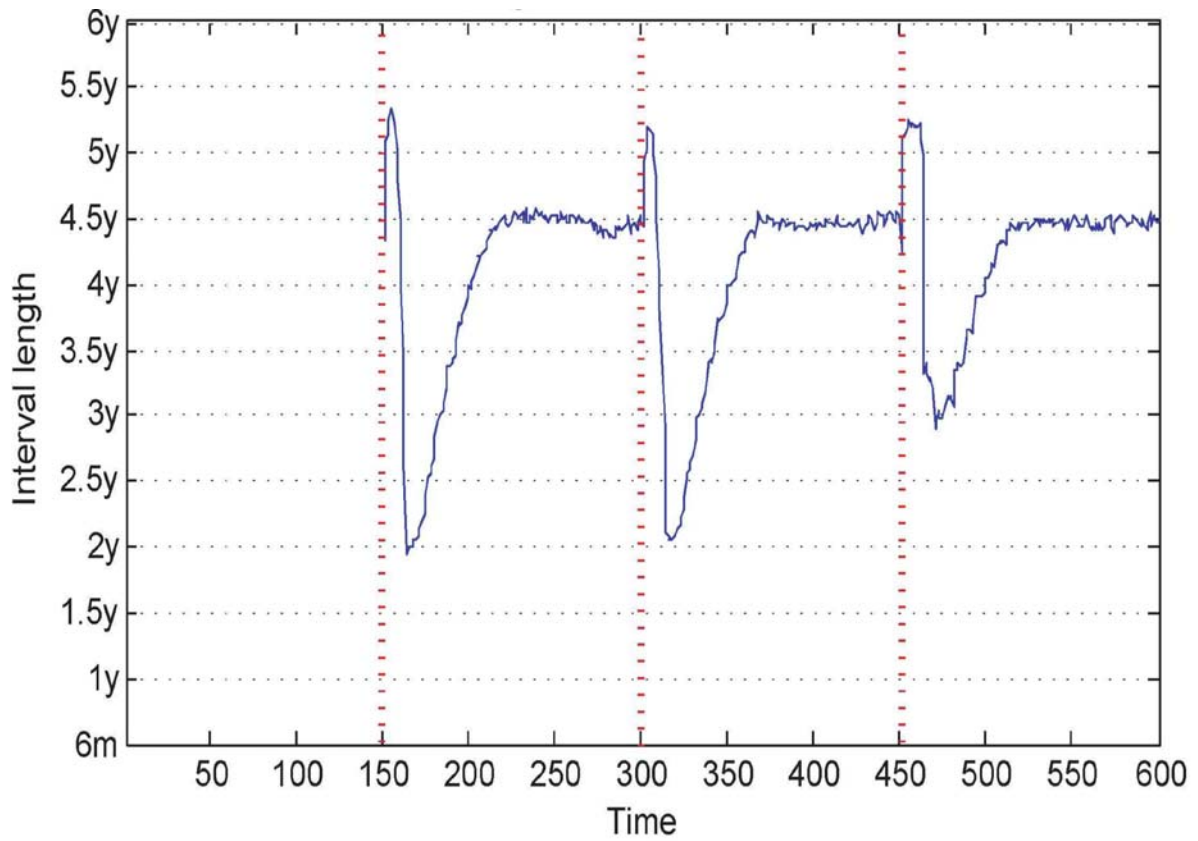


Figure 6. Parameter Evolution in the LARX Model



Notes: In each column we plot the time evolution 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 of inflation; the bottom row is the standard deviation of the error term.

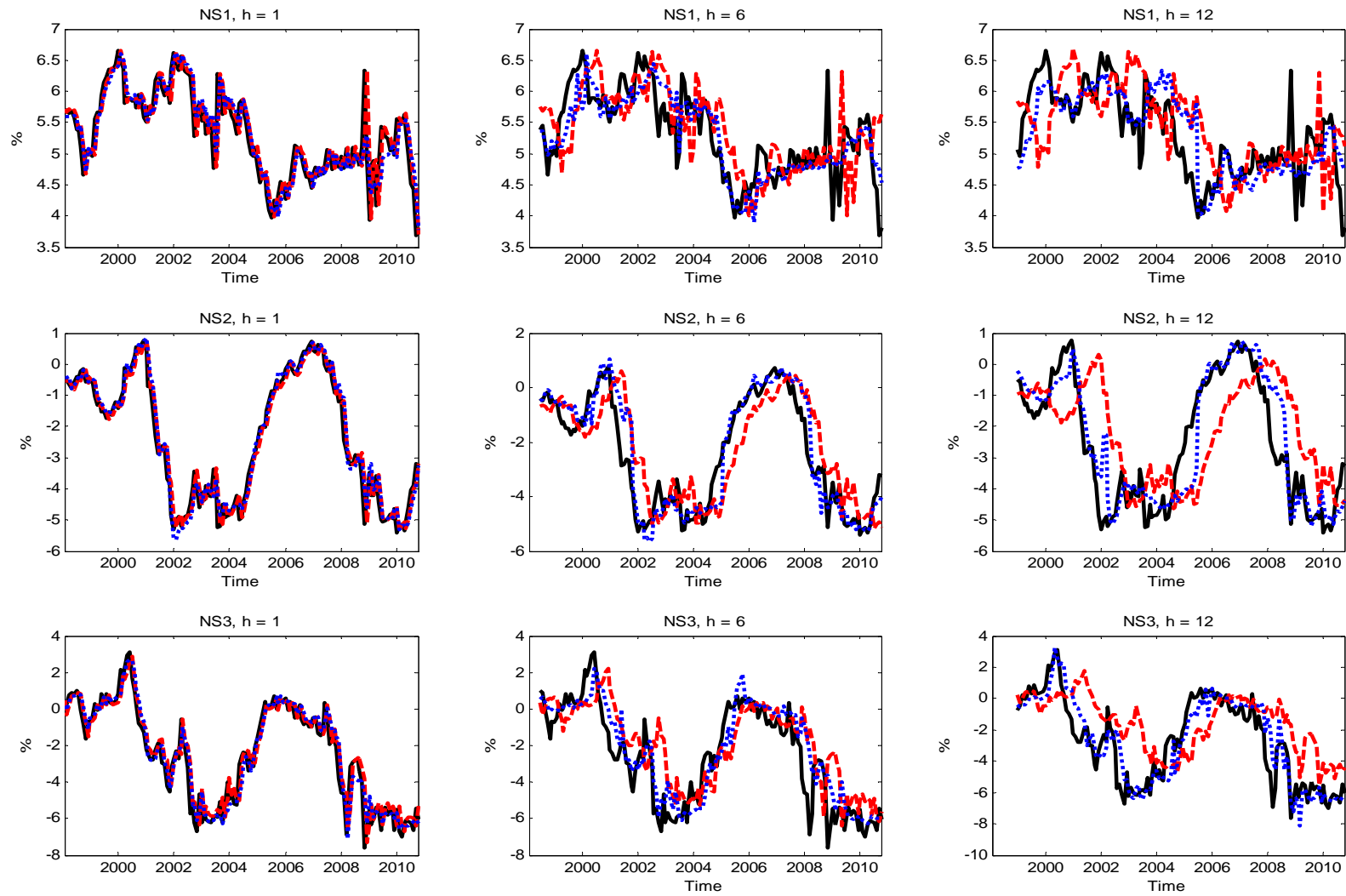
Figure 7. Average Lengths of Selected Subsamples



Notes:

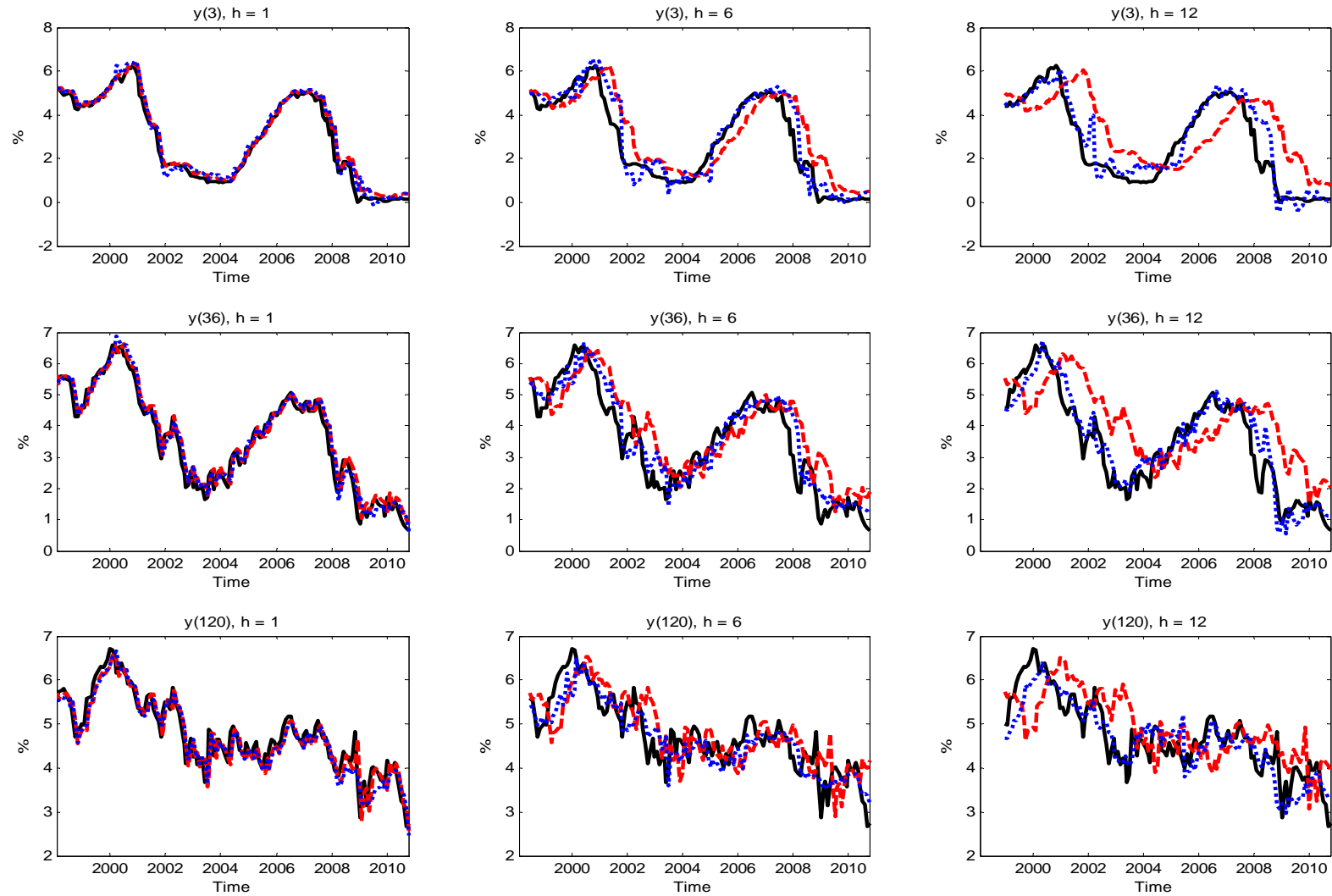
- 1) The unit of the vertical axis is a year, and the unit of the horizontal axis is a month.
- 2) The blue line indicates the average length of selected subsamples at each time point out of the results of the 1000 simulated series.

Figure 8. Plots of NS Factors with ADNS-LARX and DNS-VAR Recursive Forecasts



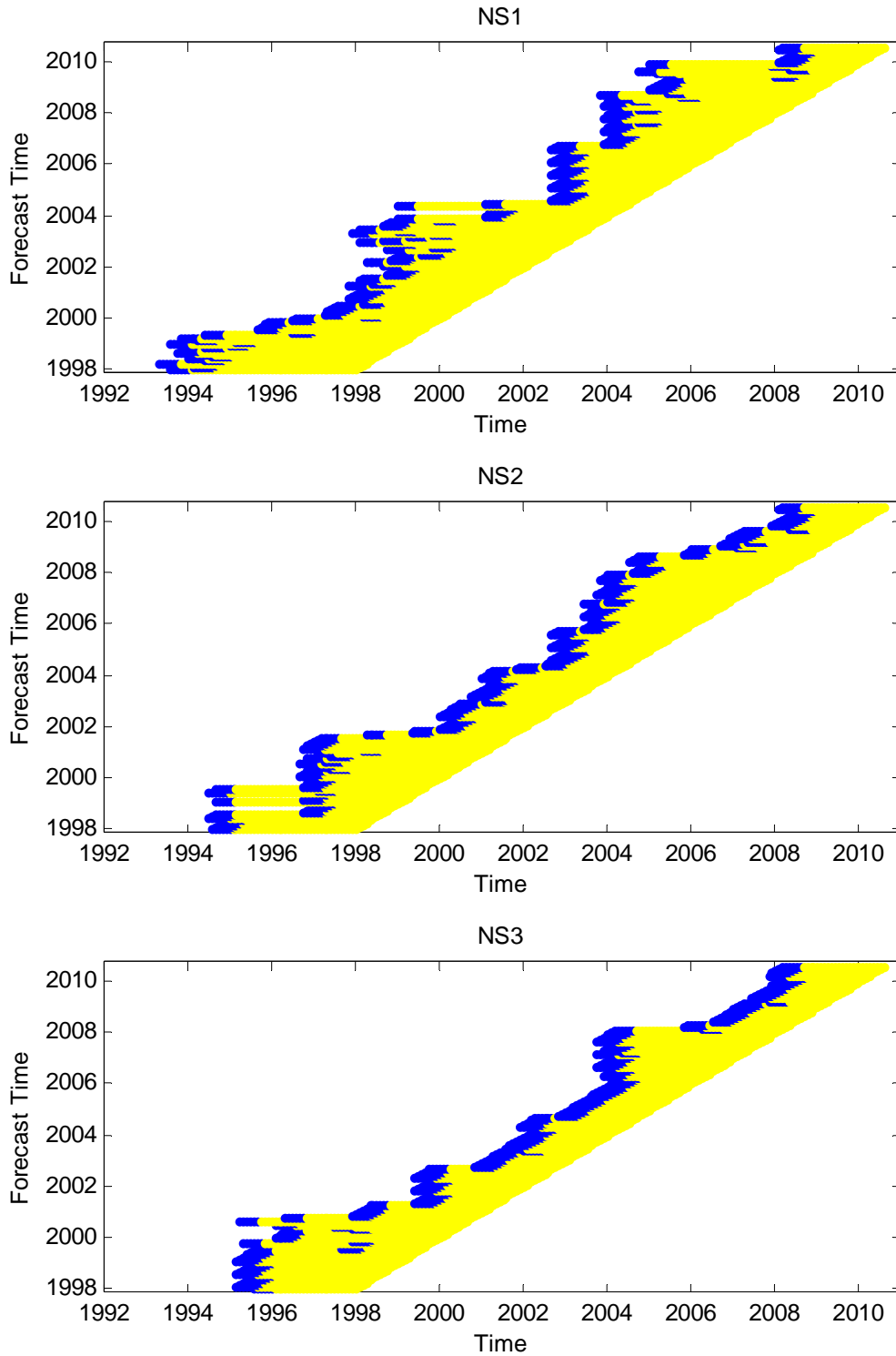
Notes: The solid lines indicate factors extracted from real yields; dotted lines are the ADNS-LARX forecasts; dashed lines are the DNS-VAR recursive forecasts.

Figure 9. Plots of Yields with ADNS-LARX and DNS-VAR Recursive Forecasts



Notes: The solid lines indicate real yields; dotted lines are the ADNS-LARX forecasts; dashed lines are the DNS-VAR recursive forecasts.

Figure 10. Breaks of Factors Detected in the ADNS-LARX Model



Notes: The vertical axis marks the time when the estimation is made. The selected stable sample interval is marked horizontally as a light solid line. The dark line marks the 6-month interval during which the most recent break is detected.