TERG

Discussion Paper No.304

Term Structure Modeling and Forecasting of Government Bond Yields

Does Macroeconomic Factors Imply Better Out-of-Sample Forecasts?

Wali Ullah and Yasumasa Matsuda

October 5, 2012

TOHOKU ECONOMICS RESEARCH GROUP

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
27-1 KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

Term Structure Modeling and Forecasting of Government Bond Yields

Does Macroeconomic Factors Imply Better Out-of-Sample Forecasts?

by
Wali Ullah and Yasumasa Matsuda
Graduate School of Economics and Management
Tohoku University
Sendai, Japan
October 5, 2012

Abstract

Accurate modeling and precise estimation of the term structure of interest rate are of crucial importance in many areas of finance and macroeconomics as it is the most important factor in the capital market and probably the economy. This study compares the in-sample fit and out-of-sample forecast accuracy of the CIR and Nelson-Siegel models. For the in-sample fit, there is a significant lack of information on the short-term CIR model. The CIR model should also be considered too poor to describe the term structure in a simulation based context. It generates a downward slope average yield curve. Contrary to CIR model, Nelson-Siegel model is not only compatible to fit attractively the yield curve but also accurately forecast the future yield for various maturities. Furthermore, the non-linear version of the Nelson-Siegel model outperforms the linearized one. In a simulation based context the Nelson-Siegel model is capable to replicate most of the stylized facts of the Japanese market yield curve. Therefore, it turns out that the Nelson-Siegel model (non-linear version) could be a good candidate among various alternatives to study the evolution of the yield curve in Japanese market.

Key Words: Non-linear Least Square, Simulation, Maximum Likelihood, In-sample fit, Forecasting, Yield Curve.

JEL Codes: C32, C53, C58, G12, G13, G17.

1. Introduction

Nothing in economy is watched much closer on a minute by minute basis than the yield curve. The central banks around the world try to manage it and everyone tries to forecast it. Its shape is a key to the profitability of many businesses and investment strategies. Equally important is the ability of the model to forecast the future term structure as it can be interpreted as a predictor of the future state of economy.[1] Therefore, accurate modeling, estimation and precise forecasting of the term structure of interest rate are of crucial importance in many areas of finance and macroeconomics.

Although the prices of zero-coupon bonds can be directly used to construct the term structure, however, due to the limited available maturity spectrum and lack of market liquidity of the zero-coupon bonds, it is essential to estimate the yields based on the observed coupon-bearing bond prices. Therefore, several term structure models have been developed in the course of time to plot the yield curve. A model that forms the basis of many other term structure models is the Vasicek (1977) model. The innovative feature of the Vasicek (1977) is that it models the interest rate as a mean reversion process. A famous extension to the Vasicek model is the Cox-Ingersoll-Ross (1985) model, which aims to cope with some of the drawbacks of the Vasicek model. The Cox *et al.* (1985) model describes the evolution of the short rates and distills the entire term structure by only one stochastic variable. Other famous extensions are the Vasicek and Fong (1982), Hull and White (1990) and Black *et al.* (1990) models.

However, more positive results have emerged recently based on the framework of Nelson and Siegel (1987). Originally intended to describe the cross sectional aspects of the yield curves, the Nelson-Siegel model imposes a parsimonious three-factor structure on the link between yields of different maturities, where the factors can be interpreted as level, slope and curvature. Though statistical in nature, the standard Nelson-Siegel model is still widely used due to its good fit of the observed term structure.[2]

This chapter discusses the Cox-Ingersoll-Ross (CIR) model and the Nelson-Siegel exponential components framework to distill the entire term structure of zero-coupon yields. Being derived from dynamic stochastic general equilibrium (DSGE) specification, the CIR model was characterized for theoretical purposes, whereas, the motivation for the Nelson-Siegel model comes from the stylized facts that can be inferred from empirical analysis. The CIR model is compared with the Nelson-Siegel model to find out which of the two classes is appropriate for forecasting purposes. The comparison between the Nelson-Siegel and the CIR models will help to find out which of the two can appropriately represent the true characteristics of the market. We also compare the in-sample fit of Nelson-Siegel model for the linear and non-linear estimation

[1] These forecasts are used by companies in their investment decisions and discounting future cash flows, consumers in their saving decisions, and economists in the policy decisions.

[2] For instance, De Pooter (2007) states that nine out of thirteen central banks that report their curve estimation methods to the Bank of International Settlements (BIS) use either the Nelson-Siegel model or its variation. Furthermore, Diebold and Li (2006) find that the dynamic reformulation of this model provides forecasts that outperform the random walk and various alternative forecasting approaches.

methods.

Furthermore, we simulate the CIR and the Nelson-Siegel models to find out whether simulation results match the larger trends and statistics (i.e., stylized facts) of the actual interest rate data. In this context, we aim to understand that:

— Which of the two classes of models well explain the entire term structure of interest rates?
— Does non-linear estimation of Nelson-Siegel model lead to a better in-sample fit than the linear estimation process?
— Does better fit imply reasonable simulation results?

The motivation to simulate interest rates may be to examine the out-of-sample performance of the two classes of term structure models. An interesting reading on this topic for the Nelson-Siegel model is in Diebold and Li (2006), which indicates that the model produces term structure forecasts at both short and long horizons with encouraging results.

The chapter contributes to the existing literature in two ways. In calibrating the multi-factor Nelson-Siegel model, we estimate the dynamic version of the model by employing the non-linear least squares estimation procedure and allow all the four parameters to vary over time.[3] We show that how the non-linearized version of the model (assuming the time-varying τ) leads to a better in-sample fit as compared to the linearized one. Secondly, we model the four time-varying factors of Nelson-Siegel model to simulate the yield curve, contrary to the previous studies in which parameter τ_t is fixed to a pre-specified value and they model three factors to forecast the term structure. Lastly, in estimating the CIR and Nelson-Siegel models, some new empirical facts will emerge from the Japanese market data. Of particular importance, short-term yields such as the three and six-month yields were essentially stuck at zero during most of the period from 2000 to 2006. It will also be interesting to figure out that how the short rate CIR model fits the very low short-term interest rate to compute the entire term structure.

We proceed as follows: in section 2 the Cox-Ingersoll-Ross (1985) model and the dynamic multi-factor Nelson-Siegel (1987) model are discussed. Section 3 describes the Japanese interest rate data and estimates the parameters of the models. We evaluate the forecasting performance of the two competing term structure models in section 4, while section 5 concludes the study.

2. Term Structure Models

The term structure of interest rates describes the relationship between interest rates and time to maturity. At a certain point of time for various maturities, the term structure can have different shapes. The curves that encounter in reality can be upward, downward sloping, flat or humped

[3] In the earlier studies, the parameter τ is pre-specified to a fixed value without estimation. For example, Diebold and Li (2006) argue that τ is to be taken as a constant with little degradation of fit, but it greatly simplifies the estimation procedure. They fix τ to 30 months that maximizes the loading of the curvature factor. Similarly, Fabozzi *et al.* (2005) set the shape parameter τ to 3 leaving the hump located at 5.38 years, arguing for the computational efficiency (no iterations through τ need to be performed). However, in some studies τ is considered as time invariant unknown parameter (does not pre-specify). Such as Diebold, *et al.* (2006) estimate τ to be 23.3 months ($\lambda = 0.077$). In Ullah *et al.* (2013), the estimated τ is 71.420 implying that the loading on the curvature factor is maximized at a maturity of about 6 years.

shape. These typical shapes can be generated by a class of functions associated with the solutions of differential or difference equations. Cox *et al.* (1985) developed a general equilibrium model with explicit analytical expression for the equilibrium interest rate dynamics and bond prices using the first order stochastic differential equation (SDE). Being a general equilibrium model, it contains all the elements of the traditional expectation hypothesis. On the other hand, Nelson and Siegel (1987) introduced a model for term structure which explains 96% of the variation of the yield curve across maturities with the help of second order differential equation.

For a zero-coupon bond with unit face value maturing in m periods and current price $P_t(m)$, the continuously compounded yield $R_t(m)$ is $R_t(m) = -m^{-1}\log[P_t(m)]$, where t denotes a moment in time. The instantaneous forward rate $f_t(m)$, which is the interest rate contracted now and to be paid for a future investment, can be obtained from the discount function as $f_t(m) = -[P_t^{'}(m)/P_t(m)]$, where $P_t^{'}(m) = \partial P_t(m)/\partial m$.[4] Furthermore, the relationship between the yield to maturity and the implied forward rates is $R_t(m) = m^{-1}\int_0^m f_t(u)du$. Given the yield curve or forward curve, we can price any coupon bond as the sum of the present values of future coupon and principal payments. This important relationship between zero-coupon and instantaneous forward rates is a critical component of the Nelson-Siegel model. Moreover, the short rate is the annualized interest rate (yield) for an infinitesimally short period of time and is defined as $r_t = R_t(0) = \lim_{m\to 0} R_t(m)$, whereas long rate is the annualized spot rate for long horizon maturity, defined as $l_t = R_t(\infty) = \lim_{m\to\infty} R_t(m)$. Based on these definitions and notations, in the next two sections we present the models.

2.1. Cox-Ingersoll-Ross (CIR) Model

Vasicek (1977) developed a one-factor model of the term structure which depends on only one uncertainty factor, i.e., the short rate. Vasicek defines the short rate process as:

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t \tag{1}$$

As with the mean reverting process, the three parameters κ, μ and σ are strictly positive and W_t is a Wiener process. A major drawback of the Vasicek model is that the model can produce negative interest rates.[5] Cox *et al.* (1985) adopt a general equilibrium approach to endogenously determine the risk-free rate. They reformulated the Vasicek model, in order to prevent the short rate from becoming negative, as:

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{2}$$

The $\kappa(\mu - r_t)dt$ is a drift term which represents the mean reversion and is similar to the drift term in the Vasicek model. The difference between the two models is the square root in the

[4] The function $P_t(m) = exp[-R_t(m)m]$, which relates the zero-coupon yield to the value of bond, is referred as discount function.

[5] If real interest rates are to be modeled, this does not necessarily have to be a big problem as real interest rates can be negative in reality. Nominal rates, on the contrary, will never be negative in practice.

second (volatility) term, which prevents the short rate from becoming negative.[6]

Furthermore, the short rate r_t as in (2) follows a non-central chi-square distribution with $(2q+2)$ degrees of freedom, and the parameter of non-centrality $2u$ is proportional to the current spot rate. The probability density of the interest rate r_{t_i} at time t_i conditional on $r_{t_{i-1}}$ at t_{i-1} is given as:

$$f_{\text{CIR}}\left(r_{t_i} | r_{t_{i-1}}; \xi, \Delta t\right) = c[\exp\left(-u-v\right)]\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2\sqrt{uv}\right) \tag{3}$$

where $\xi = (\kappa, \mu, \sigma)'$ is the parameters vector,

$$c = \frac{2\kappa}{\sigma^2[1-\exp(-\kappa\Delta t)]}$$
$$u = cr_{t_{i-1}} \exp(-\kappa\Delta t)$$
$$v = cr_{t_i}$$
$$q = \frac{2\kappa\mu}{\sigma^2} - 1$$

and $I_q\left(2\sqrt{uv}\right)$ is a modified Bessel function of the first kind of order q.

Valuing the zero-coupon bond, Cox *et al.* (1985) show that the pricing function in the CIR model can be expressed as:

$$P_t(m) = A_t(m)\exp[-B_t(m)]\, r_t \tag{4}$$

where

$$A_t(m) = \left[\frac{2\theta.\exp\left\{\frac{m}{2}(\theta+\kappa)\right\}}{2\theta + (\kappa+\theta)[\exp(m\theta)-1]}\right]^{2\kappa\mu/\sigma^2} \tag{5}$$

$$B_t(m) = \frac{2[\exp(m\theta)-1]}{2\theta + (\kappa+\theta)[\exp(m\theta)-1]} \tag{6}$$

$$\theta = \sqrt{\kappa^2 + 2\sigma^2} \tag{7}$$

The bond price in (4) is a decreasing concave function of maturity m and decreasing convex function of the short-term interest rate r_t and mean interest rate level μ. Furthermore, $P_t(m)$ is an increasing concave (decreasing convex) function of κ (the speed of adjustment parameter), if the short-term interest rate r_t is greater (less) than θ. The bond price is also an increasing

[6] When r_t approaches zero, the volatility term $\sigma\sqrt{r_t}$ approaches zero. In this case, the short rate will only be affected by the drift term, resulting the short rate to revert to the mean again. Cox *et al.* (1985) show that whenever $2\kappa\mu > \sigma^2$, the interest rate is strictly larger than zero. Furthermore, there is empirical evidence that whenever interest rates are high, the volatility is likely to be high as well, which justifies the volatility term in the CIR model.

concave function of the interest rate variance σ^2.[7]

Rewriting the expression for $P_t(m)$ in (4) and using the relation between bond pricing and yield to maturity, implies a function to compute the term structure of interest rate in the CIR model as:

$$R_t(m) = \frac{1}{m}\left[(B_t(m) r_t) - \log\big(A_t(m)\big)\right] \tag{8}$$

with $A_t(m)$, $B_t(m)$ and θ are as in (5), (6) and (7) respectively.

On a time grid $0 = t_0, t_1, t_2, \ldots\ldots$ with time step $\Delta t = t_i - t_{i-1}$, the discretized version of the CIR model is defined as:

$$r_{t+\Delta t} = r_t + \kappa(\mu - r_t)\Delta t + \sigma\sqrt{\Delta t}\sqrt{r_t}\varepsilon_t \tag{9}$$

with $\varepsilon_t \sim N(0,1)$. Various different shapes of the term structure can be computed by the CIR model by changing the parameters values in (8).

2.2. Nelson-Siegel Model

Motivation for Nelson-Siegel model comes from the expectation hypothesis. According to the expectation hypothesis, forward rates will behave in such a way that there is no arbitrage opportunity in the market. In other words, the theory suggests that implied forward rates are the rationally expected spot rates of the future periods. Nelson and Siegel (1987) propose that if spot rates are generated by a differential equation, then implied forward rates will be the solutions to this equation. Assuming a second-order differential equation, to describe the movements of the yield curve, with the assumption of real and equal roots, the solution will be the instantaneous implied forward rate function as:

$$f_t(m) = \beta_{1t} + \beta_{2t}\exp\left(\frac{-m}{\tau_t}\right) + \beta_{3t}\left[\left(\frac{m}{\tau_t}\right)\exp\left(\frac{-m}{\tau_t}\right)\right] \tag{10}$$

for $t = 1,2,\ldots,T$ and time-varying parameters vector $\psi_t = (\beta_{1t}, \beta_{2t}, \beta_{3t}, \tau_t)'$.

The model may be viewed as a constant plus a Laguerre function, that is, a polynomial times an exponential decay term, which belongs to a mathematical class of approximating functions. The solution for the yield as a function of maturity is:

$$R_t(m) = \beta_{1t} + \beta_{2t}\left[\frac{1-\exp(-m/\tau_t)}{m/\tau_t}\right] + \beta_{3t}\left[\frac{1-\exp(-m/\tau_t)}{m/\tau_t} - \exp\left(\frac{-m}{\tau_t}\right)\right] \tag{11}$$

The Nelson-Siegel specification of yield in (11) can generate several shapes of the yield curve including upward sloping, downward sloping and (inverse) humped shape with no more than one

[7] It is due to that larger σ^2 value indicates more uncertainty about future real production opportunities, and thus more uncertainty about future consumption. In such a world, risk-averse investors would value the guaranteed claim in a bond more highly.

maxima or minima. The functional form imposed on the forward interest rates as in (10) leads to a flexible, smooth parametric function of the term structure that is capable of capturing many of the typically observed shapes that the yield curve assumes over time and captures most of the properties of the term structure.

The limiting path of $R_t(m)$, as m increases, is its asymptote β_{1t}; and, when m is small, the limit is $(\beta_{1t} + \beta_{2t})$. β_{1t} is the asymptotic value of the spot rate function, which can be seen as the long-term interest rate and is assumed (required) to be positive $(\beta_{1t} > 0)$. Furthermore, the loading of β_{1t} equals one (constant and independent of m) and, therefore, the term structure at different maturities is affected by β_{1t} equally, which justifies the interpretation of β_{1t} as a level factor. The instantaneous short rate is given by $\beta_{1t} + \beta_{2t}$, which is constrained to be greater than zero. Furthermore, β_{2t} determines the rate of convergence with which the spot rate function approaches its long-term trend. The slope will be negative if $\beta_{2t} > 0$ and vice versa. The loading of β_{2t} approaches to one as $m \to 0$ and to zero as $m \to \infty$. Therefore, the yield curve is primarily affected by β_{2t} in the shorter run, so a change in β_{2t} implies a change in the slope of the term structure. Therefore, it is legitimate to interpret β_{2t} as the slope factor. The loading that comes with β_{3t} starts at 0, increases, and then decays to zero. Since, β_{3t} has the greater impact on medium-term yields and can be termed as the curvature factor, because it affects the curvature of the term structure. Furthermore, the parameter β_{3t} determines the size and the form of the hump, i.e., $\beta_{3t} > 0$ results in a hump, whereas $\beta_{3t} < 0$ produces a U-shape.

Finally, the parameter τ_t determines the maturity time at which the loading of the β_{3t} is optimal. It also specifies the location of the hump or the U-shape on the yield curve. Therefore, the range of shapes the curve can take is dependent on τ_t, it can be interpreted as a shape factor. The small values of τ_t, which have rapid decay in regressors, tend to fit low maturities interest rates quite well and larger values of τ_t lead to more appropriate fit of longer maturities spot rates. It has an interesting rule and economic interpretation as it shows a point of maturity m that separates the short rate from the medium/long-term rates.

3. Parameter Calibration and Estimation

Taking into account three dimensions–yield, maturity and time–of the data, different estimation methods can be used. To estimate the CIR model, one could choose to do a cross-sectional or time series estimation. For the Nelson-Siegel as the factors are time-dependent, one can proceed with cross-sectional or multivariate time series estimation. The differences between the estimates should be small if the employed model of the term structure is true. In this study, we estimate the CIR model using the time series data and the Nelson-Siegel model via the cross-sectional data for each observed month in the dataset.

3.1. Data

The data we use are monthly spot rates for zero-coupon and coupon-bearing bonds, generated using the pricing data of Japanese bonds and treasury bills. The standard way of measuring the term structure of interest rates is by means of the spot rates on zero-coupon bonds, however due to limited maturity spectrum and lack of market liquidity of zero-coupon bonds, longer maturity rates need to be derived from coupon-bearing treasury notes and bonds. In practice, we can therefore not observe the entire term structure of interest rates directly, but we need to estimate it using approximation methods.

We use the end-of-month price quotes (bid-ask average) for Japanese Government bonds, from January 2000 to December 2011, taken from the Japan Securities Dealers Association (JSDA) bonds files. In total, there are 144 months in the dataset. Following Fama and Bliss (1987) method, in the first stage, each month we calculate one day continuously compounded unsmoothed forward rates for the available maturities from the price data and in second stage, we sum the daily forward rates to generate end of month term structure of yield for all the available maturities. Furthermore, we pool the data into fixed maturities. Because not every month has the same maturities available, we linearly interpolate nearby maturities to pool into fixed quarterly maturities of 3, 6, 9, 12, 15, 18, 21, 24, …,300 months (100 maturities).

<<Table 1>>

<<Figure 1>>

In table 1, the descriptive statistics for the zero-coupon yields is presented. It shows that the average yield curve is upward sloping as the mean yield is increasing with maturity. Furthermore, the long rates are less volatile and persistent than short rates. It also seems that the skewness has downward trend with the maturity. Moreover, kurtosis of the short rates is lower than those of the long rates. Figure 1 provides a three-dimensional plot of the Japanese yield curve data. It is clearly visible that during 2000 to 2006 the short rates are nearly zero and on ward from 2006 there is an increasing trend in the yield for all the maturities.

3.2. Calibration of the Cox-Ingersoll-Ross Model

The parameters vector of the Cox-Ingersoll-Ross model $\xi = (\kappa, \mu, \sigma)'$, as introduced in (2), is estimated using the time series data. To estimate the parameters vector ξ by maximum likelihood method, we use the CIR density given in (3). For T be the number of observations, e.g., the number of months the interest rate is observed, the likelihood function is given by:

$$L(\xi) = \prod_{i=1}^{T} f_{CIR}\left(r_{t_i} \middle| r_{t_{i-1}}; \xi, \Delta t\right) \tag{12}$$

for $i = 1,2, \dots\dots, T$. Moreover, maximizing the log-likelihood function is often easier than maximizing the likelihood function itself, we take natural logarithm on both sides in (12), resulting in:

$$\log[L(\xi)] = T\log(c) + \sum_{i=1}^{T} \left[-u - v + \frac{q}{2}\log\left(\frac{v}{u}\right) + \log\left(I_q\left(2\sqrt{uv}\right)\right)\right] \tag{13}$$

Maximizing (13) over its parameter space yields maximum likelihood estimates $\hat{\xi}$.[8] Matlab built-in function `fminsearch` is used to minimize the negative log-likelihood function to obtain $\hat{\xi}$. However, direct implementation of the Bessel function $I_q(2\sqrt{uv})$ into Matlab causes the program to crash. A failure occurs because the Bessel function diverges to plus infinity on a high pace. To cope with this problem, scaled Bessel function [denoted by $I_q^{scaled}(2\sqrt{uv})$], defined as $I_q(2\sqrt{uv})[\exp(-2\sqrt{uv})]$, is used. To take the scaled Bessel function into account, the log-likelihood function in (13) is adjusted as:[9]

$$\log[L(\xi)] = T\log(c) + \sum_{i=0}^{T} \left[-u - v + \frac{q}{2}\log\left(\frac{v}{u}\right) + \log\left(I_q^{scaled}\left(2\sqrt{uv}\right)\right) + 2\sqrt{uv}\right] \tag{14}$$

We use the OLS estimators as the start values of the discrete version of the CIR model (9) for the optimization problem defined in (14). To estimate the parameter vector ξ, using (2.14), one can use the time series data of three months, six months, one year or two years maturity yields. Obviously, taking different yield data implies different parameter estimates. We choose to calibrate the model on the two years maturity yields, for two reasons. On the one hand, the CIR model is a short rate model, so the time to maturity should not be too large. On the other hand, taking a short maturity time, say three or six months, might yield strange estimates because of the extremely low interest rates and high volatilities in the initial years of the data from 2000 to the end of 2006.[10] Moreover, the data is on a monthly interval, the time step is set equal to 1/12.

The results of initial estimates of OLS along with the global optimal estimates, using the maximum likelihood method, are depicted in the first panel of table 2. Given the initial estimates, the maximum likelihood estimates (MLE) in panel 1 of table 2 shows that the fitted yield curve is upward sloping.

<<Table 2>>

Figure 2 (upper pane) plots the average observed and the estimated yield curve. It is clearly visible that the CIR model plots an upward sloping yield curve like the observed one. In the perfect case, the two curves would match exactly. However, we observe that estimated yield is closer to the actual yield curve up to two years maturity and the discrepancy between the two is

[8] Note that, as the logarithmic function is a monotonically increasing function, maximization of the likelihood function also maximizes the log-likelihood function. That is, the location of the maximum does not change.

[9] In (14) the term $2\sqrt{uv}$ appears because $[\exp(-2\sqrt{uv})]$ in the scaled Bessel function should be canceled out to keep the log-likelihood function the same.

[10] We also tried the 3 months, 6 months, one year and 18 months short rates and the results are almost same with the 24 months short rates results. However, the 24 months yield data fits the estimated yield curve a slightly better than the 3, 6, 12, and 18 months at short maturity. Estimated results are reported in appendix A.

the increasing function of maturity beyond two years, as the residuals curve is upward sloping. It may be largely due to the low interest rates from 2000 to 2006. In order to get deeper insight of the behavior of the yield curve during the prolonged period of zero policy rate, we also estimated the CIR model for the two sub-periods, i.e., sub-period 1 (January 2000 to December 2006) and sub-period 2 (January 2007 to December 2011). In the second panel of table 2, we provide the initial and MLE estimates for the two subsets of data, i.e., the zero interest rate period (2000 to 2006) and the non-zero interest rate period (2007 to 2011). Furthermore, the estimated yield curves for both sub-periods are depicted in the lower two panes of figure 2.

<<Figure 2>>

The maximum likelihood estimates for the first sub-period show that the fitted yield curve is negatively sloped, however for the second sub-period the estimated yield curve has an upward slope. Furthermore, the plots of estimated yield curve for both the sub-periods in figure 2 also support this view.

3.3. Estimation of the Nelson-Siegel Model

The Nelson-Siegel model in (11) forms the basis for our estimation procedure. For estimating the parameters of the model, we consider the functional form as:

$$R_t(m) = \Lambda(\tau_t)\beta_t + \varepsilon_t \tag{15}$$

where $R_t(m)$ denotes the (N×1) vector of yield rates at time t for N distinct maturities, $\Lambda(\tau_t)$ is (N×3) matrix of loadings and $\varepsilon_t \sim N(0, \sigma^2 I_N)$ is the error term, which accounts for whatever is not captured in the function $R_t(m)$ about how bonds are priced. The $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ is the vector of unknown parameters. Furthermore, τ_t is also unknown parameter.

Contrary to the prior studies, we do not fix τ_t to a pre-specified value, but allow it to vary over time and can be optimally determined in the estimation process in order to obtain a better in-sample fit. As the dynamic Nelson-Siegel function of spot rates results in a non-linear model, we employ the non-linear least squares method to estimate the model parameters $\psi_t = (\beta_{1t}, \beta_{2t}, \beta_{3t}, \tau_t)'$ for each month t. To minimize the sum of squared zero-yield errors, the objective function $F(\beta_t, \tau_t)$ is given by:

$$F(\beta_t, \tau_t) = [R_t(m) - \Lambda(\tau_t)\beta_t]^2 \tag{16}$$

We derive the analytical gradient $\nabla F(\beta_t, \tau_t)$ for the objective function in (16) and solve numerically for the optimal $\hat{\psi}_t$.[11] The analytical gradient of $F(\beta_t, \tau_t)$ is reported in appendix B.

[11] The optimization problem stated in (16) is non-convex and may have multiple local optima, which increases the dependence of the numeric solution on the starting values. Arbitrarily choosing the start parameters possibly may not reach to a global optimum. This phenomena is avoided by applying the one-dimensional grid search to the system to estimate τ_t (denoted as $\tilde{\tau}_t$) and substituted in the (11) to linearize the dynamic model. Subsequently, OLS is employed to estimate the parameter vector β_t (denoted as $\tilde{\beta}_t$). The grid search $\tilde{\tau}_t$ and the OLS estimated $\tilde{\beta}_{1t}$, $\tilde{\beta}_{2t}$ and $\tilde{\beta}_{3t}$ are used as the initial values to find the numeric solution of optimization problem defined in (16).

Moreover, following Nelson and Siegel (1987), we set τ_t to the median value estimated in the non-linearized version of Nelson-Siegel model (in previous stage) and estimate it by the ordinary least squares (OLS) in order to make a comparison between the linearized and the non-linearized versions of the model.

Applying the non-linear least squares to the yield data for each month gives us a time series of estimated parameters vector $\hat{\psi}_t$ and the corresponding panel of fitted yields $\hat{R}_t(m)$ and residuals $\hat{\varepsilon}_t$ (pricing errors). The first panel of table 3 shows the descriptive statistics of the estimates of the Nelson-Siegel model of the non-linear least squares method.

The estimated vector of parameters $\hat{\psi}_t$ is highly statistically significant.[12] From the autocorrelations in the table 3 (panel 1) of the four factors, we can see that the $\hat{\beta}_{3t}$ and $\hat{\beta}_{1t}$ are the most persistent, and that the second factor is a bit less persistent than the first. It suggests that long rates are slightly more persistent than short rates. Although the lag autocorrelation is reasonably high, the Augmented Dickey–Fuller (1979) test for unit root suggests that all the estimated factors $\hat{\beta}_{1t}$,$\hat{\beta}_{2t}$,$\hat{\beta}_{3t}$ and $\hat{\tau}_t$ are *I(0)* process and stationary at level.[13] However, $\hat{\beta}_{1t}$ solely determines the long run limiting behavior of the Nelson-Siegel model. The results also indicate that the residuals autocorrelation is low, justifying the use of non-linear least squares method. The average R^2 and residuals indicate that the average yield curve is fitted very well.

<<Table 3>>

Furthermore, the time series plot of the $\hat{\tau}_t$ in figure 3 shows that the optimal point of β_{3t} loading ranges from 1.6 to 10 years. It indicates that there is a large degree of variability in the $\hat{\tau}_t$ over the period selected. Testing the sample with the median value of $\hat{\tau}_t$ leads to a small loss of accuracy of the fitted curve but there is a large variation in the $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$.[14] The descriptive statistic results of estimated $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ for the fixed value of τ (median value of $\tau = 38.068$), estimated by OLS, are presented in the second panel of table 3.

The degree of loss of fit ranges from 1.4% to 5.7%. Comparing the results in panel 1 and 2 of table 3, there is significant difference in the estimated factors of Nelson-Siegel model for the two estimation processes. The linearized version of model either under-estimate or over-estimate the actual yield curve, whereas the non-linear estimation application leads to a reasonable fit of the yield curve. It suggests that standardizing the parameter τ_t to a prespecified value, not only reduces the degree of fit but also leads to a significant biased in the estimated parameters $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$.

Furthermore, to empirically test whether the factors β_{1t},β_{2t} and β_{3t} are legitimately called

[12] The p-value of individual t-statistic (not reported) is less than 0.03 in almost every period for of all the four factors.

[13] Based on the SIC criteria, optimal lag 3 has been selected for all the four variables in employing the augmented Dickey–Fuller unit-root test. The MacKinnon critical values for rejection of hypothesis of a unit root are -4.023at the one percent level, -3.441 at the five percent level and -3.145 at the ten percent level.

[14] The median value of $\hat{\tau}_t$ is 38.068.

a level, slope and curvature factors respectively, as suggested in Diebold and Li (2006), we construct a level, slope and curvature from the observed zero-coupon yields data and compare them with $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ (estimated with time-varying τ_t) respectively. The level of the yield curve (L_t) is defined as the 25-year yield. We compute the slope (S_t) as the difference between the 25-year and three-month yield and the curvature (C_t) is worked out as two times the two-year yield minus the sum of the 25-year and three month zero-coupon yields. The pairwise correlation of empirically defined factors and estimated (model based) factors is $\rho(L_t, \hat{\beta}_{1t}) = 0.694$, $\rho(S_t, \hat{\beta}_{2t}) = -0.741$ and $\rho(C_t, \hat{\beta}_{3t}) = 0.660$. Pairwise correlations between the estimated factors and the empirically defined level, slope and curvature is almost smaller by 0.28 points than the results of earlier empirical studies, particularly for the US and Canadian markets.[15] Furthermore, to be precise, the estimated correlation and the time series plot in figure 3 show that $\hat{\beta}_{1t}$, $\hat{\beta}_{2t}$ and $\hat{\beta}_{3t}$ may truly be called level, slope and curvature factors respectively, as the estimated factors and their empirical proxies seem to follow the same pattern.

<<Figure 3>>

Using the estimates of Nelson-Siegel model for both time-varying and fixed τ, in figure 4, we plot the implied average fitted yield curves, the actual yield curve and the residuals. It seems that the curve fits pretty well and the two vary quite closely for time-varying τ. It does, however, have difficulties at some dates, especially when yields are dispersed, with multiple interior minima and maxima. For the fixed τ the discrepancy between the actual and estimated average yield curve is clearly visible. It under-estimates the actual yield up to 30 months maturity and over-estimates beyond 30 months. Similarly, the average residuals plot in the right panel of figure 4 also supports this argument.

<<Figure 4>>

Furthermore, table 4 and figure 5 present the descriptive statistics and the three dimensional plot of the residuals of Nelson-Siegel model estimation by non-linear least squares (for the time-varying τ_t) respectively. It turns out that the fit is more appealing in most cases. Some months, however, especially those with multiple maxima and/or minima are not fitted very well. Multiple maxima and/or minima occur in the term structure of months in the mid-2005 and onward, which becomes apparent by the large residuals in these months.

<<Table 4>>

<<Figure 5>>

---

[15] Diebold and Li (2006) perform a similar exercise based on zero-coupon yields generated using end-of-month price quotes for U.S. treasuries, from 1985:01 through 2000:12. Their estimated correlations are $\rho(L_t, \hat{\beta}_{1t}) = 0.97$, $\rho(S_t, \hat{\beta}_{2t}) = -0.99$ and $\rho(C_t, \hat{\beta}_{3t}) = 0.99$. Similarly, Elen van (2010) used the monthly Canadian zero-coupon yields from 1986:01 to 2009:012 and has reported the correlations as $\rho(L_t, \hat{\beta}_{1t}) = 0.943$, $\rho(S_t, \hat{\beta}_{2t}) = -0.929$ and $\rho(C_t, \hat{\beta}_{3t}) = 0.784$.

In summary, there is a significant lack of information on the short-term CIR model to fit the term structure of interest rate. It is not capable to fit the yield curve as the discrepancy between the two curves is significantly large. Contrarily, the Nelson-Siegel model provides an evolution of the term structure closer to reality. It distills the term structure of interest rate quite well and can describe the evolution and the trends of the market. Fixing the τ to the median value leads to fit the yield curve better than the CIR model but not than the time-varying τ estimation process (non-linear least squares) of the Nelson-Siegel model.

4. Term Structure Forecasting

A good approximation to yield-curve dynamics should not only fit well in-sample, but also produces satisfactorily out-of-sample forecasts. In this section, we simulate the interest rates to find out whether the simulated yields for various maturities based on the CIR and Nelson-Siegel models can replicate the stylized facts of the actual observed yields data. The stylized facts derived from the actual yields data for Japanese bonds are:

1. The average yield curve is upward sloping and concave.
2. Short rates are more volatile than long rates.
3. Long rates are less persistent than short rates.
4. Skewness has the downward trend with the maturity.
5. Kurtosis of the short rates are lower than those of the long rates.

4.1. Forecasting with the Cox-Ingersoll-Ross Model

Using the parameters in panel 1 of table 2, we simulate the short rates using the discrete version of CIR model as in (2.9) for 10,000 times. The starting point of the short rates simulation process is the two-year yield at December 2011, being 0.071. Using the simulated short rates, the entire term structure of yield is computed by using equation (2.8), that is, we compute 10,000 matrices of (144×100), containing yields for all maturity times and for all months.

Table 5 displays descriptive statistics that are of interest (e.g., mean, variance and autocorrelations) of the simulated yields for various maturities. This table may be compared with the statistical properties of actual yields in table 1 (data section).

<<Table 5>>

Summary statistics in table 5 indicate that the CIR model is not capable of replicating the interest rates' general trends. The CIR model generates the same skewness coefficients, the same kurtosis and the same autocorrelations for all maturity times. The short rates seem more volatile than the long rates, although the volatility is underestimated for all maturity times compared to the actual yields data. Moreover, the mean has a downward trend with increasing maturity. Figure 6 shows a plot of the downward shaped average yield curve (averaged over simulation times), implying that the simulated yield curve is not in line with the first stylized fact. The figure also shows that the CIR model is capable to produce term structure's other shapes.

<<**Figure 6**>>

One may conclude that the CIR model performs unsatisfactorily and seems not useful in the simulation based context. As opposed to the CIR model, the Nelson-Siegel model does not fall within the standard class of affine term structure models. Therefore, yields forecasts and their stylized facts simulated with the Nelson-Siegel model will likely be significantly different from the yields produced by the CIR model.

4.2. Forecasting with the Nelson-Siegel Model

Since the four parameters of the Nelson-Siegel model give a full description of the term structure of interest rate, one can model them and can use various methodologies to make out-of-sample forecast of the yield curve.[16] Here, the four time-varying estimated Nelson–Siegel factors are modeled as univariate AR(1) processes to simulate the term structure of interest rate.[17] The yield forecasts based on underlying univariate AR(1) factor specifications are:

$$\hat{R}_t(m) = \Lambda(\hat{\tau}_t)\hat{\beta}_t \tag{17}$$

$$\hat{\psi}_t = A_0 + A_1\hat{\psi}_{t-1} + \varepsilon_t \tag{18}$$

where A_0 is (4×1) vector of constants, A_1 is (4×4) diagonal matrix, $\hat{\psi}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\tau}_t)'$ and $\varepsilon_t \backsim N(0, \Sigma)$ is (4×1) error vector. For comparison, we also include the VAR(1) forecasts of yield because the pairwise correlation between estimated factors is reasonably high. This might produce out-of-sample forecasts with greater accuracy. The multivariate VAR(1) model specification is the same as in (18), but we modify A_1 to be (4×4) full matrix rather than a diagonal matrix.

Estimation of AR(1) and VAR(1) models specified in (18) is straight forward. We estimate the parameters vector A_0 and matrix A_1 of both AR(1) and VAR(1) using the time series of $\hat{\psi}_t$ that we obtained from the non-linear least squares regression on (15) by employing the maximum likelihood method, assuming the normal density for ε_t. We use a forecasting period of ten years with a time step of one month. That is, we simulate 120 months, starting with the January 2012 until December 2021. Using the AR(1) and VAR (1) estimated parameters, we simulate the time series of size 120 months for 10,000 times.

Table 6 displays summary statistics of the four simulated factors for both AR(1) and VAR (1) specifications, averaged over number of simulations. This table may be compared with the actual estimated factors from table 3 (panel 1).

<<**Table 6**>>

Comparing the simulated Nelson-Siegel factors of AR(1) and VAR(1) models with the estimated factors in panel 1 of table 3, the results show that in terms of most descriptive statistical

[16] It is concluded in the previous section that the non-linear estimation (with time-varying τ) leads to a better fit of the yield curve; therefore, non-linear least squares estimated parameters are modeled to carry out the simulation exercise.

[17] Following Diebold and Li (2006), we also computed out-of-sample forecasts for one month, 6 months and 1 year. The summary results are given in appendix C for reference.

properties, particularly the mean, skewness and kurtosis, the VAR(1) simulated factors and estimated factors are close alternatives. However, relatively the estimated factors are less persistent than the simulated factors. In terms of lag autocorrelation, the estimated factors are almost similar to the AR(1) results but regarding the mean and other descriptive features the AR(1) overestimates $\hat{\beta}_2$ and $\hat{\beta}_3$ and accurately estimates the $\hat{\beta}_1$ and $\hat{\tau}$.

Averaged over the number of simulations and the different months, both the simulated yield curves are upward sloping (figure 7). Comparing the simulated yield curves with the actual in figure 7, one notices the curves to be much alike. This may be attributed to the fact that the standard Nelson-Siegel is not only capable to generate a better in-sample fit but also performs satisfactorily in out-of-sample forecasts. At lower maturities, the VAR(1) simulated average yield curve is a bit nearer to the actual yield curve but at longer maturities both the VAR(1) and AR(1) are identical. Overall the results show that VAR(1) can replicate the properties of the estimated yield features better than the AR(1) specification.

<<Figure 7>>

To check for the other stylized facts, we compute yields for all maturities, by substituting the simulated vector $\hat{\psi}_t$ (at each simulation) in (17), for all 120 different months. Accordingly, we compute 10,000 matrices – one for each scenario (simulation) – of dimensions (120×100), containing the yields on every month for all maturity times. Then, we compute the statistical properties that are of interest (e.g., variances and autocorrelations) of the simulated yields for all maturities. Table 7 shows the descriptive statistics of the simulated yields for maturities of 3, 6, 12, 18, 24, 36, 60, 120, 180, 240, and 300 months for AR(1) and VAR(1) specifications, that can be compared to the actual yield statistical properties in table 1 (section 3.1).

Here, it can be seen that the simulated short rates of both AR(1) and VAR(1) indeed are more volatile than the long rates. It also seems that in simulation the skewness catches the downward trend with maturity in both cases. Moreover, kurtosis of the simulated short rates are lower than those of the simulated long rates, as can also be found in the observed nominal yields.

<<Table 7>>

The numeric values of the average yield of actual yield data for various maturities resembles with the VAR (1) simulated yields. The volatilities of both AR(1) and VAR(1) are much smaller than the actual yield. The actual volatilities vary within the range of 0.207 and 0.348, whereas the simulated yields unconditional volatility of VAR(1) model vary between 0.002 and 0.004 and between 0.002 and 0.003 for AR(1) specification. Numeric values for the skewness coefficients and kurtosis, however, deviate from the observed yields. The Japanese data shows skewness coefficients between -1.934 and 1.360, the simulation shows values somewhere between -0.201 and -0.045 for AR(1) and between -0.500 and -0.086 for the VAR(1) model. Furthermore, the kurtosis ranging from approximately 2.079 and 8.291, while the simulation produces kurtosis ranging from roughly 1.970 up to 3.984 for both AR(1) and VAR (1) specifications. One may also deduce from table 7 that the simulated yield short rates are more persistent than the long rates as we observe in the nominal data.

In summary, we conclude that the CIR model cannot replicate the interest rates' general trends and should be considered weak to describe the term structure in the simulation based context. On the other hand, the out-of-sample forecast results of the Nelson-Siegel seem reasonably well. In a simulation based context, the Nelson-Siegel model is capable to replicate most of the stylized facts of the Japanese market yield curve and the VAR(1) based specification of factors is able to replicate the properties of the estimated factors as well as actual yield data better than the AR (1) model of the factors.

5. Conclusion

The term structure of interest rates is the most important factor in the capital markets and probably the economy. It is widely used for pricing contingent claims, determining the cost of capital and managing financial risk. In this study, we implement the CIR and the Nelson-Siegel models and compare the in-sample fit as well as the out-of-sample forecast performance using monthly Japanese government bonds zero-coupon data (yield to maturity) from January 2000 until December 2011.

For the in-sample fit, the results show that there is a significant lack of information on the short-term CIR model. The CIR model plots upward sloping yield curve, however, the discrepancy between the actual and the estimated is an increasing function of maturity beyond two years maturity. Contrary to CIR model, the Nelson-Siegel model provides an evolution of the term structure closer to reality. The Nelson-Siegel model is capable to distill the term structure of interest rate quite well and describe the evolution and the trends of the market. Furthermore, fixing the shape parameter τ to the median value leads to a better yield curve fit than the CIR model but not as striking as the time-varying τ estimation process (non-linear least squares) does.

Regarding the term structure forecast, we conclude that the CIR model cannot accomplish to replicate the interest rates' general trends. The CIR model generates the same skewness, kurtosis and autocorrelations for all maturity times. The volatility is underestimated for all maturity times and more importantly, it produces a downward slope average yield curve, implying that CIR model should be considered too poor to describe the term structure evolution in the simulation based context. On the other hand, the out-of-sample forecast results of the Nelson-Siegel model seem reasonably well. The Nelson-Siegel model is capable to replicate most of the stylized facts of the Japanese market yield curve. Between the AR(1) and VAR(1) specification of factors, the descriptive features of the actual yield data and estimated factors are more closely in line with the VAR(1) simulated yields features.

Summarizing, it turns out that the model proposed by Nelson and Siegel (1987) is compatible to fit attractively the yield curve (in-sample fit) and accurately forecast the future yields for various maturities. These successes account for the continued popularity of statistical class of models and its use by central banks around the world. Furthermore, the Nelson-Siegel model (non-linear version) could be a good candidate to study the evolution of the yield curve in Japanese market.

References

Black F, Derman E, Toy E. 1990. A one-factor model of interest rates and its application to Treasury bond options. *Financial Analyst Journal* 46: 33 - 39.

Bliss R. 1997. Testing term structure estimation methods. *Advances in Futures and Options Research* 9: 97–231.

Cox JC, Ingersoll JE, Ross SA. 1985. A theory of the term structure of interest rates. *Econometrica* 3: 385 - 407.

de Jong F. 2000. Time series and cross section information in affine term structure models. *Journal of Business and Economic Statistics* 18: 300–314.

De Pooter M. 2007. Examining the Nelson-Siegel class of term structure models. *Tinbergen Institute Discussion paper*, No. 2007-043/4.

Dickey DA, Fuller WA. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431.

Diebold FX, Li C. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130: 337-364.

Elen van EALJ. 2010. Term Structure Forecasting: Does a Good Fit Imply Reasonable Simulation Results? A thesis submitted to Tilburg School of Economics and Management, Tilburg University. [online] Available at < http://arno.uvt.nl/show.cgi?fid=112924> [Accessed 11th August, 2011].

Fabozzi FJ, Martellini L, Priaulet P. 2005. Predictability in the shape of the term structure of interest rates. *Journal of Fixed Income* 15-1: 40 - 53.

Fama E, Bliss R. 1987. The Information in long-maturity forward rates. *American Economic Review* 77: 680–692.

Hull J, White A. 1990. Pricing interest rate derivative securities. *Review of Financial Studies* 3: 573–592.

Nelson C. Siegel A. 1987. Parsimonious modeling of yield curves. *The Journal of Business* 60: 473-489.

Ullah W, Tsukuda Y, Matsuda Y. 2013. Term structure forecasting of government bond yields with latent and macroeconomic factors: Does macroeconomic factors imply better out-of-sample forecasts? *Forthcoming in Journal of Forecasting*.

Vasicek OA, Fong HG. 1982. Term structure modeling using exponential splines. *Journal of Finance* 37: 339–348.

Vasicek OA. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5: 177 - 188.

Appendix A

CIR Model Results for 3, 6, 12 and 18 Months Maturity Data

The results of initial estimates of OLS along with the MLE optimal estimates using the dataset for 3 months, 6 months, 12 months, and 18 months maturity periods for the entire sample (2000:01–2011:12) are depicted in table A-1. The results of MLE show that the average fitted yield curve is upward sloping. Figure A-1 plots the average observed yields and the estimated yield curves for all the four maturities data. It shows that the CIR model plots an upward sloping yield curve like the observed positively sloped average yield curve. However, the discrepancy between the estimated curves for all the four data sets and average observed yield curve is very high.

<<Table A-1>>

Furthermore, we estimate the CIR model for the two sub-periods, sub-period 1(2000:01–2006:12) and sub-period 2 (2007:01–2011:12) to observe the yield curve behavior during the prolonged period of zero policy rates. In table A-2, we provide the initial estimates and MLE estimated parameters for the two subsets of data, i.e., the zero interest rate period (2000–2006) and the non-zero interest rate period (2007–2011). Furthermore, the estimated yield curves for both the sub-periods are depicted in figure A-2.

<<Figure A-1>>
<<Table A-2>>

The maximum likelihood estimates for the first sub-period shows that the fitted yield curve is negatively sloped, however for the second sub-period the estimated yield curve has an upward slope for all the four maturities data sets.

Overall the results of 3 months, 6 months, 12 months, and 18 months maturities data sets generate the same yield curve as we have estimated using the two years maturity data for the overall sample as well as for the two sub-periods, however, the 24 months yield data fits the estimated yield curve slightly better than the 3 months, 6 months, 12 months, and 18 months at short maturities.

<<Figure A-2>>

Appendix B

Derivation of Analytical Gradient $\nabla F(\boldsymbol{\beta}, \boldsymbol{\tau})$ for the Non-Linear Ordinary Least Square of the Nelson-Siegel Model

To minimize the sum of squared zero-coupon yield errors, the objective function $F(\beta, \tau)$ is as given in (16):

$$F(\beta, \tau) = [R(m) - \Lambda(\tau)\beta]^2 \tag{A-1}$$

Differentiate the objective function in (A-1) w.r.t β and τ,

$$\frac{\partial F}{\partial \beta_1} = [-2(R(m) - b)] = 0 \tag{A-2}$$

$$\frac{\partial F}{\partial \beta_2} = [-2\tau a.(R(m) - b)] = 0 \tag{A-3}$$

$$\frac{\partial F}{\partial \beta_3} = [-2[\tau a - \exp(-m/\tau)].(R(m) - b)] = 0 \tag{A-4}$$

$$\begin{aligned}\frac{\partial F}{\partial \tau} = \Bigg[2\Bigg\{&-\beta_2\left(a + \frac{\exp\,(-m/\tau)}{\tau}\right) \\ &-\beta_3\left(a - \frac{\exp\,(-m/\tau)}{\tau} - \frac{\exp\,(-m/\tau).m}{\tau^2}\right)\Bigg\}.(R(m) - b)\Bigg] = 0\end{aligned} \tag{A-5}$$

where

$$a = \frac{[1 - \exp(-m/\tau)]}{m}$$

$$b = -\beta_1 - \beta_2 a\tau - \beta_3[a\tau - \exp(-m/\tau)]$$

The system of equations derived analytically in (A-2), (A-3), (A-4) and (A-5) is non-linear and can be solved numerically. The numerical solution of the system implies to the Nelson-Siegel estimated factors vector $\hat{\psi}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\tau}_t)'$.

Appendix C

Out-of-Sample Forecast Performance of the Nelson-Siegel Model

We follow the Diebold and Li (2006) method and model the estimated four time-varying factors of Nelson-Siegel model as first order auto-regressive and vector auto-regressive and make out of sample forecast for one month, 6 months and 1 year horizons.[18] The yield forecasts based on underlying univariate AR(1) factor specifications are:

$$\hat{R}_{t+h}(m) = \Lambda(\hat{\tau}_{t+h})\hat{\beta}_{t+h} \qquad \text{(A-6)}$$
$$\hat{\psi}_{t+h} = A_0 + A_1\hat{\psi}_t + \varepsilon_{t+h} \qquad \text{(A-7)}$$

where A_0 is (4×1) vector of constants, A_1 is (4×4) diagonal matrix, $\hat{\psi}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\tau}_t)'$ and $\varepsilon_{t+h} \backsim N(0, \Sigma)$ is (4×1) error vector. A_0 and A_1 are obtained by regressing $\hat{\psi}_t$ on $\hat{\psi}_{t-h}$. The multivariate VAR(1) model specification is same as in (A-7) but we modify A_1 to be (4×4) full matrix rather than a diagonal matrix.

We estimate and forecast recursively, using data from January 2000 to the time that the forecast is made, beginning in January 2008 and extending through December 2011. Subsequently, we substitute the forecasted factors $\hat{\psi}_{t+h}$ at time t in (A-6) to get the forecasted yield denoted as $\hat{R}_{t,t+h}(m)$.

In tables A-3, A-4 and A-5, we compute the descriptive statistics of *h-month-ahead* out-of sample forecasting results of the dynamic Nelson–Siegel models of AR(1) and VAR(1) representation of $\hat{\psi}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\tau}_t)'$, for maturities of 3, 6, 12, 18, 24, 36, 60, 120, 180, 240 and 300 months for the forecast horizons of $h = 1, 6$ and 12 months.

We define forecast errors at $t + h$ as $[R_{t+h}(m) - \hat{R}_{t,t+h}(m)]$, where $\hat{R}_{t,t+h}(m)$ is the forecasted yield in period t for $t + h$ period and is not the Nelson–Siegel fitted yield. $R_{t+h}(m)$ is the actual yield in period $t + h$. We examine a number of descriptive statistics for the forecast errors, including mean, standard deviation, mean absolute error (MAE), root mean squared error (RMSE) and autocorrelations at various displacements.

The results of one month ahead forecast of AR(1) and VAR(1) representation are reported in table A-3. The one month ahead forecasting results appear suboptimal as the forecast errors appear serially correlated. The average forecast errors and RMSE are much smaller than that of the related work such as Bliss (1997), de Jong (2000) and Diebold and Li (2006). In relative terms, RMSE comparison at various maturities reveals that AR(1) forecasts are slightly better than the VAR(1), however in term of serial correlation of errors the VAR(1) outperform the AR(1) specification.

<<Table A-3>>

[18] Diebold and Li (2006) model the three estimated factors of Nelson-Siegel model $\hat{\beta}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t})'$ as they assume the shape parameter τ_t is constant. Contrarily, we model the four estimated factors of Nelson-Siegel model $\hat{\psi}_t = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\tau}_t)'$, assuming a time-varying τ_t.

The results in table A-4 and A-5 of 6 months and one year ahead forecast respectively, reveal that matters worsen radically with longer horizons forecast. For 6 months ahead forecast, the AR(1) forecasts are slightly better than the VAR(1), while for the 12 months ahead, the VAR(1) performs better than the AR(1) in terms of lower RMSE. However, in regard of auto-correlation of the forecast errors, VAR(1) outperforms AR(1) for all maturities in both 6 and 12 months ahead forecasts.

<<Table A-4>>

<<Table A-5>>

Furthermore, we also compute the Trace Root Mean Squared Prediction Error (TRMSPE) which combines the forecast errors of all maturities and summarizes the performance of each model, thereby allowing for a direct comparison between the models.[19] In table A-6, we report the TRMSPE for both the specifications of yield curve factors, i.e., AR(1) and VAR(1) for all the three forecasts horizons.

<<Table A-6>>

The performances of AR(1) is to some extent superior to that of the VAR(1) model of factors in terms of TRMSPE for the one month and six months ahead forecasts horizons, while the VAR(1) outperform the AR(1) for twelve months ahead forecasts. It suggests that for longer horizons forecasts the multivariate VAR(1) specification of factors can forecast the future yields with greater accuracy than the univariate AR(1) model of factors.

In summary, the out-of-sample forecast results of the Nelson-Siegel seem reasonably well in terms of lower forecast errors, however the errors are serially correlated. These results are slightly different from Dieobld and Li (2006). In term of lower RMSE, our results for all the three horizons forecast are preferred than that of related studies. Diebold and Li (2006) have a great success in forecasts, particularly in terms of the errors persistency, using a different dataset with maturities up to 10-year, whereas we have maturities up to 25-year. The original Nelson-Siegel framework might forecast the long maturities sub-optimally. The serial correlation of forecast errors may likely come from a variety of sources, some of which could be eliminated, such as, pricing errors due to illiquidity may be highly persistent and could be reduced by including variables that may explain mispricing as suggested by Dieobld and Li (2006).

---

[19] Given a sample of T out-of-sample forecasts of N distinct maturities with $h-$months ahead forecast horizon, we compute the TRMSPE as follows:

$$TRMSPE = \sqrt{\frac{1}{NT}\sum_{m=1}^{N}\sum_{t=1}^{T}\left[R_{t+h}(m) - \hat{R}_{t,t+h}(m)\right]^2}$$

where $\hat{R}_{t,t+h}(m)$ is the forecasted yield in period t for $t+h$ period, $[R_{t+h}(m) - \hat{R}_{t,t+h}(m)]$ is the forecast errors at $t+h$ for yield.

Tables

Table 1: Descriptive Statistics of Yield Curve Data

Maturity	Mean	S. Dev.	Max.	Min.	Skewness	Kurtosis	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
3	0.167	0.348	0.692	0.002	1.346	3.259	0.892	0.530	0.077
6	0.164	0.345	0.733	0.004	1.367	3.469	0.877	0.548	0.081
9	0.176	0.339	0.770	0.003	1.348	3.412	0.874	0.555	0.092
12	0.224	0.327	0.812	0.004	1.003	2.600	0.878	0.450	-0.001
15	0.250	0.327	0.855	0.003	0.956	2.487	0.870	0.459	0.021
18	0.276	0.304	0.990	0.013	0.974	2.589	0.873	0.455	0.018
21	0.303	0.303	0.990	0.027	0.932	2.475	0.877	0.451	0.022
24	0.327	0.292	1.027	0.019	0.896	2.382	0.875	0.434	0.025
30	0.387	0.284	1.117	0.027	0.871	2.368	0.865	0.403	0.026
36	0.446	0.281	1.186	0.078	0.815	2.315	0.862	0.383	0.035
48	0.594	0.280	1.368	0.121	0.653	2.133	0.855	0.326	0.027
60	0.730	0.273	1.517	0.161	0.509	2.079	0.856	0.269	0.027
72	0.864	0.265	1.627	0.216	0.365	2.137	0.849	0.210	0.025
84	1.011	0.262	1.759	0.285	0.214	2.234	0.842	0.129	0.035
96	1.165	0.260	1.878	0.382	-0.009	2.418	0.830	0.066	0.051
108	1.302	0.246	1.951	0.474	-0.224	2.784	0.832	0.056	0.091
120	1.424	0.231	1.998	0.549	-0.535	3.457	0.830	0.042	0.102
180	1.801	0.217	2.24	0.758	-1.388	6.203	0.841	-0.009	0.183
240	2.061	0.209	2.525	0.934	-1.934	8.291	0.850	-0.018	0.152
300	2.267	0.207	2.860	1.070	-1.774	7.983	0.874	-0.114	-0.045
Level	2.267	0.207	2.860	1.070	-1.774	7.983	0.874	-0.114	-0.045
Slope	2.099	0.311	2.842	1.031	-0.571	4.081	0.874	-0.043	-0.292
Curvature	-1.781	0.382	-0.993	-2.489	0.293	1.972	0.867	-0.017	-0.037

Note: The table shows descriptive statistics for monthly yields at different maturities and for the yield curve level, slope and curvature, where we define the level as the 25-year yield, the slope as the difference between the 25-year and 3-month yields, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 25-year yields. The last three columns contain sample autocorrelations at displacements of 1, 12 and 24 months. The sample period is 2000:01–2011:12. The number of observations is 144.

Table 2: Results of the MLE Estimation of the CIR Model

	$\hat{\kappa}$	$\hat{\mu}$	$\hat{\sigma}$	log L
Panel 1. Full Period Sample Results (2000:01–2011:12)				
Initial (OLS)	0.9287	0.0030	0.0809	
MLE	1.6149	0.0031	0.0775	6702.800
Panel 2. Results for Two Sub-Periods Samples				
Sub-Period I (2000:01–2006:12)				
Initial (OLS)	0.6185	0.0035	0.0760	
MLE	1.4591	0.0030	0.0738	3881.000
Sub-Period II (2007:01–2011:12)				
Initial (OLS)	1.6540	0.0033	0.0879	
MLE	2.1960	0.0035	0.0838	2935.000

Note: The table presents the initial OLS and MLE estimated results of $\hat{\xi}$ vector using the time series data of two years maturity. log L denotes the log likelihood value of the MLE estimation. Panel 1 consists the results of the full sample period, 2000:01–2011:12 (144 observations), while panel 2 presents the results for two sub-periods, i.e., sub-period 1 (2000:01–2006:12) and sub-period 2 (2007:01–2011:12). The number of observations for the first sub-period and second sub-period is 84 and 60 respectively.

Table 3: Descriptive Statistics of the Nelson-Siegel Estimated Factors

	$\hat{\beta}_{1t}$	$\hat{\beta}_{2t}$	$\hat{\beta}_{3t}$	$\hat{\tau}_t$	$\hat{\varepsilon}_t$	R^2
Panel 1. Non-Linearized Version of the Model (Time-varying τ_t)						
Mean	2.940	-2.759	-2.426	46.876	0.000	0.996
Std. Dev.	0.417	0.391	1.925	6.156	0.000	0.002
Maximum	3.805	-1.374	5.201	119.999	0.000	0.999
Minimum	1.219	-3.671	-4.676	19.348	0.000	0.987
Skewness	-1.566	0.891	1.420	1.530	0.068	-1.355
Kurtosis	6.690	4.943	5.156	4.996	2.796	6.116
$\hat{\rho}$ (1)	0.802	0.784	0.840	0.688	0.015	0.497
$\hat{\rho}$ (12)	0.055	-0.027	0.112	0.127	-0.067	-0.070
$\hat{\rho}$ (24)	-0.118	-0.355	-0.208	-0.128	-0.066	-0.040
ADF Stat.	-4.255	-4.147	-3.163	-5.129	-11.789	-
Panel 2. Linearized Version of the Model ($\tau = 38.068$)						
Mean	2.055	-2.989	-2.831	-	0.000	0.956
Std. Dev.	0.118	0.177	0.775	-	0.015	0.012
Maximum	3.094	-1.266	3.036	-	0.071	0.988
Minimum	1.124	-3.275	-3.395	-	-0.083	0.931
Skewness	0.614	0.388	-0.722	-	0.205	-0.893
Kurtosis	2.133	1.789	2.544	-	2.371	3.940
$\hat{\rho}$ (1)	0.866	0.857	0.860	-	0.215	0.436
$\hat{\rho}$ (12)	0.275	0.399	0.439	-	-0.178	0.014
$\hat{\rho}$ (24)	-0.168	-0.091	0.053	-	--0.125	-0.134
ADF Stat.	-3.355	-3.324	-3.297	-	-7.756	-

Note: The table presents descriptive statistics for Nelson-Siegel estimated factors, R^2 and $\hat{\varepsilon}$ averaged over the different maturity times using monthly yield data 2000:01–2011:12. Panel 1 presents the features of the results obtained from non-linearized version of the Nelson-Siegel model by applying non-linear least squares method, while panel 2 shows the features of the results estimated by ordinary least squares (OLS) methods for pre-specified value (median value obtained from non-linear estimation) of the shape parameter ($\hat{\tau} = 38.068$). $\hat{\rho}\ (i)$ denotes the sample autocorrelations at displacements of 1, 12, and 24 months. The last row contains augmented Dickey–Fuller (ADF) unit root test statistics. The number of observations is 144.

Table 4: Descriptive Statistic of the Nelson-Siegel Yield Curve Residuals for Time-varying τ

Maturity	Mean	S. Dev.	MAE	RMSE	Skewness	Kurtosis	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
3	-0.004	0.023	0.022	0.023	0.278	1.370	0.740	0.229	0.034
6	-0.011	0.019	0.020	0.022	0.874	2.381	0.579	0.038	0.061
9	-0.009	0.017	0.017	0.020	0.769	2.508	0.671	-0.036	-0.119
12	0.002	0.018	0.015	0.018	-0.174	1.833	0.587	0.300	0.111
15	0.008	0.018	0.017	0.019	-0.569	1.983	0.610	0.332	0.155
18	0.010	0.018	0.019	0.021	-0.753	2.243	0.572	0.211	0.122
21	0.011	0.018	0.019	0.021	-0.799	2.298	0.650	0.079	0.200
24	0.008	0.018	0.017	0.020	-0.551	2.044	0.623	0.119	0.167
30	0.007	0.019	0.018	0.020	-0.395	1.786	0.694	0.227	-0.090
36	0.001	0.022	0.020	0.022	-0.034	1.379	0.806	0.373	-0.096
48	0.008	0.020	0.018	0.020	-0.096	1.463	0.775	0.300	0.017
60	-0.008	0.020	0.019	0.021	0.568	1.826	0.743	0.308	0.111
72	-0.014	0.018	0.020	0.022	1.088	2.841	0.756	0.156	-0.072
84	-0.006	0.021	0.020	0.022	0.462	1.643	0.855	0.152	0.077
96	0.004	0.024	0.023	0.024	-0.255	1.283	0.891	0.420	0.072
108	0.008	0.019	0.018	0.020	-0.492	1.814	0.758	0.519	0.161
120	0.016	0.016	0.021	0.023	-1.437	3.975	0.425	0.188	0.022
180	-0.003	0.021	0.019	0.021	0.272	1.543	0.869	0.435	0.174
240	-0.005	0.017	0.016	0.018	0.353	1.844	0.795	0.334	-0.004
300	0.005	0.025	0.024	0.025	-0.385	1.265	0.856	0.403	0.029

Note: The table presents summary statistics of the residuals $\hat{\varepsilon}$ for different maturity times of the Nelson–Siegel model using monthly yield data 2000:01–2011:12 for time-varying τ. MAE is mean absolute errors, RMSE is the root mean squared errors and $\hat{\rho}\ (i)$ denotes the sample autocorrelations at displacements of 1,12, and 24 months. The number of observations is 144.

Table 5: Descriptive Statistics of the Simulated Yields Using the CIR Model

Maturity	Mean	S. Dev.	Max	Min	Skewness	Kurtosis	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
3	0.314	0.204	1.835	0.016	1.455	6.221	-0.008	-0.005	0.017
6	0.314	0.170	1.582	0.037	1.455	6.221	-0.008	-0.005	0.017
12	0.313	0.123	1.229	0.059	1.455	6.221	-0.008	-0.005	0.017
18	0.312	0.093	1.007	0.120	1.455	6.221	-0.008	-0.005	0.017
24	0.312	0.074	0.861	0.160	1.455	6.221	-0.008	-0.005	0.017
36	0.311	0.051	0.689	0.206	1.455	6.221	-0.008	-0.005	0.017
60	0.311	0.031	0.539	0.247	1.455	6.221	-0.008	-0.005	0.017
120	0.310	0.015	0.425	0.279	1.455	6.221	-0.008	-0.005	0.017
180	0.310	0.010	0.386	0.289	1.455	6.221	-0.008	-0.005	0.017
240	0.310	0.008	0.367	0.294	1.455	6.221	-0.008	-0.005	0.017
300	0.310	0.006	0.356	0.297	1.455	6.221	-0.008	-0.005	0.017

Note: The table shows descriptive statistics for simulated yields at different maturities for the CIR model. The entire term structure of yield is computed by the CIR yield curve model using the simulated short rates. The simulation exercise is done 10,000 times for 144 months. The last three columns contain the first, 12th and 24th order sample autocorrelation coefficients. The number of observations is 10,000.

Table 6: Descriptive Statistics of the Simulated Nelson-Siegel Factors

	AR(1)				VAR(1)			
	$\hat{\beta}_{1t}$	$\hat{\beta}_{2t}$	$\hat{\beta}_{3t}$	$\hat{\tau}_t$	$\hat{\beta}_{1t}$	$\hat{\beta}_{2t}$	$\hat{\beta}_{3t}$	$\hat{\tau}_t$
Mean	2.959	-2.799	-2.932	3.552	2.939	-2.753	-2.621	3.792
Std. Dev.	0.004	0.004	0.016	0.018	0.005	0.005	0.019	0.015
Maximum	2.966	-2.787	-2.892	3.596	2.953	-2.742	-2.583	3.824
Minimum	2.948	-2.807	-2.967	3.504	2.928	-2.764	-2.662	3.760
Skewness	-0.494	0.592	-0.101	-0.159	0.343	-0.178	-0.025	-0.015
Kurtosis	3.227	3.024	2.290	2.929	2.941	2.368	2.254	2.432
$\hat{\rho}$ (1)	0.786	0.359	0.115	-0.129	0.830	0.395	0.118	0.013
$\hat{\rho}$ (12)	0.881	0.453	0.123	-0.125	0.866	0.582	0.212	-0.292
$\hat{\rho}$ (24)	0.796	0.481	0.140	-0.070	0.622	0.333	0.016	-0.147

Note: The table presents descriptive statistics of the simulated Nelson-Siegel factors averaged over number of simulations for both AR(1) and VAR(1) specifications. The four factors of the Nelson-Siegel specification are modeled as first order AR and VAR to forecast the yield curve for 120 months, 2012:01–2021:12, for 10,000 times. The last three rows contain their first, 12^{th} and 24^{th} order sample autocorrelation coefficients. The computation of descriptive statistics is based on 120 observations.

Table 7: Descriptive Statistics of Simulated Yields Using the Nelson-Siegel Model

Maturity	Mean	S. Dev.	Max	Min	Skewness	Kurtosis	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
Simulated Yields Descriptive Statistics for AR (1) Specification									
3	0.192	0.003	0.196	0.189	-0.045	1.979	0.866	-0.195	-0.175
6	0.202	0.003	0.205	0.199	0.012	2.042	0.851	-0.254	-0.130
12	0.229	0.003	0.233	0.225	-0.105	2.338	0.837	-0.235	-0.116
18	0.265	0.003	0.269	0.260	-0.232	2.441	0.837	-0.159	-0.147
24	0.310	0.003	0.314	0.304	-0.290	2.427	0.840	-0.100	-0.178
36	0.414	0.003	0.418	0.408	-0.307	2.361	0.833	-0.037	-0.211
60	0.655	0.003	0.660	0.649	-0.198	2.353	0.745	-0.020	-0.187
120	1.235	0.003	1.241	1.228	-0.183	2.843	0.448	-0.033	0.037
180	1.657	0.002	1.665	1.648	-0.339	3.059	0.558	0.048	0.017
240	1.939	0.002	1.949	1.930	-0.291	3.029	0.652	0.089	-0.006
300	2.129	0.002	2.140	2.119	-0.204	3.986	0.708	0.107	-0.013
Simulated Yields Descriptive Statistics for VAR(1) Specification									
3	0.158	0.004	0.161	0.154	-0.086	2.214	0.865	-0.041	-0.175
6	0.160	0.004	0.164	0.156	-0.407	2.475	0.875	0.008	-0.199
12	0.176	0.003	0.181	0.171	-0.557	2.659	0.891	0.075	-0.188
18	0.206	0.002	0.211	0.199	-0.502	2.567	0.900	0.108	-0.160
24	0.246	0.002	0.252	0.239	-0.425	2.429	0.905	0.125	-0.134
36	0.349	0.002	0.354	0.342	-0.292	2.182	0.903	0.137	-0.091
60	0.601	0.002	0.606	0.595	-0.115	2.027	0.849	0.121	-0.017
120	1.221	0.002	1.227	1.215	0.017	2.349	0.632	0.019	-0.038
180	1.664	0.002	1.671	1.658	0.025	2.312	0.629	0.001	-0.131
240	1.955	0.002	1.962	1.947	-0.095	2.430	0.657	0.001	-0.180
300	2.148	0.002	2.155	2.140	-0.252	2.628	0.684	0.010	-0.202

Note: The table shows descriptive statistics for monthly simulated yields at different maturities for both AR(1) and VAR(1) specifications of the four factors vector $\hat{\psi}_t$ of the Nelson-Siegel Model. The four simulated factors are substituted in (2.17) to compute the simulated yields for various maturities for 120 months, 2012:01–2021:12, for 10,000 times. The last three columns contain the first, 12^{th} and 24^{th} order sample autocorrelation coefficients. The computation of descriptive statistics is based on 120 observations.

Table A-1: Results of the MLE Estimation of the CIR Model

Maturity		$\hat{\kappa}$	$\hat{\mu}$	$\hat{\sigma}$	log L
3 Months	Initial (OLS)	0.8729	0.0017	0.0983	
	MLE	1.4762	0.0017	0.0743	5969.100
6 Months	Initial (OLS)	1.1527	0.0017	0.1350	
	MLE	1.9030	0.0017	0.0821	5794.000
12 Months	Initial (OLS)	0.7615	0.0021	0.0982	
	MLE	1.5163	0.0022	0.0788	6190.400
18 Months	Initial (OLS)	0.8642	0.0026	0.0876	
	MLE	1.6859	0.0027	0.0807	6542.700

Note: The table presents the initial OLS and MLE estimated results of $\hat{\xi}$ vector using the time series data of 3 months, 6 months, 12 months, and 18 months maturities from 2000:01–2011:12. log L denotes the log likelihood value of the MLE estimation. The number of observations is 144.

Table A-2: Results of the MLE Estimation of the CIR Model for Sub-Periods

Maturity		$\hat{\kappa}$	$\hat{\mu}$	$\hat{\sigma}$	log L
		Sub-Period I (2000:01– 2006:12)			
3 Months	Initial (OLS)	1.3230	0.0011	0.1119	
	MLE	3.7744	0.0008	0.0825	3344.000
6 Months	Initial (OLS)	1.3230	0.0011	0.1119	
	MLE	3.7744	0.0008	0.0825	4121.000
12 Months	Initial (OLS)	0.6122	0.0024	0.1116	
	MLE	2.2455	0.0017	0.0885	3371.000
18 Months	Initial (OLS)	0.6577	0.0028	0.0899	
	MLE	1.9660	0.0023	0.0829	3708.600
		Sub-Period II (2007:01– 2011:12)			
3 Months	Initial (OLS)	0.8952	0.0020	0.0771	
	MLE	1.0691	0.0022	0.0696	2676.200
6 Months	Initial (OLS)	1.1414	0.0023	0.0779	
	MLE	1.3698	0.0024	0.0700	2709.500
12 Months	Initial (OLS)	1.4333	0.0027	0.0770	
	MLE	1.8056	0.0028	0.0701	2856.600
18 Months	Initial (OLS)	1.5149	0.0031	0.0852	
	MLE	2.0789	0.0032	0.0807	2910.600

Note: The table presents the initial OLS and MLE estimated results of $\hat{\xi}$ vector using the time series data of 3 months, 6 months, 12 months, and 18 months maturities for two sub-periods, i.e., sub-period 1 (2000:01 –2006:12) and sub-period 2 (2007:01 – 2011:12). log L denotes the log likelihood value of the MLE estimation. The number of observations for the first sub–period and second sub–period is 84 and 60 respectively.

Table A-3: Out-of-Sample 1 Month Ahead Forecasting Results

Maturity	Mean	Std. Dev.	MAE	RMSE	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
Forecast Summary for AR(1) Specification							
3	0.049	0.152	0.102	0.047	0.865	-0.067	0.000
6	0.022	0.143	0.102	0.039	0.850	-0.076	0.000
12	-0.022	0.148	0.126	0.028	0.892	-0.073	0.000
18	-0.046	0.181	0.163	0.037	0.870	-0.059	0.000
24	-0.079	0.197	0.187	0.039	0.856	-0.029	0.000
36	-0.120	0.227	0.227	0.050	0.821	0.000	0.000
60	-0.152	0.251	0.255	0.077	0.768	0.059	0.000
120	-0.003	0.188	0.145	0.065	0.547	0.093	0.000
180	0.098	0.169	0.157	0.052	0.495	0.045	0.000
240	0.128	0.162	0.172	0.045	0.550	-0.018	0.000
300	0.087	0.145	0.139	0.028	0.643	-0.067	0.000
Forecast Summary for VAR(1) Specification							
3	0.048	0.208	0.140	0.079	0.829	-0.048	0.000
6	-0.053	0.321	0.245	0.143	0.825	0.063	0.000
12	-0.222	0.552	0.147	0.515	0.841	0.112	0.000
18	-0.348	0.764	0.235	0.426	0.846	0.119	0.000
24	-0.460	0.928	0.383	0.073	0.847	0.123	0.000
36	-0.612	1.166	0.396	0.527	0.848	0.124	0.000
60	-0.735	1.375	1.080	1.012	0.847	0.126	0.000
120	-0.515	1.213	0.596	0.557	0.83	0.128	0.000
180	-0.277	0.944	0.761	0.376	0.826	0.120	0.000
240	-0.141	0.744	0.593	0.806	0.816	0.117	0.000
300	-0.107	0.598	0.477	0.478	0.826	0.111	0.000

The table presents the results of out-of-sample 1-month-ahead forecasting using AR (1) and VAR (1) specification of the estimated factors. We estimate all models recursively from 2000:1 to the time that the forecast is made, beginning in 2008:1 and extending through 2011:12. We define forecast errors at $t+1$ as $R_{t+1}(m) - \hat{R}_{t,t+1}(m)$, where $\hat{R}_{t,t+1}(m)$ is the $t+1$ month ahead forecasted yield at period t, and we report the mean, standard deviation, mean absolute errors and root mean squared errors of the forecast errors, as well as their first, 12^{th} and 24^{th} order sample autocorrelation coefficients.

Table A-4: Out-of-Sample 6 Months Ahead Forecasting Results

Maturity	Mean	Std. Dev.	MAE	RMSE	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
Forecast Summary for AR(1) Specification							
3	0.096	0.184	0.122	0.077	0.883	-0.067	0.000
6	0.078	0.172	0.116	0.066	0.867	-0.071	0.000
12	0.050	0.177	0.126	0.061	0.889	-0.065	0.000
18	0.037	0.208	0.154	0.082	0.872	-0.046	0.000
24	0.013	0.224	0.173	0.088	0.852	-0.017	0.000
36	-0.018	0.254	0.205	0.102	0.809	0.012	0.000
60	-0.049	0.278	0.234	0.107	0.755	0.070	0.000
120	0.070	0.210	0.169	0.080	0.571	0.104	0.000
180	0.145	0.185	0.191	0.070	0.527	0.060	0.000
240	0.159	0.176	0.200	0.059	0.587	-0.026	0.000
300	0.110	0.166	0.163	0.044	0.689	-0.081	0.000
Forecast Summary for VAR(1) Specification							
3	-0.660	0.451	0.698	0.095	0.820	0.040	0.000
6	-0.647	0.435	0.680	0.063	0.808	0.065	0.000
12	-0.622	0.434	0.661	0.057	0.813	0.075	0.000
18	-0.589	0.454	0.647	0.144	0.828	0.083	0.000
24	-0.573	0.463	0.642	0.132	0.830	0.090	0.000
36	-0.539	0.484	0.633	0.142	0.828	0.092	0.000
60	-0.474	0.502	0.597	0.140	0.820	0.101	0.000
120	-0.194	0.413	0.366	0.247	0.724	0.127	0.000
180	-0.008	0.363	0.265	0.271	0.666	0.116	0.000
240	0.083	0.312	0.228	0.239	0.636	0.100	0.000
300	0.086	0.270	0.201	0.183	0.643	0.070	0.000

The table presents the results of out-of-sample 6-month-ahead forecasting using AR (1) and VAR (1) specification of the estimated factors. We estimate all models recursively from 2000:1 to the time that the forecast is made, beginning in 2008:1 and extending through 2011:12. We define forecast errors at $t+6$ as $R_{t+6}(m) - \hat{R}_{t,t+6}(m)$, where $\hat{R}_{t,t+6}(m)$ is the $t+6$ months ahead forecasted yield at period t, and we report the mean, standard deviation, mean absolute errors and root mean squared errors of the forecast errors, as well as their first, 12^{th} and 24^{th} order sample autocorrelation coefficients.

Table A-5: Out-of-Sample 12 Months Ahead Forecasting Results

Maturity	Mean	Std. Dev.	MAE	RMSE	$\hat{\rho}$ (1)	$\hat{\rho}$ (12)	$\hat{\rho}$ (24)
			Forecast Summary for AR(1) Specification				
3	0.093	0.197	0.130	0.083	0.848	-0.003	0.000
6	0.075	0.183	0.125	0.070	0.874	-0.074	0.000
12	0.046	0.188	0.138	0.061	0.897	-0.070	0.000
18	0.032	0.216	0.167	0.075	0.896	-0.054	0.000
24	0.006	0.230	0.187	0.077	0.882	-0.029	0.000
36	-0.028	0.258	0.221	0.084	0.881	-0.068	0.000
60	-0.065	0.281	0.246	0.091	0.798	0.048	0.000
120	0.050	0.211	0.169	0.074	0.612	0.078	0.000
180	0.129	0.186	0.183	0.066	0.557	0.034	0.000
240	0.150	0.179	0.196	0.057	0.612	-0.049	0.000
300	0.108	0.170	0.163	0.045	0.716	-0.098	0.000
			Forecast Summary for VAR(1) Specification				
3	-0.081	0.150	0.135	0.041	0.590	-0.070	0.000
6	-0.091	0.135	0.129	0.034	0.510	-0.080	0.000
12	-0.106	0.135	0.144	0.031	0.599	-0.059	0.000
18	-0.109	0.162	0.171	0.033	0.699	-0.024	0.000
24	-0.127	0.185	0.198	0.038	0.722	0.026	0.000
36	-0.149	0.236	0.251	0.059	0.746	0.066	0.000
60	-0.172	0.309	0.305	0.111	0.758	0.087	0.000
120	-0.035	0.299	0.229	0.175	0.676	0.015	0.000
180	0.064	0.271	0.199	0.177	0.639	-0.049	0.000
240	0.100	0.236	0.191	0.139	0.609	-0.073	0.000
300	0.067	0.208	0.164	0.099	0.621	-0.073	0.000

The table presents the results of out-of-sample 12-month-ahead forecasting using AR (1) and VAR (1) specification of the estimated factors. We estimate all models recursively from 2000:1 to the time that the forecast is made, beginning in 2008:1 and extending through 2011:12. We define forecast errors at $t+12$ as $R_{t+12}(m) - \hat{R}_{t,t+12}(m)$, where $\hat{R}_{t,t+12}(m)$ is the $t+12$ months ahead forecasted yield at period t, and we report the mean, standard deviation, mean absolute errors and root mean squared errors of the forecast errors, as well as their first, 12th and 24th order sample autocorrelation coefficients.

Table A-6: TRMSPE Results for Out-of-Sample Forecasts Accuracy Comparisons

TRMSPE	1 Month Forecasts	6 Months Forecasts	12 Months Forecasts
AR(1) Model of Factors	0.046	0.076	0.079
VAR(1) Model of Factors	0.054	0.085	0.055

Note: The table reports the Trace Root Mean Squared Prediction Error (TRMSPE) results of out-of-sample forecasts accuracy comparison for horizons of one, 6, and 12 months for both AR(1) and VAR(1) specification of factors.

Figures

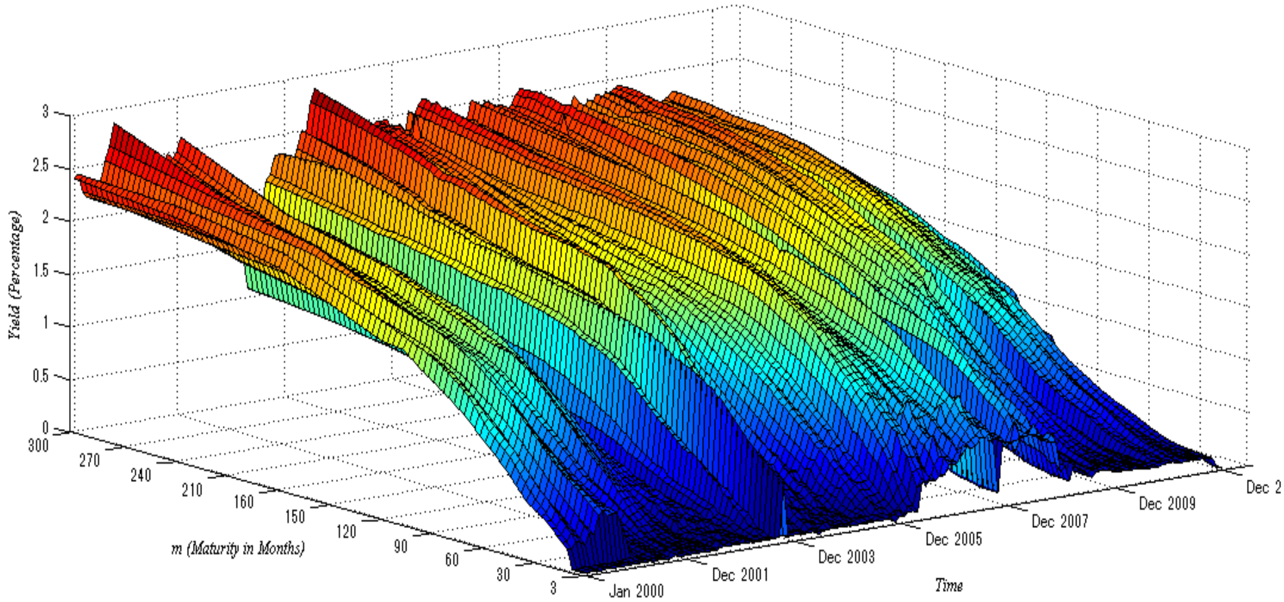

Figure 1. Yield Curves, 2000:01–2011:12.
The sample consists of monthly yield data 2000:01–2011:12 (144 months) at fixed quarterly maturities of 3, 6, 9, 12, 15, 18, 21, 24 …300 months (100 maturities).

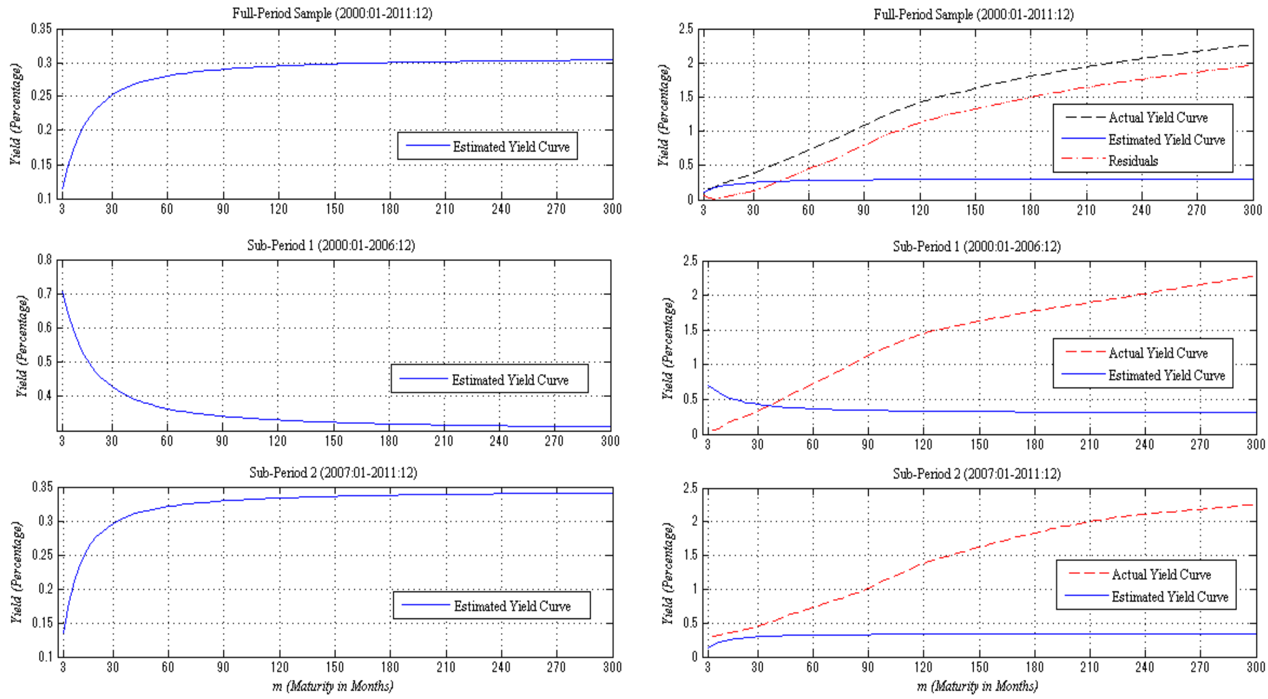

Figure 2: Fitted Yield Curve with the CIR Model

Actual average (data-based) and fitted (model-based) yield curve along the residuals for the entire sample (2000:01–2011:12) and two sub-periods, i.e., sub-period 1 (2000:01–2006:12) and sub-period 2 (2007:01–2011:12) are plotted. The fitted yield curves are obtained by evaluating the CIR function at the MLE estimated $\hat{\kappa}, \hat{\mu}$ and $\hat{\sigma}$ from the table 2.

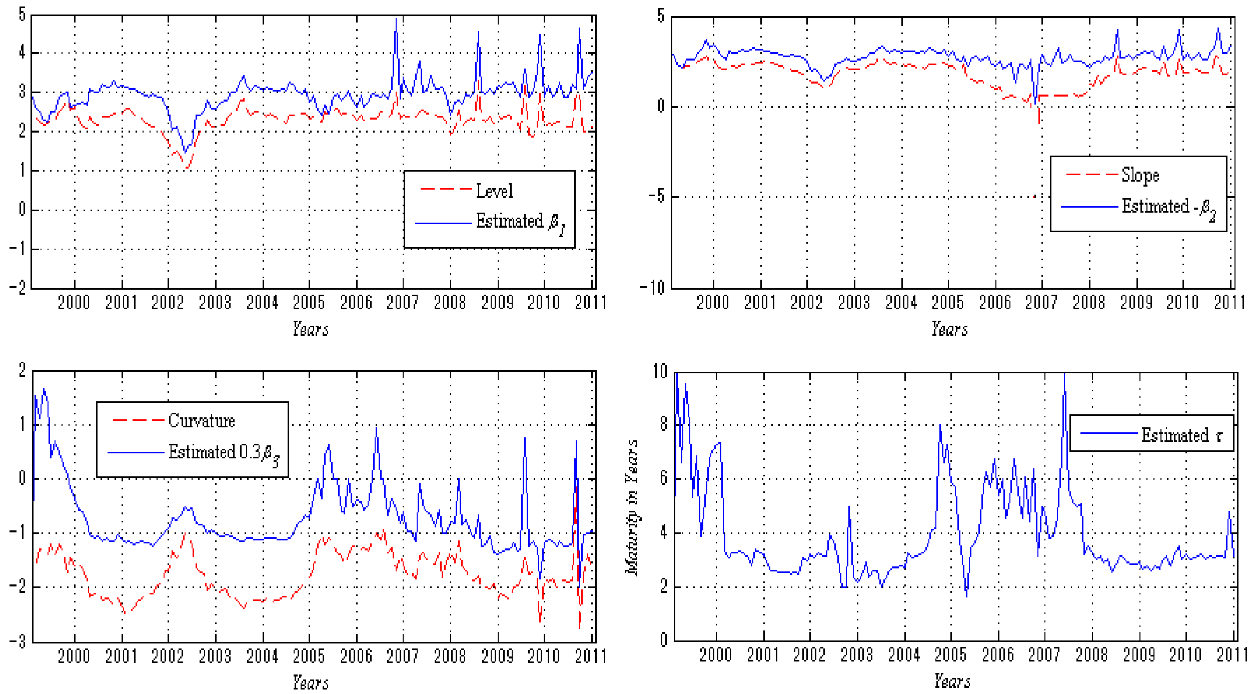

Figure 3: Time Series Plot of Nelson-Siegel Estimated Factors and Empirical Level, Slope and Curvature

Model-based level, slope and curvature (i.e., estimated factors vector $\hat{\beta}_t$) for time-varying τ_t vs. data-based level, slope and curvature are plotted, where level is defined as the 25-year yield, slope as the difference between the 25-year and 3-month yields and curvature as two times the 2-year yield minus the sum of the 25-years and 3- month zero-coupon yields. Rescaling of estimated factors is based on Diebold and Li (2006).

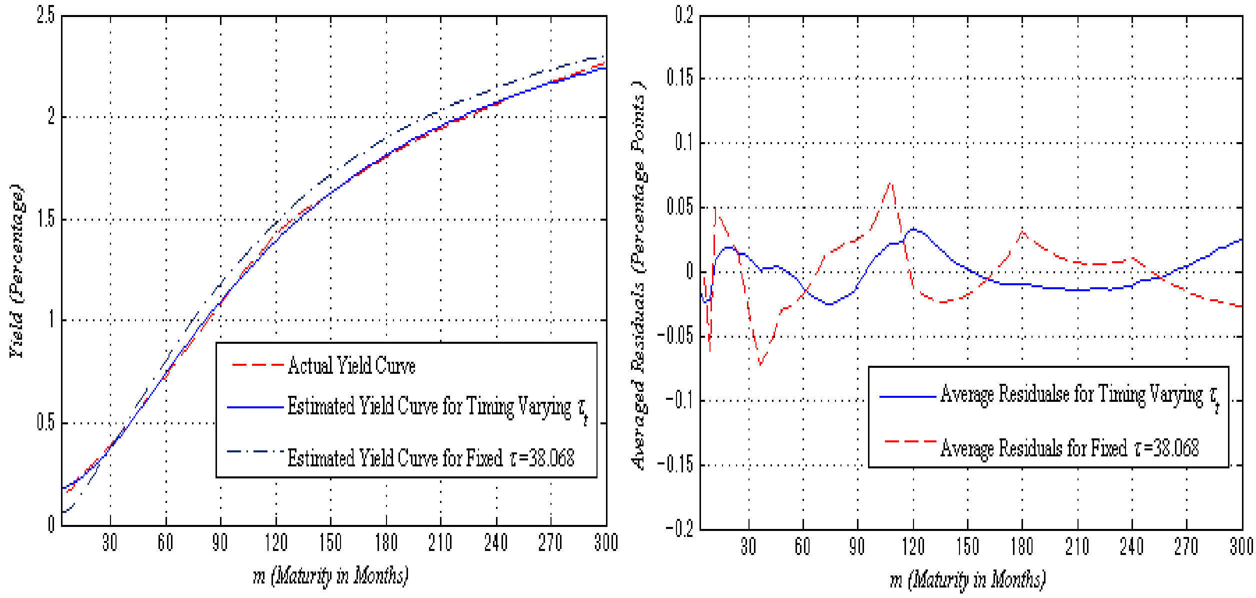

Figure 4: Average Fitted Yield Curve and Residuals of the Nelson–Siegel Model
Actual (data-based) and estimated (model-based) average yield curves and average residuals for both time-varying $\hat{\tau}_t$ and fixed $\hat{\tau} = 38.068$ are plotted. The fitted yield curves are obtained by taking average of the estimated yield of the Nelson-Siegel model over 144 months. Similarly, the residuals are also averaged over 144 months for the various maturities.

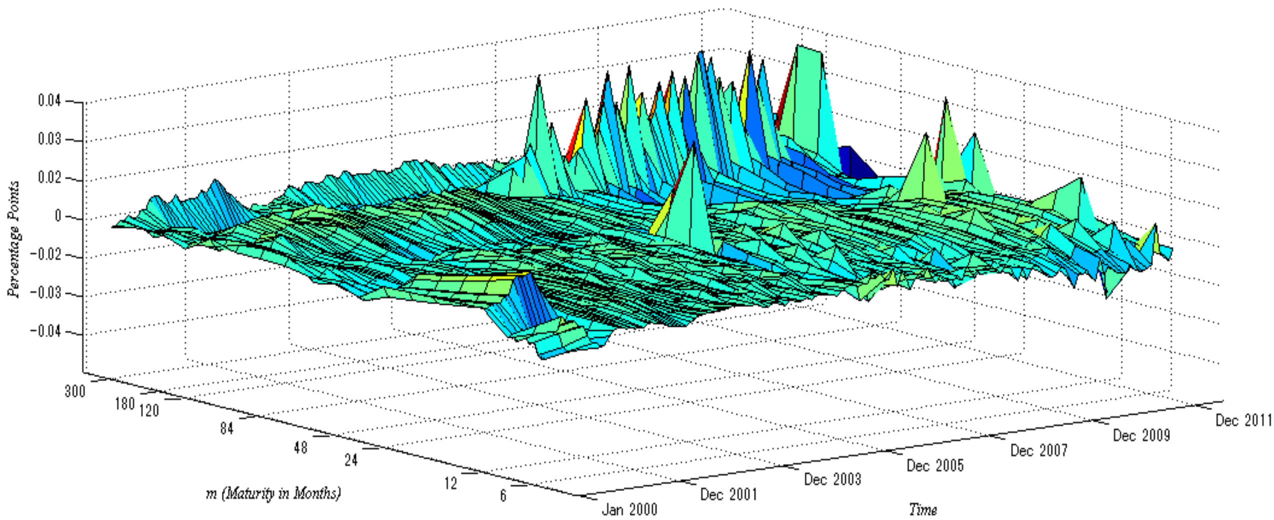

Figure 5: Nelson–Siegel Model based Yield Curves Residuals, 2000:01–2011:12 for Time-varying τ.
The sample consists of monthly residuals, obtained from the non-linear least squares estimation of the Nelson-Siegel model using the data 2000:01–2011:12 (144 months), at fixed quarterly maturities of 3, 6, 9, 12, 15, 18,…300 months.

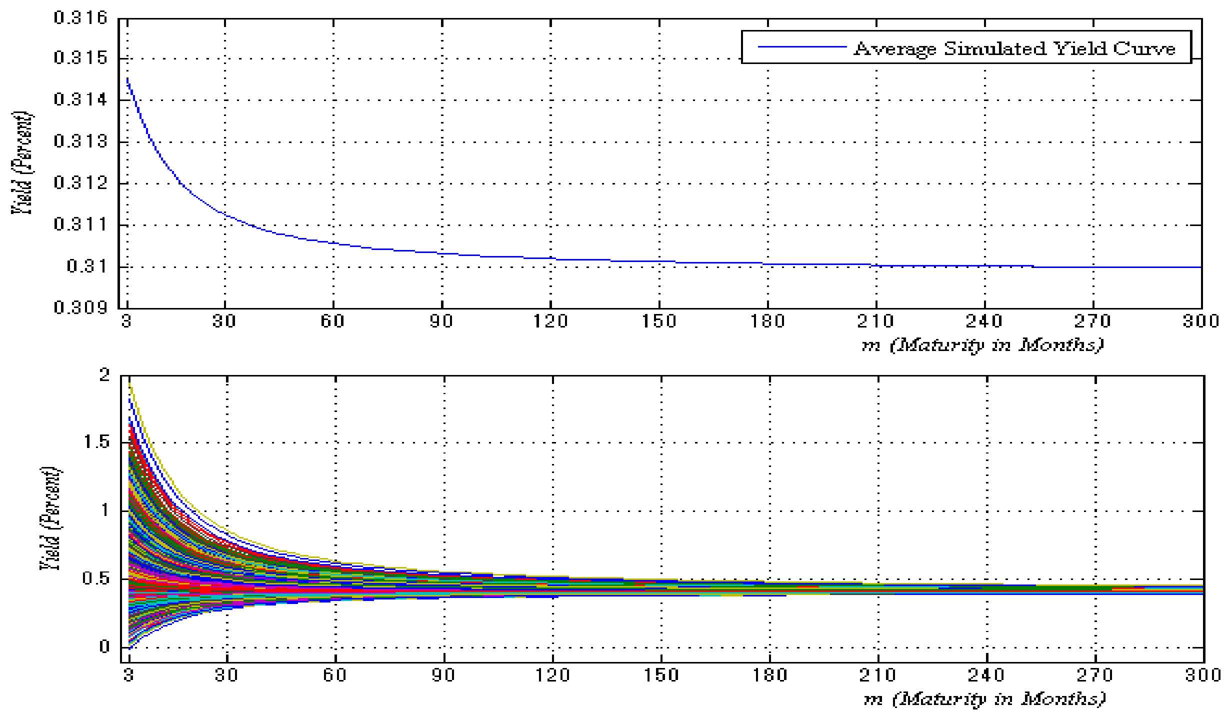

Figure 6:Average and All Simulated Yield Curves with the CIR Model
The entire term structure of yield is computed by the CIR yield curve model using the simulated short rates. The simulation exercise is done 10,000 times for 144 months. The 10000 simulated yield curves along with average simulated yield curve are plotted at fixed quarterly maturities of 3, 6, 9, 12, 15, 18, 21, 24 …300 months (100 maturities).

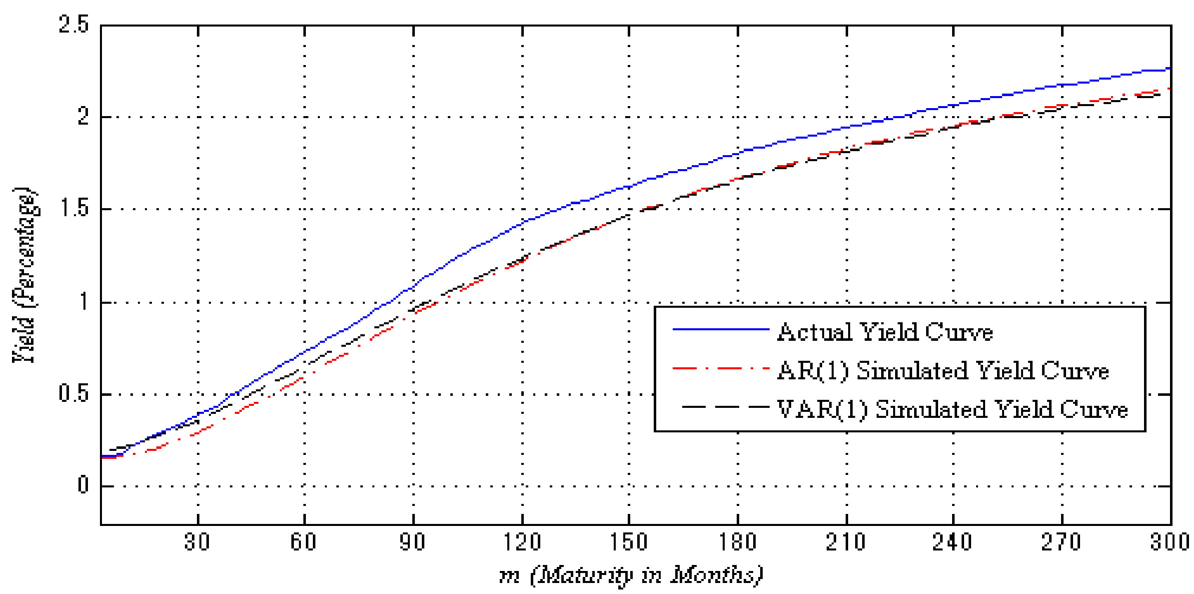

Figure 7: Simulated Average Yield Curves with the Nelson-Siegel Model
The four factors of the Nelson-Siegel specification are modeled as first order AR and VAR to forecast the yield curve for 120 months, 2012:01–2021:12, for 10,000 times. The average simulated yield curves for both AR(1) and VAR(1) specifications are obtained by averaging the simulated yields over different months as well as number of simulations. Actual (data-based) average yield curve is also plotted for comparison. All three yield curve are plotted at fixed quarterly maturities of 3, 6, 9, 12, 15, 18, 21, 24 …300 months (100 maturities).

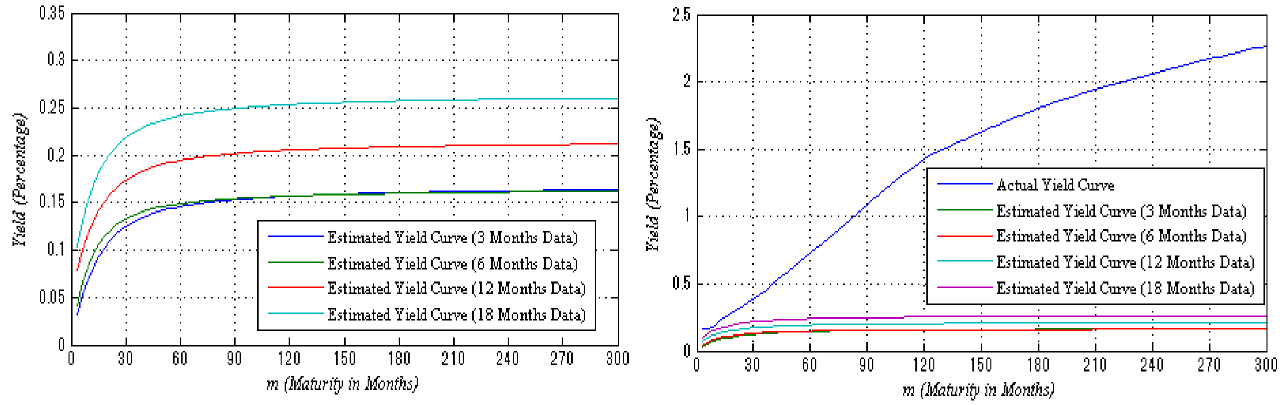

Figure A-1: Fitted Yield Curves with the CIR Model
Actual average (data-based) and fitted (model-based) yield curves for various maturities are plotted. The fitted yield curves are obtained by evaluating the CIR function at the MLE estimated $\hat{\kappa}, \hat{\mu}$ and $\hat{\sigma}$ from table A-1.

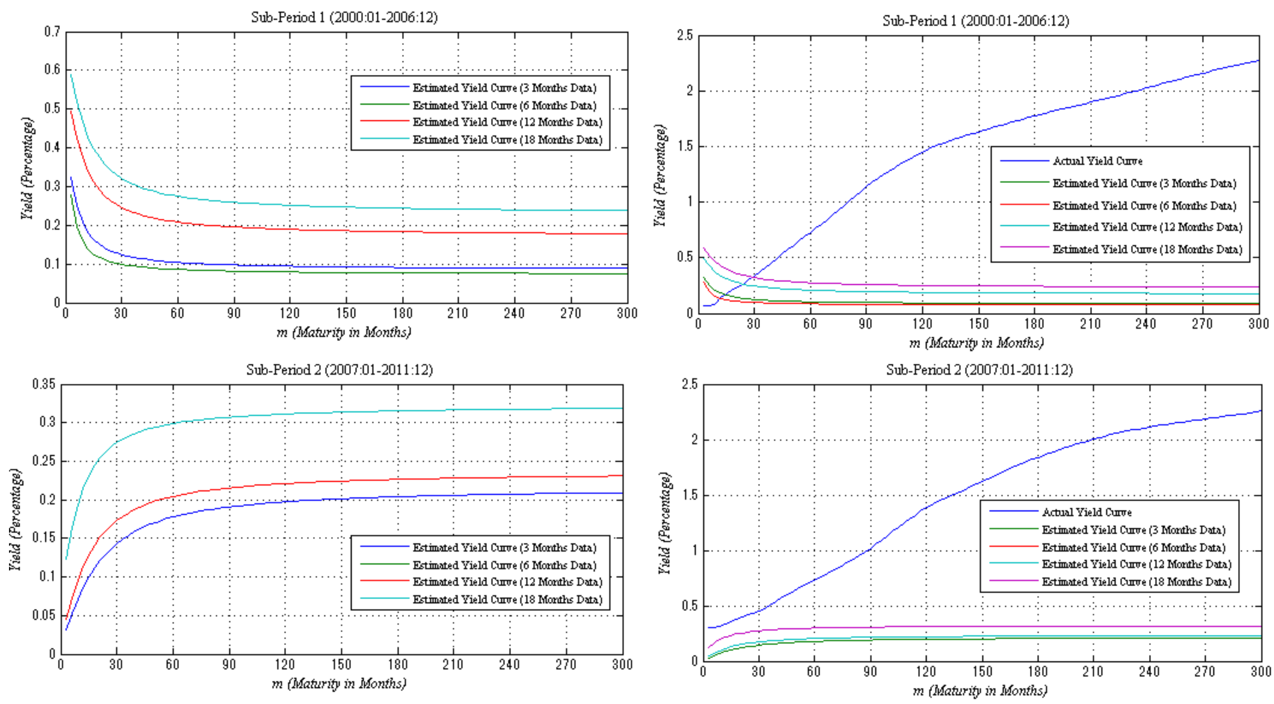

Figure A-2: Fitted Yield Curve with the CIR Model for Two Sub-Periods
Actual average (data-based) and fitted (model-based) yield curves for two sub-periods, i.e., sub-period 1 (2000:01 –2006:12) and sub-period 2 (2007:01 – 2011:12) using the time series data of 3 months, 6 months, 12 months, and 18 months maturities are plotted. The fitted yield curves are obtained by evaluating the CIR function at the MLE estimated $\hat{\kappa}, \hat{\mu}$ and $\hat{\sigma}$ from table A-2.