Geometric information of yield curve, unspanned stochastic volatility, and affine Heath-Jarrow-Morton models

Qingbin Wang
University at Albany, State University of New York, dylen1@yahoo.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Economics Commons, and the Finance and Financial Management Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1303

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
GEOMETRIC INFORMATION OF YIELD CURVE,
UNSPANNED STOCHASTIC VOLATILITY,
AND AFFINE HEATH-JARROW-MORTON MODELS

by

Qingbin Wang

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Economics
July 2014
GEOMETRIC INFORMATION OF YIELD CURVE, UNSPANNED STOCHASTIC VOLATILITY, AND AFFINE HEATH-JARROW-MORTON MODELS

by

Qingbin Wang

COPYRIGHT 2014
To my family
and
in memory of my father
Acknowledgment

I am indebted to a great many people who have contributed to this dissertation in different ways. Without their supports and encouragements, it will not only be harder but also take longer to write and finish this dissertation.

My deepest gratitudes go to my advisor Professor Bruce Dieffenbach. His immense knowledge, caring, and patience have supported me for all these years of study and research. His guidance and inspiring comments helped me through all the time of writing this paper. I could not imagine how this dissertation can be finished without him.

I also want to extend my sincere thanks to my committee members, Professor John Jones and Professor Terrence Kinal, for their valuable suggestions and questions. It has been a great honor to have them on my committee.

I am grateful to the faculty of the Department of Economics, University at Albany, for their teaching and mentoring over my graduate studies, which I will always cherish in my future.

I owe hugely to my parents, Xianying Pu and Zhaohu Wang. Their unconditional love and unselfish dedication to me are beyond words and have encouraged me to move forward in my life. I also want to extend my gratitudes and appreciations to my sisters, Hongxia Wang and Ping Wang, and my brother-in-law, Xingfeng Niu, for them taking good care of my mother after my father passed away. I dedicate this dissertation to them.

This dissertation is also dedicated to my unborn baby girl for her bringing new meanings and joys to my life. She will always be the softest point of my heart.

Last, but not least, I express my gratitudes to my dear wife, Yunshu Li, for her love, support, and encouragement, which in the end make this dissertation possible. I dedicate this dissertation to her too.
Abstract

The differences between the daily routine of fitting the yield curve (or equivalently, the forward rate curve) and the dynamic affine term structure modeling practice serve their own purposes even though they both work on the same object, the term structure of interest rates. Curve fitting models use parameterizations of the yield curve with parameters labeled as level, slope, and curvature depending on their parameter loadings. They are static and descriptive, while the dynamic affine term structure models are dynamic and theoretically consistent, or arbitrage-free. Recently, attempts to use the geometric information of curve fitting models to improve the empirical performance of affine term structure models have been carried out. This paper contributes to this line of research by joining together the curve fitting models and dynamic affine models.

Unlike most affine short rate models, we develop multifactor affine Heath-Jarrow-Morton (HJM) forward rate models with general forms of volatility. This makes those attempts with short rate models special cases of our models. The factors in our models have clear geometric meanings such as level, slope, and curvature. Our model framework is of great flexibility and can be easily extended from Gaussian to various forms of stochastic volatility, either spanned or unspanned. Depending on the forward rate volatility, the factors in our model can form either composite (correlated) or independent level, slope, and curvature of the forward rate curve.

The paper also proposes a consistent Augmented Nelson-Siegel (ANS) model, which is more parsimonious and has three factors represent independent level, slope, and curvature of the forward rate curve. Various model specifications are implemented in two empirical studies. One uses U.S. Treasury yields and the other uses LIBOR and swap rates.

We also suggest a statistic called information utilization, which is used to study how investors respond to market situations. The in-depth examination of the extracted factors and their geometric meanings shows that the movement patterns of the yield factors behave differently after the 2008 financial crisis, a clear indication of structural change. The analysis of information utilization also shows its potential to identify business cycle.
## Contents

1 Introduction  
   1

2 Heath, Jarrow and Morton (1992) Framework  
   10
   2.1 Interest rate model basics ........................................ 10
   2.2 Nelson-Siegel Model and Svensson Model ......................... 12
   2.3 Heath, Jarrow and Morton (1992) Framework .................... 13

3 Bjork and Christensen (1999) Consistent Conditions  
   16

4 Consistent Conditions for Affine Parameterization  
   18

5 Affine HJM Models with Stochastic Volatility  
   20
   5.1 Affine HJM models with Gaussian Volatility .................. 21
   5.2 Affine HJM Models with Stochastic Volatility ............... 31
   5.3 Consistent Augmented Nelson-Siegel Model .................... 42

6 Unspannedness of Risk Factors: a Further Note  
   51

7 Pricing Zero-coupon Bonds  
   52
   7.1 Bond Price in n-factor HJM Model with USV .................. 53
   7.2 Bond Price in Consistent Augmented Nelson-Siegel Model with USV .................. 55

8 Market Price of Risk  
   56
   8.1 Completely Affine Model ........................................ 57
   8.2 Essentially Affine Model ....................................... 57
   8.3 Semi-affine Model .............................................. 58
   8.4 Extended Affine Model ....................................... 58
   8.5 Market Price of Risk for Our Models .......................... 60

9 Empirical Analysis  
   62
   9.1 Discretization of State Process ................................. 63
   9.2 Extended Kalman Filter ....................................... 65
   9.3 Application to U.S. Treasury Yields ............................ 68
   9.4 Application on LIBOR and Swap Rates .......................... 80
1 Introduction

There is a discrepancy between the daily routine of fitting the yield curve \(^1\) and the affine term structure modeling practice. Although they both work on the same object, the term structure of interest rates, curve fitting models are static and descriptive, while affine term structure models are dynamic and theoretically consistent, or arbitrage-free. Recently, attempts to use the geometric information of curve fitting models to improve the empirical performance of affine term structure models have been carried out by researchers, such as Diebold and Li (2006) and Christensen, Diebold, and Rudebush (2010). This project contributes to this line of research.

Unlike them, we develop multifactor affine Heath-Jarrow-Morton models where the factors \(^2\) have clear geometric meanings, such as the level, slope and curvature of the forward rate curve. The model framework in the paper is of great flexibility and can easily be extended from Gaussian to various types of stochastic volatility, either spanned or unspanned. Based on the general model framework, we also study a consistent Augmented Nelson-Siegel (ANS) model with three factors being the level, slope, and curvature of the forward rate curve. The paper limits itself to the unspanned stochastic volatility, which means the interest rate risk cannot be hedged with positions only containing bonds. Unspanned stochastic volatility has caught attentions in research like Collin-Dufresne and Goldstein (2002), Heidari and Wu (2009), and Trolle and Schwartz (2009) to name a few. The models developed in this paper are theoretically consistent, which means arbitrage opportunity has been ruled out, and also in affine forms of the state variables. Therefore, they can be used to price bonds and interest rate derivatives directly and also offer closed-form solutions for European-type interest rate derivatives.

Since the seminal work of Black and Scholes (1972), last four decades have seen tremendous development in modern finance. Most of dynamic asset pricing models, term structure of interest rates models included, are affine models. The popularity of affine models are largely due to the mathematical convenience, such as analytical tractability and the existence of closed-form solutions for European-type interest rate derivatives.

\(^1\) As shown in Section 2.1, the information contained in the yield curve is equivalent to that contained in the forward rate curve. Knowing one means knowing the other. Thus, we use them interchangeably throughout this paper.

\(^2\) In most affine term structure models, the factors are the same as the state variables. In this paper, they are different. Please refer to Definition 5.1 and 5.2 for details.
Models of the term structure of interest rates can be categorized into two groups: equilibrium models and no-arbitrage models. Equilibrium models don’t fit the initial term structure and focus on the time evolution of the term structure. Most equilibrium models start with specifying the short rate as an affine function of one or more state variables. Together with equilibrium condition, bond price is derived as an exponential affine function of those state variables. Popular short rate models include Merton (1973), Vasicek (1977), Cox, Ingersoll, and Ross (1985a) (CIR), and Duffie and Kan (1996). The first three models have the short rate as the only state variable with Merton (1973) and Vasicek (1977) having constant diffusion term, or Gaussian volatility. CIR model, however, specifies the short rate as a square-root process with the square-root of the short rate being the diffusion term. CIR model overcomes the shortcoming of Merton (1973) and Vasicek (1977), where the short rate can become negative. Duffie and Kan (1996) extended the CIR model to a multifactor case. They specified the short rate as an affine function of multiple factors.

Because equilibrium models don’t fit the initial term structure perfectly, they are not suitable to price interest rate derivatives. On the other hand, no-arbitrage models fit the initial term structure as given information. The first generation of no-arbitrage models are just the equilibrium short rate models with time inhomogeneous parameters in order to perfectly fit the initial term structure. For instance, Hull and White (1990) extended Vasicek (1977) by making the mean of the short rate time-inhomogeneous.

Heath, Jarrow, and Morton (1992) (HJM) took a leap forward by directly modeling the whole forward rate curve. HJM took the whole forward rate curve as a state variable. Under the risk-neutral probability measure, the no arbitrage constraint in the HJM model implies that the drift is completely determined by the forward rate volatility. By construction, the initial forward rate curve is fitted perfectly. As the short rate is the short end of the whole forward rate curve, all short rate arbitrage-free models are just special cases of the HJM forward rate model, at least theoretically. Heath, Jarrow, and Morton (1992) is more of a model framework than just a single model as it gives the condition that all interest rate models should satisfy. This theoretical generality gains its popularity in theoretical and practical research. Therefore, we also construct our models within the HJM model framework.

Affine term structure models such as Duffie and Kan (1996) have been a dominant
class in both theoretical and empirical analysis. These models specify the short rate as an affine function of some state variables and obtain the bond price as an exponential affine function of those state variables. However, the advantage of the analytical tractability of affine models comes at the cost of their poor flexibility to explain empirical observations and their unsatisfactory forecast performance as documented in Duffee (2002). Besides, due to the potential over-parameterization, the likelihood function is usually very flat which makes the estimation and interpretation more difficult. As a solution, Duffee (2002) and Dai and Singleton (2002) simply set some parameters as zero. The resulting more parsimonious models typically improve on the empirical analysis. However, these restrictions are just imposed subjectively without interpretations.

To improve the empirical performance of affine term structure models, some research utilizes the parameterizations in curve fitting models. The term structure of interest rates or the yield curve are not directly observed in the market. Instead, some smooth curves, or parameterizations, are fitted to the discretely observed bond prices by minimizing the fitting errors. Among them, Nelson and Siegel (1987), Svensson (1994), and cubic splines are widely used by central banks and institutional investors. The parameters in the Nelson and Siegel (1987) model and Svensson (1994) model are called the level, slope, and curvature of the forward curve according to their coefficients in the model. When these parameters are viewed in a time series manner, they are called factors or state variables and their coefficients are called factor loadings. The geometric information of the yield curve is the information embedded in the shape of the yield curve, which is carried by the three factors named above. The shape of the yield curve depends on the yield formula which is a function of the factors. It means that, to obtain a certain shape of the yield curve, some constraints need to be imposed on an interest rate model.

Since the curve fitting process is performed on a daily basis, curve fitting models are static and descriptive. When the time series evolution of the yield curve is being studied, these models create a discrepancy between the static parameterizations and the dynamic interest rate models. Filipovic (1999) and Bjork and Christensen (1999) showed that there is no Ito’s process that is consistent with either the Nelson and Siegel (1987) model or the Svensson (1994) model. It means that if we want to use the information in the geometric shape of the Nelson and Siegel (1987) model and the Svensson (1994) model, a certain augmentation is required to make them consistent with some interest rate models.
Krippner (2006), Diebold and Li (2006), and Christensen, Diebold, and Rudebush (2010) added dynamics to the parameters to build dynamic Nelson-Siegel models. After the dynamization, the parameters are usually called state variables. Krippner (2006) started with a three-factor HJM model by specifying the forward rate volatility and also made assumptions on the expected short rate. With the relationship between the forward rate and the short rate, Krippner (2006) cast the forward rate formula into the Nelson-Siegel format. As Bjork and Christensen (1999) and Filipovic (1999) proved that no Ito’s process is consistent with the Nelson-Siegel model, unsurprisingly, Krippner (2006) model has an extra term added to the original Nelson-Siegel model. He called it augmentation term. A similar term also appears in Christensen, Diebold, and Rudebush (2010). Strictly speaking, Krippner (2006) does not belong to the HJM model framework because he made an additional assumption on the expected short rate. In the HJM framework, the short rate is also determined by the volatility structure under the risk-neutral probability measure. Therefore, Krippner (2006) is more like a short rate model with a specification of forward rate volatility. But it is theoretically consistent because it is derived under the HJM framework and, thus, the arbitrage opportunity is ruled out.

Diebold and Li (2006) suggested their dynamic Nelson-Siegel model by imposing an arbitrary \( AR(1) \) process on the parameters of the original Nelson-Siegel model. Their model has no theoretical consistency, or could not rule out the arbitrage opportunity. The authors explained that, since market has eliminated the arbitrage opportunity, the yield curve observed is already arbitrage free. Christensen, Diebold, and Rudebush (2010) filled the theoretical gap in Diebold and Li (2006). They constructed a short rate model in the affine form of Duffie and Kan (1996) and used the result that the bond price is an exponential affine function of the state variables. Thus, the yield becomes an affine function of the state variables. They rearranged the yield function and cast it into a form called generalized Nelson-Siegel model, which includes the form of the original Nelson-Siegel model and an additional term to rule out of the arbitrage opportunity. Since their bond price is derived under the affine model framework, the model is theoretically consistent.

Along the line of Diebold and Li (2006), de Pooter, Ravazzolo, and Dijk (2010) added macro-factors into the dynamic Nelson-Siegel model. Pooter (2007) compared the performance of different versions of the Nelson-Siegel family model and suggested
that consistent Nelson-Siegel and Svensson family models, in the sense of Bjork and Christensen (1999), perform better than the inconsistent models. But the consistent Nelson-Siegel models in Pooter (2007) are only consistent with one-factor models even though the author still treated the parameters as multiple state variables. All models in this line of research have Gaussian volatility except Koopman, Mallee, and Wel (2010), which introduced a GARCH type stochastic volatility.

As being said, most models in those papers, except Christensen, Diebold, and Rudebush (2010), are theoretically inconsistent. They all work on the bond market only. Due to the lack of theoretical consistency, they could not be used directly to price interest rate derivatives. Just like the parameters in the original Nelson-Siegel model, the state variables in Christensen, Diebold, and Rudebush (2010) have the geometric meanings of level, slope, and curvature of the yield curve. Therefore, the adoption of modified Nelson-Siegel model is like using the geometric information of the yield curve. The inclusion of this "additional" geometric information enables those models outperform the traditional affine term structure models.

This is exactly why we choose to build our models on the basis of the consistent parameterizations of forward rate curve. On one hand, we can gain a better empirical performance by imposing consistent parameterizations. Meanwhile, we can maintain the yield as an affine function of the state variables (parameters after dynamization). Our consistent parameterizations include a very general form and an Augmented Nelson-Siegel form. Unlike Christensen, Diebold, and Rudebush (2010), which followed the short rate model framework, we follow the Heath, Jarrow, and Morton (1992) forward rate model framework. Since we focus on affine models, we first simplify the consistent conditions in Bjork and Christensen (1999) to limit ourselves within the affine consistent parameterizations. Then, with the consistent conditions for affine parameterizations, we derive the dynamics of state variables.

In the HJM framework, the forward rate curve is well-known to be infinite dimensional. If a parameterization form is found to be consistent with a forward rate model, the infinite dimensional forward rate curve then can be represented with a finite dimensional state space system. The consistent conditions in Bjork and Christensen (1999), including a drift condition and a volatility condition, specify what the drift and volatility terms of the state variables (parameters in dynamic setting) should satisfy when a parameterization
is consistent with an interest rate model. The geometric meaning of the consistent conditions tell that, for a parameterization of forward curve and an interest rate model to be consistent, the forward rate model's drift and volatility terms should be in the tangential manifold generated by the manifold of the parameterized forward curve. Now we have two consistent dynamic systems, a forward rate model and a dynamized parameterization of forward rate curve which is an affine function of some state variables (dynamized parameters). The forward rate model now can be expressed as an affine function of some state variables and, thus, the parameterization of the forward curve becomes dynamic.

Chiarella and Kwon (2003) proposed another way to transform the HJM model into a finite dimensional state space system. They rearranged the stochastic differential equation followed by the forward rate and regrouped the terms according to the state variables defined. This method can be called the "forward method". Our method is a kind of "backward method" as we choose some consistent parameterization in advance and then derive the dynamics of the state variables using the consistent conditions. In Chiarella and Kwon (2003), the state variables do not have any geometric meanings. The advantage of our method is that the factors in our models have geometric meanings based on the forward rate parameterization proposed.

As the parameters in the original Nelson-Siegel model can be interpreted as the level, slope and curvature of the forward rate curve, factors in our multifactor general volatility models can be combined to get the geometric meanings. That is why they are called "composite" geometric meanings. On the other hand, in our Augmented Nelson-Siegel model, the factors have the meanings of independent level, slope, and curvature of the forward rate curve.

Trolle and Schwartz (2009) followed Chiarella and Kwon (2003) and their factors, accordingly, do not have any geometric meaning. As shown later, our models have a more general form of forward rate volatility than Trolle and Schwartz (2009) and the factors in our model have clear geometric meanings. Even with the forward rate volatility specified in Trolle and Schwartz (2009), it can be shown that the model constructed with our method have fewer state variables than Trolle and Schwartz (2009) besides the factor geometric meanings. Those added geometric meanings of factors not only improve the empirical performance of our models but also bring convenience to deeper factor analysis. More importantly, Trolle and Schwartz (2009) claimed their three state variables with
contemporaneous randomness as the level, slope and curvature of the forward rate curve. Their specification, however, can only generate the slope and curvature components and could not generate the level component. All differences mentioned above make Trolle and Schwartz (2009) model a special case of our general volatility model.

The paper contributes to the existent literature in the following aspects. First, our model framework links together two modeling practices, the curve fitting practice and affine term structure modeling practice. This feature enables us to incorporate the geometric information in curve fitting models into the construction of affine models. The geometric information of the forward rate curve imposes extra constraints on the coefficients in our models. Compared with curve fitting models, such as Nelson-Siegel models or Svensson models, our models are consistent in a sense that they rule out the arbitrage opportunity. This property enables our models to price interest rate derivatives directly. On the other hand, compared with ordinary affine term structure models, our models take into consideration the geometric information gained from curve fitting models. Including the "additional" geometric information is just like imposing extra constraints on the coefficients in ordinary affine term structure models. As Christensen, Diebold, and Rudebush (2010) suggested, adding those constraints improves the empirical performance of affine term structure models.

Secondly, Christensen, Diebold, and Rudebush (2010) and most other affine term structure models commonly start with specifying the short rate as an affine function of some unobservable state variables. Unlike them, we develop HJM forward rate models by specifying various forms of forward rate volatility, including a very general form and some special forms. The proposed affine HJM model can jointly price bonds and interest rate derivatives. The volatility structure defined in our models can accommodate level, slope and humped shape. This volatility form, in turn, helps us obtain the state variables that can be explained as the level, slope and curvature of the forward rate (equivalently, the yield curve) formula. Just like that short rate model can be interpreted as a special case of the forward rate model, Christensen, Diebold, and Rudebush (2010) and Krippner (2006) can also be interpreted as special cases of our model.

Thirdly, we propose a general, yet flexible, model framework with stochastic volatility. While maintaining the bond yield formula unchanged, or without affecting the shape of the forward rate curve, the model can be easily extended from Gaussian volatility to
various forms of stochastic volatility, either spanned or unspanned, or even something like unspanned macro-risk factors in Joslin, Priebsch, and Singleton (2014). Specifically, under the unspanned stochastic volatility, the volatility variables do not enter the bond yield formula. This characteristic enables us to incorporate observable variables, such as macroeconomic variables and even news announcements into the specification of volatility dynamics. In this way, the model can be used to explain how the term structure can be affected by the observable variables selected without having those variables in the bond yield formula.

Fourthly, since our model is theoretically consistent, it is ready to price interest rate derivatives. Our model specification is in a more general form than Trolle and Schwartz (2009). The volatility specification in Trolle and Schwartz (2009) could not generate the level component of the forward rate curve even though they claim one of the state variables represents the level factor. Moreover, the state variables in our model have more clear geometric meanings than those in Trolle and Schwartz (2009).

Fifthly, based on our general model framework, we derive a three-factor consistent Augmented Nelson-Siegel model as a special case. Compared with other no-arbitrage dynamic Nelson-Siegel models, our Augmented Nelson-Siegel model has different interpretation of the augmented term, which is completely dynamic. Christensen, Diebold, and Rudebush (2010) had an augmented term integrated into an explicit form due to its Gaussian volatility. This means the augmented term in Christensen, Diebold, and Rudebush (2010) does not change over time. The augmented term in our model, however, is expressed as the sum of auxiliary state variables, which are dynamic but don’t have contemporaneous randomness. This augmented Nelson-Siegel model is also ready to price interest rate derivatives.

In addition, when the unspanned stochastic volatility is introduced, the volatility variable in our models happen to be a component of the long-run mean of the state variables with contemporaneous randomness. This property clearly differs from other models with unspanned stochastic volatility, such as Casassus, Collin-Dufresne, and Goldstein (2005) and Trolle and Schwartz (2009). Casassus, Collin-Dufresne, and Goldstein (2005) had the long-run mean of the short rate as another state variable but without contemporaneous randomness, which is similar to our auxiliary state variables. Trolle and Schwartz (2009) had state variables with zero long-run mean.
Intuitively, with the long-run mean of the state variables proportionate to the volatility variables, our models indicate that the volatility state variables affect the bond yield indirectly through their effects on other state variables. Of course, this relation is in terms of the long-run mean of other state variables, which form the yield formula. Thus, the propagation mechanism embedded in our model framework goes from the factors that affect the volatility variables to the volatility variables and then to the bond yield through the state variables in the bond yield formula.

It becomes more interesting when the volatility state variables incorporate variables like macro-economic variables, news announcement, and other volatility proxies. The property mentioned above stands out our models from other macro-finance term structure models, which have macro-economic variables directly enter into the bond yield formula, such as Ang and Piazzesi (2003) and Rudebusch and Wu (2008). It also implicitly shows that the volatility of the yield is related to the level of the yield.

The rest of the paper proceeds as follows. Section 2 introduces the Heath, Jarrow, and Morton (1992) model framework. Section 3 explains the consistent conditions in Bjork and Christensen (1999). Section 4 obtains the consistent conditions for affine models. Section 5 derives various affine HJM models with Gaussian and stochastic volatilities. Multifactor HJM models with a general form of volatility and the Augmented Nelson-Siegel model are also derived in this section. Section 6 further discusses the unspannedness of risk factors. Section 7 derives the bond prices for various models obtained in Section 5. Various forms of market price of risk are discussed in Section 8. In Section 9, two empirical studies are carried out to test the models we derived in the paper. Section 10 concludes the paper and briefly introduces some possible extensions.

**Remark 1.1 Asset Pricing Trinity.** Before getting into details of various models, it would be helpful to explain how our models are proposed since it is different from how most term structure models are constructed. In asset pricing, especially in interest rate modeling, closely related are three components, which I call the trinity of asset pricing with each component on one vertex. The three components include a microscopic object, some underlying factors, and a pricing formula of an asset. The microscopic object determines what kind of modeling framework is in use and the underlying factors (state variables) are the driving force of the system, while the pricing formula of an asset is the target of interest. In the middle of the trinity is the no-arbitrage condition which unifies the three
components together.

Take the short rate model for an instance, the microscopic object is the short rate which is usually specified as an affine function of the underlying factors. As the driving force of the dynamic system, the underlying factors, or state variables, are specified as well. Then, by imposing the no-arbitrage constraint, the pricing formula of an asset is derived as the result. This is the usual way of building an asset pricing model. With no-arbitrage condition in the center, the three components are united together as a complete entity. Theoretically, if we know two of the three components, the third one can be derived. Unlike the usual way, in order to utilized the geometric information of the yield curve, we start with specifying the microscopic object, or the forward rate volatility. Then, we propose the pricing formula of the asset, or the forward rate formula. Finally, the state variable dynamics is derived with the imposition of no-arbitrage condition.

2 Heath, Jarrow and Morton (1992) Framework

This section first introduces the basics of interest rate models and then presents the Heath, Jarrow, and Morton (1992) framework.

2.1 Interest rate model basics

A probability space is defined as \((\Omega, \mathcal{F}, \mathbf{P})\) with a state space \(\Omega\), a \(\sigma\)-field \(\mathcal{F}\), and a probability measure \(\mathbf{P}\) that maps \(\mathcal{F}\) into \([0, 1]\). A filtration is an increasing sequence of \(\sigma\)-fields, for \(s \leq t\), \(\mathcal{F}_s \subseteq \mathcal{F}_t\), which describes the evolution of information sets. No arbitrage implies the existence of an equivalent martingale measure \(Q\). When a money market account is used as the discount factor, this equivalent martingale measure is also called the risk-neutral probability measure. The models discussed in the following sections are under the risk-neutral probability measure \(Q\) unless stated otherwise.

The term structure of interest rates or the yield curve can be described in different but equivalent ways.

1. Zero-coupon bond price \(P(t, T)\). A zero-coupon bond, also known as discount bond, is a financial contract that guarantees a payoff of one unit at its maturity time. A zero-coupon bond which matures at time \(T > 0\) is usually called a \(T\)-bond. The price of a zero-coupon bond is the time \(t\) value of one unit at maturity time \(T\),
which is written as $P(t, T)$. Clearly, we have $P(T, T) = 1$.

2. Yield to maturity $y(t, T)$. The yield $y(t, T)$ on a zero-coupon bond is the effective compounded interest paid on the bond held until maturity.

3. Instantaneous forward rates $f(t, T)$. The instantaneous forward rate is the time $t$ cost of an instantaneous borrowing at a future time $T$, or the cost at time $t$ of borrowing over an infinitesimal time interval $[T, T + dt]$. In a similar sense to the yield, the instantaneous forward rate $f(t, T)$ is a “constant interest rate” over time interval $[T, T + dt]$ determined at time $t$.

4. Short rate $r(t)$. The short rate, $r(t)$, is the return rate on an instantaneous lending at time $t$.

By definitions above, we have the following identities.

$$y(t, T) = -\frac{\log P(t, T)}{T - t}.$$  \hspace{1cm} (2.1)

and

$$f(t, T) = -\frac{\partial \log P(t, T)}{\partial T}.$$  \hspace{1cm} (2.2)

i.e.

$$P(t, T) = \exp \left( -\int_t^T f(t, s) ds \right).$$  \hspace{1cm} (2.3)

$$y(t, T) = \frac{1}{T - t} \int_t^T f(t, s) ds$$  \hspace{1cm} (2.4)

and

$$r(t) = f(t, t).$$  \hspace{1cm} (2.5)

These relations are not pricing equations but identities obtained from their definitions. It means that these equations cannot be used to price zero-coupon bonds. Instead, they can facilitate the analysis when different values, such as the yields, forward rates, and bond price, are required. These identities also indicate that, once we know any value of them, other values can be derived equivalently. For this reason, the yield curve, forward rate curve, bond prices are used interchangeably in this paper.
2.2 Nelson-Siegel Model and Svensson Model

To fit the forward rate curve on a daily basis, different parameterizations have been developed. Bank of International Settlement documents various parameterizations used by the central banks around the world (BIS (2005)). Two popular ones include the Nelson-Siegel model (Nelson and Siegel (1987)) and Svensson model (Svensson (1994)).

Nelson and Siegel (1987) proposed a parsimonious parameterization for the forward rate curve as

\[
f(t, T) = \beta_1 + \beta_2 e^{-a(T-t)} + \beta_3(T - t)e^{-a(T-t)}, \tag{2.6}
\]

or, in zero-coupon bond yield form,

\[
y(t, T) = \beta_1 + \beta_2 \left( 1 - e^{-a(T-t)} \right) + \beta_3 \left( \frac{1 - e^{-a(T-t)}}{a(T-t)} - e^{-a(T-t)} \right), \tag{2.7}
\]

where \(\beta_1, \beta_2, \beta_3,\) and \(a\) are parameters to be estimated daily to fit the market observed yields. As the parameters are estimated every day, the model is actually a static model. However, when the parameters, \(\beta_1, \beta_2,\) and \(\beta_3,\) are viewed in a time series manner, they can be interpreted as factors or state variables, which drive the forward rate curve. Thus, their coefficients, \(1, e^{-a(T-t)},\) and \((T - t)e^{-a(T-t)},\) can be interpreted as factor loadings. These factor loadings are the functions of level, downward slope and curvature in terms of time-to-maturity \((T - t).\) Figure-1 shows what these factor loadings look like. Based on the interpretation of factor loadings, \(\beta_1, \beta_2,\) and \(\beta_3\) are labeled as the level, slope and curvature factors of the forward rate curve, or the yield curve. The Nelson-Siegel model in equation (2.6) has one level, slope, and curvature. Similarly, Svensson (1994) suggested another parameterization by adding a second curvature factor as

\[
f(t, T) = \beta_1 + \left( \beta_2 + \beta_3(T - t)e^{-a_1(T-t)} \right) + \beta_4(T - t)e^{-a_2(T-t)}, \tag{2.8}
\]

where, \(\beta_4\) is the second curvature factor. The same interpretation of the factors, factor loadings, and geometric information can be applied to the Svensson model as to Nelson-Siegel model. In a similar way, Svensson (1994) model can be written as zero-coupon bond yields just like that for Nelson and Siegel (1987) model in equation (2.7).

Due to their static implementation, neither the Nelson-Siegel nor the Svensson model is consistent with any interest rate model in a sense of ruling out arbitrage opportunity.
Studies have shown that no Ito’s process can generate a forward rate process as either
the Nelson-Siegel model in equation (2.6) or the Svensson model in (2.8). Please refer to
Filipovic (1999) and Bjork and Christensen (1999) for details. Therefore, to utilize the
geometric information buried in the forward rate curve and obtain a consistent interest
rate model, we have to augment the original Nelson-Siegel model or Svensson model by
adding some extra terms. These extra terms will help rule out the arbitrage opportunity
while we maintain the geometric meanings of the factors. Later in this paper, we propose
various specifications based on one general model framework to accommodate composite
level, slope and curvature factors and independent level, slope and curvature factors.
In this way, we can give the factors geometric meanings, which is important for factor
analysis, and also can decompose the forward rate curve to examine how each factor
behaves over time.

2.3 Heath, Jarrow and Morton (1992) Framework

The Heath, Jarrow, and Morton (1992) approach is a major breakthrough in pricing
fixed-income products. They proposed a model framework that encompasses all the
important models in the past, such as Vasicek (1977), Ho and Lee (1986), and Hull
and White (1990). There are two key inputs in the Heath, Jarrow, and Morton (1992)
framework: the initial forward curve and the forward rate volatility. Specifically, the
evolution of the entire forward rate curve is defined as a multidimensional diffusion process
under some probability measure. Here we choose the risk-neutral probability measure $Q$:

$$f(t, T) = f(0, T) + \int_0^t \mu(s, T)ds + \sum_{i=1}^n \int_0^t \sigma_i(s, T)dW_i(s), \quad (2.9)$$

or, in differential form:

$$df(t, T) = \mu(t, T)dt + \sum_{i=1}^n \sigma_i(t, T)dW_i(t). \quad (2.10)$$

where $f(0, T)$ is the initial forward rate curve, $W = (W_1, W_2, ... W_n)$ is a $N$-dimensional
vector of standard Brownian motions.

By construction, the initial forward curve is perfectly fitted. More importantly, under
the risk-neutral probability measure $Q$, the no-arbitrage condition implies that the drift
term \( \mu(t, T) \) is completely determined by the volatility \( \sigma_i(t, T) \) with:

\[
\mu(t, T) = \sum_{i=1}^{n} \sigma_i(t, T) \int_t^T \sigma_i(t, u) du,
\]

or, equivalently,

\[
\int_t^T \mu(t, s) ds = \frac{1}{2} \sum_{i=1}^{n} \left( \int_t^T \sigma_i(t, u) du \right)^2.
\]

Then,

\[
f(t, T) = f(0, T) + \sum_{i=1}^{n} \int_0^t \sigma_i(s, T) \left( \int_s^T \sigma_i(s, u) du \right) ds + \sum_{i=1}^{n} \int_0^t \sigma_i(s, T) dW_i(s).
\]

(2.13)

With the identities between \( P(t, T), f(t, T) \) and \( r(t) \), by using Ito’s lemma, the instantaneous short rate becomes:

\[
r(t) = f(t, t)
\]

(2.14)

\[
= f(0, t) + \sum_{i=1}^{n} \int_0^t \sigma_i(s, t) \left( \int_s^t \sigma_i(s, u) du \right) ds + \sum_{i=1}^{n} \int_0^t \sigma_i(s, t) dW_i(s).
\]

With equation (2.12), applying Ito’s lemma to equation (2.14) yields the short rate dynamics

\[
dr(t) = df(t, T)|_{T=t} + \frac{\partial f(t, T)}{\partial T}|_{T=t} dt
\]

\[
= \left[ \frac{\partial f(0, t)}{\partial t} + \sum_{i=1}^{n} \left( \int_0^t \frac{\partial \sigma_i(t, s)}{\partial t} \left( \int_s^t \sigma_i(s, u) du \right) ds + \int_0^t \sigma_i^2(s, t) ds + \int_0^t \frac{\partial \sigma_i(t, s)}{\partial t} dW_i(s) \right) \right] dt
\]

\[
+ \sum_{i=1}^{n} \sigma_i(t, t) dW_i(t),
\]

(2.15)

Equation (2.15) shows that, for an HJM model defined in equation (2.10), an equivalent short rate model defined in equation (2.15) can be found. This one-to-one mapping reveals that the HJM model is more of a model framework such that a short rate model can be derived from it.

Equation (2.14) indicates that the short rate \( r(t) \) is determined by the volatility structure and the initial forward rate with maturity \( t \). Generally speaking, the short rate in (2.14) or in (2.15) is not a Markovian process due to the last integral in equation (2.15). This non-Markov property makes the short rate path-dependent. Markovian
transformation of the HJM model has been studied by Carverhill (1994), Ritchken and Sankarasubramanian (1995), Inui and Kijima (1998), and Chiarella and Kwon (2001). They suggested a separable forward rate volatility as:

$$
\sigma_i(t, T) = \xi_i(t) \eta_i(T),
$$

(2.16)

where $\xi_i(t)$ and $\eta_i(T)$ are strictly positive and deterministic functions of time.

Equation (2.13) shows that, with standard Brownian motions, a multifactor HJM model can be viewed as the sum of one-factor HJM models generated by each associated volatility term. So, from now on we will focus on a one-factor HJM for illustration purposes, since it is straightforward to extend it to a multifactor model.

With the identity between the bond price and forward rate, the bond price dynamics becomes

$$
\frac{dP(t, T)}{P(t, T)} = r(t)dt + \sum_{i=1}^{n} \left( - \int_{t}^{T} \sigma_i(t, u)du \right) dW_i(t)
$$

(2.17)

Equation (2.17) can be used in deriving the bond price after we obtain the affine realization of the HJM model. It shows that the volatility of the relative bond price change is the cumulative sum of the volatility of forward rates spanned by the bond maturity. This relation can also be used in approximating the swaption price with options on zero-coupon bonds when the stochastic duration of the coupon bonds is calculated.

In practical applications, it is more convenient to use the time-to-maturity, $x = T - t$, rather than the time-of-maturity, $T$, in the formula of forward rate or bond yield such that we can have the same maturity-spectrum of forward rates at each time. Based on Brace and Musiela (1994),

$$
r(t, x) = f(t, t + x).
$$

By applying Ito’s lemma, the dynamics of forward rate becomes

$$
\frac{dr(t, x)}{dt} = \left[ \frac{\partial r(t, x)}{\partial x} + \sigma_0(t, x) \int_{0}^{T} \sigma_0(t, u)du \right] dt + \sigma_0(t, x) dW(t),
$$

(2.18)

where,

$$
\sigma_0(t, x) = \sigma(t, t + x).
$$

3Forward rate $r(t, x)$ is double-indexed with $t$ for time and $x$ for time-to-maturity. It is different from the short rate $r(t)$, which only has one index.
It can be rewritten as:

\[
dr(t, x) = \mu_0(t, x)dt + \sigma_0(t, x)dW(t),
\]

where,

\[
\mu_0(t, x) = \frac{\partial r(t, x)}{\partial x} + \sigma_0(t, x) \int_0^x \sigma_0(t, u)du.
\]

(2.19)

(2.20)

3 Bjork and Christensen (1999) Consistent Conditions

Bjork and Christensen (1999) studied the consistent conditions between an interest rate model (dynamic), \( \mathcal{M} \), and a family of forward rate curves, \( \mathcal{G} \), generated by a parameterized smooth function (static). With those consistent conditions, a static parameterization can be dynamized into a dynamic interest rate model. Some of their results are highlighted in this section.

A pair \( \{\mathcal{M}, \mathcal{G}\} \) consists of a dynamic interest rate model \( \mathcal{M} \) and a family of static parameterized forward curves \( \mathcal{G} \). A natural question is: can future forward rate process generated by the interest rate model, \( \mathcal{M} \), fall in \( \mathcal{G} \), given that the initial forward curve is in \( \mathcal{G} \)? Bjork and Christensen (1999) and Filipovic (1999) proved that no Ito’s process is consistent with either Nelson-Siegel or Svensson model. It means that no interest rate model can produce forward rate processes that fall in the family of forward rate curves formed by the original Nelson-Siegel or Svensson model. That is to say, no interest rate model is consistent with either of those two original models.

Under the HJM framework, an interest rate model \( \mathcal{M} \) is defined by the forward rate dynamics in (2.18). A mapping of parameterized forward rate curve is defined as

\[
G: \mathcal{Z} \to C[0, \infty],
\]

with \( \mathcal{Z} \in \mathbb{R}^d \) being the parameter space. The forward rate with time-to-maturity \( \tau \) is written as \( G(\tau; z) \).

Given a mapping of parameterized forward rate curve \( G: \mathcal{Z} \to C[0, \infty] \), the forward curve manifold \( \mathcal{G} \) is defined as

\[
\mathcal{G} = \text{Im} G,
\]
$G = \{ G(\cdot, z) \in C[0, \infty] : z \in \mathbb{Z} \}$

A pair $\{ \mathcal{M}, \mathcal{G} \}$ is consistent if all the forward rate curves generated by the interest rate model $\mathcal{M}$ fall into the family of forward curves $\mathcal{G}$, given that the initial forward curve is in $\mathcal{G}$. Otherwise, the pair $\{ \mathcal{M}, \mathcal{G} \}$ is inconsistent.

**Definition 3.1** Given a forward curve manifold $\mathcal{G}$ and a forward rate process $r(t, x)$ generated by an interest rate model $\mathcal{M}$, $\mathcal{G}$ is said to be invariant under the action of $r$ if, for every initial time $s$, $r(s, \cdot) \in \mathcal{G}$ implies $r(t, \cdot) \in \mathcal{G}$ with $\forall t \geq s$.

Thus, a pair $\{ \mathcal{M}, \mathcal{G} \}$ is consistent if and only if the manifold $\mathcal{G}$ is invariant under the action of $r$. The question is under what conditions $\mathcal{G}$ is $r$-invariant. In Bjork and Christensen (1999), the parameterized forward curve $G(x, z)$ takes a generic form and is not limited to be affine in parameters. $dG(x, z)$ will have an extra term due to Ito’s lemma. If $G(x, z)$ is an affine function of $z$, the extra term from Ito’s lemma disappears and $dG(x, z)$ takes the same form as in regular differential. But, with Stratonovich calculus, the chain rule in ordinary calculus can be used directly in stochastic calculus.

Under Stratonovich calculus, the forward rate process $r(t, x)$ in model $\mathcal{M}$ becomes,

$$dr(t, x) = \left[ \frac{\partial r(t, x)}{\partial x} + \sigma_0(\cdot, x) \int_0^x \sigma_0(\cdot, u) du \right] dt + \varphi(t) dt + \sigma_0(\cdot, x) \circ dW(t), \tag{3.1}$$

where,

$$\varphi(t) dt = -\frac{1}{2} d \langle \sigma_0(\cdot, x), W(t) \rangle$$

and $d \langle \cdot, \cdot \rangle$ denotes a quadratic variation and $\circ$ denotes Stratonovich calculus. The $\cdot$ in $\sigma_0(\cdot, x)$ means that $\sigma_0(\cdot, x)$ depends on $t$ through some state variables. Please refer to Appendix A.1 for details.

With Stratonovich calculus, the $r$-invariance can be rewritten as follows:

**Definition 3.2** Given a forward curve manifold $\mathcal{G}$ and a forward rate process $r(t, x)$, we say that $\mathcal{G}$ is invariant under the action of $r$ if, for each initial forward curve $r_0 \in \mathcal{G}$,
there exists a stochastic process $z(t)$ defined as

$$dz(t) = \gamma_0(t, z(t))dt + \psi_0(t, z(t)) \circ dW(t),$$

such that, for $t \geq 0$,

$$r(t, x) = G(x, z(t)).$$

Now we have the main theorem for consistency in Bjork and Christensen (1999).

**Theorem 3.3** A forward curve manifold $G$ is $r$-invariant for the forward rate process $r(t, x)$ in model $M$ defined by (3.1) if and only if

$$G_z(\cdot, z) + \sigma_0(t, x) \int_0^x \sigma_0(t, u)du + \varphi(t) = G_z(\cdot, z)'\gamma_0(t, z(t)),$$

and

$$\sigma_0(t, x) = G_z(\cdot, z)'\psi_0(t, z(t)),$$

for all $(t, z) \in \mathbb{R}_+ \times \mathcal{Z}$. and $G_z(\cdot, z)'$ is the transpose of $\partial G / \partial z$. $dz(t)$ is defined in equation (3.2).

Equation (3.4) is the consistent drift condition.

Equation (3.5) is the consistent volatility condition.

Note that all differentials are under Stratonovich calculus.

4 Consistent Conditions for Affine Parameterization

Since our focus will remain on affine models, we simplify the consistent conditions in Bjork and Christensen (1999) and restrict ourselves within the affine forms of $G(x, z)$. If $G(x, z)$ is affine in $z$, the additional term from Ito’s lemma, $\frac{1}{2}G''(x, z)(dz)^2 = 0$, disappears. The Stratonovich calculus term becomes the ordinary differential term. The consistent conditions with $G(x, z)$ being affine are expressed in the following theorem.

**Theorem 4.1** When a parameterized forward rate curve $G(x, z)$ is in affine form of $z$, the forward curve manifold $G$ is $r$-invariant, with $r(t, x)$ defined by (2.18), if and only if,

---

4Through the paper, notation $z(t)$ is the same as $z_t$. 

18
for each initial forward curve \( r_0 \in \mathcal{G} \), there exists a stochastic process \( z(t) \) defined as

\[
dz(t) = \gamma(t, z(t))dt + \psi(t, z(t))dW(t),
\]

such that

\[
G_x(\cdot; z) + \sigma_0(t, x) \int_0^x \sigma_0(t, u)du = G_z(\cdot, z)' \gamma_0(t, z(t)),
\]

\[
\sigma_0(t, x) = G_z(\cdot, z)' \psi_0(t, z(t)),
\]

for all \((t, z) \in \mathbb{R}_+ \times \mathbb{Z}\) and \( G_z(\cdot, z)' \) is the transpose of \( \partial G/\partial z \).

Equation (4.2) is the consistent drift condition for affine \( G(x, z) \).

Equation (4.3) is the consistent volatility condition for affine \( G(x, z) \).

Note that all the processes, \( r(t, x) \) and \( G(x, z) \), are under Ito’s dynamics now. There is no extra term \( \varphi(t) \) in the consistent drift condition. The consistent drift and volatility conditions for the affine \( G(x, z) \), together with (3.3), will be used frequently in the following sections to build different forward rate models.

It is worth mentioning that equation (3.3) transforms an infinite dimensional forward rate process in equation (2.18) into a finite dimensional state space system. Just as indicated in Bjork and Christensen (1999), the finite dimensional transformation of a HJM model is not unique. It depends on a pair \{\( G(x, z) \), \( dz(t) \)\} to determine what the finite dimensional transformation looks like.

Remark 4.2 The consistent drift and volatility conditions in (4.3) and (2.16) have some significances in the specification analysis of affine term structure models. Starting with the short rate specification, Dai and Singleton (2000) studied the invariant transformations of affine models and suggested canonical specifications for different affine models. Unlike Dai and Singleton (2000), we start with the forward rate model and provide the conditions that an affine parameterization of forward curve and state variables should satisfy to be consistent with the forward rate model. Then, the affine parameterization of forward curve and the state variables form a consistent affine interest rate model.

Just like the short rate is the short end of the whole forward rate curve, there might exist a link between the specification analysis in Dai and Singleton (2000) and the consistent conditions we have here. This link might unify the two analyses together. Unlike the
short rate models, where the short rate dynamics is first prescribed, the consistent drift and volatility conditions are derived from the forward rate model framework, which only requires the specification of the forward rate volatility. Comparing the dynamics of the forward rate and the short rate in (2.10) and (2.15), we can find that the short rate dynamics is also the short end of the forward rate process. Therefore, the consistent drift and volatility conditions for affine parameterizations may be worthy of further investigation in the specification analysis of affine models.

\textbf{Proposition 4.3} If two affine parameterizations for forward curve are consistent with the same forward rate model through the consistent drift and volatility conditions as in Theorem 4.1, the two affine parameterizations are said to be observationally equivalent.

5 Affine HJM Models with Stochastic Volatility

Empirical studies have pointed out two important issues about interest rate volatility. First, interest rate volatility is stochastic and can depend on the level of the interest rate. In addition, the volatility is in humped shape, which means the volatility function is increasing in the short end of the yield curve, reaching its peak somewhere in the middle, and then decreasing in the long end. Some recently studies, Collin-Dufresne and Goldstein (2002), Heidari and Wu (2003), and Li and Zhao (2006), find the existence of unspanned stochastic volatility, i.e., the volatility that cannot be hedged with a position only containing bonds. Due to the affine structure of the model, these features can be incorporated in our model.

The empirical success of dynamic Nelson-Siegel family models inspires us to keep our models within that family but with some extensions. As Bjork and Christensen (1999) suggested, multiple parameterized forward curves may be consistent with a given interest rate model. It attracts us to simplify the specifications of volatility structure and, meanwhile, maintain a similar form of the forward rate curve. It is especially the case when we introduce multifactor HJM models, where different components may overlap in shaping the forward curve.

Our models with general volatility follow a forward rate volatility structure with \( L(\text{evel}), S(\text{lope}), \text{ and } C(\text{urvature}) \) in terms of time-to-maturity \( x \). It is called \( LSC \)
volatility:
\[
\sigma_0(t, x) = [\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}] \sqrt{h_t}, \tag{5.1}
\]
where, \(\sigma_1\) and \(\sigma_2\) are constant, \(h_t\) is a stochastic process which can maintain Markov property of the model as defined in equation (2.16). This volatility structure has components of level, slope, and curvature in terms of time-to-maturity \(x\). It also has a component of stochastic volatility, which can be either spanned, unspanned, or mixed, depending on the specification of \(\sqrt{h_t}\). Later, we show that the forward rate volatility structure of level, slope, and curvature can be transmitted to the forward rate curve with the same geometric meanings. The decomposition of this volatility specification is discussed in the following section with Gaussian volatility (\(\sqrt{h_t} = 1\)) and the same decomposition can be applied to the cases with stochastic volatility. The Gaussian models discussed here will not be implemented in the empirical analysis but they certainly can be used. We leave this for further research since this paper focuses on unspanned stochastic volatility.

5.1 Affine HJM models with Gaussian Volatility

When \(\sqrt{h_t} = 1\), we get Gaussian deterministic LSC volatility as
\[
\sigma_0(t, x) = [\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}]. \tag{5.2}
\]
Now the model becomes a Gaussian model. Here are some spacial examples,

1. Ho and Lee. (1986) model: L(evel) volatility

\[
\sigma_0(t, x) = \sigma_1, \text{ if } \sigma_2 = 0, \sigma_3 = 0.
\]

2. Hull and White (1990) model: S(lope) volatility

\[
\sigma_0(t, x) = \sigma_2 e^{-ax}, \text{ if } \sigma_1 = \sigma_3 = 0.
\]

3. Mercurio and Moraleda (2000) model: S(lope) and C(urvature), or SC volatility

\[
\sigma_0(t, x) = [(\sigma_2 + \sigma_3 x)e^{-ax}], \text{ if } \sigma_1 = 0. \tag{5.3}
\]
The following discussion shows that the level, slope and curvature components of the forward rate volatility can be transmitted in the same shape to the forward rate curve. More importantly, the level, slope, and curvature can also be combined in any way desired. Only one-factor model is discussed here for illustration purposes. In multifactor cases, each factor can have different combination of level, slope and curvature. For example, a three-factor model can have $LSC$ volatility for each factor. Or, the first factor has $LSC$ volatility, the second factor $S$ volatility, and the third factor $C$ volatility. Various combinations can also be studied as natural extensions of this paper.

Mercurio and Moraleda (2000) studied a one-factor HJM model with volatility in equation (5.3) and focused only on the interest rate derivative market. Compared with equation (5.3), (5.2) has an extra term, $\sigma_1$, which serves as the level of the volatility. Clearly, the volatility specification in equation (5.2) nests all other volatility specifications.

Interestingly, if we rewrite equation (5.2) as

$$\sigma_0(t, x) = \sigma_1 + \sigma_2 e^{-\alpha x} + \sigma_3 x e^{-\alpha x},$$

it actually shows that the volatility specification in equation (5.2) nests a constant (Ho-Lee), an exponentially decaying (Hull-White) component, and a humped component (Mercurio and Moraleda). Moreover, it can be viewed as a more general form since it combines the Ho-Lee volatility ($\sigma_1$), Hull-White volatility ($\sigma_2 e^{-\alpha x}$), and Mercurio-Moraleda volatility ($\sigma_2 + \sigma_3 x e^{-\alpha x}$). The three coefficients for $\sigma_1$, $\sigma_2$, and $\sigma_3$ are 1, $e^{-\alpha x}$, and $xe^{-\alpha x}$, which have shapes of level, slope, and curvature in terms of time-to-maturity $x$. Even though the level of the volatility, $\sigma_1$, can be achieved by setting $\sigma_3 = \alpha = 0$ in volatility form ($\sigma_2 + \sigma_3 x e^{-\alpha x}$), the volatility in equation (5.2) itself is a more general form since it combines the Ho-Lee, Hull-White, and Mercurio-Moraleda models.

Trolle and Schwartz (2009) adopted the volatility form in Mercurio and Moraleda (2000) and added a stochastic term to it. We extend Trolle and Schwartz (2009) to the volatility structure in equation (5.2) and also add a stochastic term, $\sqrt{\mu_t}$. As shown above, our model nests Trolle and Schwartz (2009).

In multifactor case, if each factor have the same form of volatility structure in (5.2) with different values of $\sigma_i$s, the consistent forward rate parameterization for each factor will have the same shape. It means that the cross-sectional relation among the forward...
rates can be captured with one factor. But empirical analysis shows that three factors are required to well capture the time series evolution of the forward rate curve. Therefore, even though one factor with volatility in (5.2) may be enough to capture the cross section of forward rates, it is not enough to capture the time series evolution of the forward rate curve. This observation also inspires a simpler volatility structure in the multifactor case, where the cross-sectional relation among the forward rates is captured with each factor bestowed with only one term in equation (5.2). It results in our consistent Augmented Nelson-Siegel model in the following section.

In (5.2), the three coefficients of $\sigma, s_1, e^{-ax}$, and $xe^{-ax}$, representing the shape of level, slope, and curvature in terms of $x$, are interestingly the same as the three parameter loadings in the Nelson-Siegel model. Trolle and Schwartz (2009) adopted the volatility structure in Mercurio and Moraleda (2000), which does not include a level component in the volatility and thus cannot generate the level component in the forward rate curve. But our volatility structure in equation (5.4) overcomes this shortcoming by adding $\sigma_1$. It is more suitable to obtain a consistent modified Nelson-Siegel model. This is another reason why we choose the volatility structure in equation (5.2) as our general form of volatility.

Falini (2010) studied a three-factor HJM model with volatility in equation (5.2) and extends to spanned stochastic volatility. Trolle and Schwartz (2009) and Falini (2010) built their models following Chiarella and Kwon (2003), where state variables don’t have clear geometric meanings. This paper, on the other hand, extends their work to a more general form of volatility structure with unspanned stochastic volatility. Meanwhile, with Theorem 4.1, the state variables in our model have clear geometric meanings just like those in the Nelson-Siegel model.

5.1.1 Affine HJM models with Gaussian Volatility

Before moving further, we declare some definitions since throughout this paper, they will be repeatedly used and they have some differences from their usual meanings.

**Definition 5.1** State variable are defined as the variables which drive the yield curve to move forward.
Definition 5.2 Factors are defined as the state variables which have contemporaneous Brownian motions. Only factors introduce randomness to the system.

Definition 5.3 Auxiliary state variables are those state variables without contemporaneous Brownian motions.

By definitions above, factors are state variables but not vice versa. This makes our models different from the usual affine term structure models, where the dimension of the state variable is the same as that of the factor. In our affine HJM forward rate models, the dimension of the state variable is larger than that of the factor. The auxiliary state variables help rule out the arbitrage opportunity. These definitions enable us to fully decompose the yield curve according to the geometric meanings of the factors since they introduce the randomness to the system.

Now with the consistent drift and volatility conditions for affine \( G(z, x) \), we can have our consistent affine parameterization for the forward rate curve.

Proposition 5.4 An interest rate model \( M \) defined by equation (5.2) with LSC volatility is consistent with a family \( G \) of forward curves generated by the parameterization

\[
G(z, x) = [\sigma_1 + (\sigma_2 + \sigma_3x)e^{-ax}]z_1 + z_2 + z_3x \\
+ z_4e^{-ax} + z_5e^{-2ax} + z_6xe^{-ax} + z_7xe^{-2ax} + z_8x^2e^{-ax} + z_9x^2e^{-2ax},
\]

(5.5)

with the dynamics of state variables as

\[
dz(t) = \begin{pmatrix}
\frac{(\sigma_2 + \sigma_3)}{a^2} - az_1 \\
\sigma_1z_1 + z_3 \\
\sigma_1^2 \\
-\frac{\sigma_1(2\sigma_2 + \sigma_3)}{a^2} + \sigma_3z_1 - az_4 + z_6 \\
-\frac{\sigma_2(\sigma_2 + \sigma_3)}{a^2} - 2az_5 + z_7 \\
\sigma_1\frac{(\sigma_2 - \sigma_1)}{a} - az_6 + 2z_8 \\
-\frac{\sigma_3(2\sigma_2 + \sigma_1)}{a^2} - 2az_7 + 2z_9 \\
\sigma_1\sigma_3 - az_8 \\
-\frac{\sigma_1^2}{a} - 2az_9
\end{pmatrix} dt + \begin{pmatrix}
1 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix} dW(t),
\]

(5.6)
and the forward rate curve as

\[ r(t, x) = G(z, x), \tag{5.7} \]

with the initial forward cure \( G(z_0, x) = r(0, x) \) and \( z = (z_1, \ldots, z_9)' \).

**Proof.** For a one-factor HJM model, \( \mathcal{M} \), with volatility in equation (5.2)

\[ \sigma_0(t, x) = \sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax}, \]

we have a forward rate process

\[ dr(t, x) = \left\{ \frac{\partial r(t, x)}{\partial x} + \sigma_0(t, x) \int_0^x \sigma_0(t, u) du \right\} dt + \sigma_0(t, x) dW(t). \tag{5.8} \]

We can confirm that the consistent drift condition for the affine parameterization \( G(z, x) \) defined by equation (5.5) as follows.

\[
G_x(\cdot; z) + \sigma_0(t, x) \int_0^x \sigma_0(t, u) du \\
= \left( \frac{(a\sigma_2 + \sigma_3)}{a^2} - az_1 \right) \left[ \sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax} \right] + (a\sigma_1 z_1 + z_3) + \sigma_1^2 \\
+ \left( -\frac{\sigma_1(a\sigma_2 + \sigma_3)}{a^2} + \sigma_3 z_1 - a z_4 + z_6 \right) e^{-ax} + \left( -\frac{\sigma_2(a\sigma_2 + \sigma_3)}{a^2} - 2az_5 + z_7 \right) e^{-2ax} \\
+ \left( \frac{\sigma_1(a\sigma_2 - \sigma_3)}{a} - az_6 + 2z_8 \right) xe^{-ax} + \left( -\frac{\sigma_3(2a\sigma_2 + \sigma_3)}{a^2} - 2az_7 + 2z_9 \right) x^2 e^{-2ax} \\
+ (\sigma_1 \sigma_3 - az_8) x^2 e^{-ax} + \left( -\frac{\sigma_3^2}{a} - 2az_9 \right) x^2 e^{-2ax}
\]

and

\[
G_z(\cdot; z)' \gamma_0(t, z(t)) \\
= \left( \left[ \sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax} \right], 1, xe^{-ax}, x e^{-2ax}, xe^{-ax}, xe^{-2ax}, x^2 e^{-ax}, x^2 e^{-2ax} \right) \gamma_0(t, z(t)),
\]

with

\[
G_x(\cdot; z) + \sigma_0(t, x) \int_0^x \sigma_0(t, u) du = G_z(\cdot; z)' \gamma_0(t, z(t)),
\]
when

\[
\gamma_0(t, z(t)) = \begin{pmatrix}
\frac{(\sigma_2 + \sigma_3)}{a^2} - a z_1 \\
\sigma_1 z_1 + z_3 \\
\sigma_1^2 \\
-\frac{\sigma_1 (\sigma_2 + \sigma_1)}{a^2} + \sigma_3 z_1 - a z_4 + z_6 \\
-\frac{\sigma_2 (\sigma_2 + \sigma_1)}{a^2} - 2 a z_5 + z_7 \\
\frac{\sigma_1 (\sigma_2 - \sigma_3)}{a} - a z_6 + 2 z_8 \\
-\frac{\sigma_3 (2 \sigma_2 + \sigma_3)}{a^2} - 2 a z_7 + 2 z_9 \\
\sigma_1 \sigma_3 - a z_8 \\
-\frac{\sigma_3^2}{a} - 2 a z_9
\end{pmatrix}.
\]

The confirmation of the consistent volatility condition comes simpler as

\[
(\sigma_1 + \sigma_2 x) e^{-ax}
\]

\[
= \left( [\sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax}], 1, x, e^{-ax}, e^{-2ax}, xe^{-ax}, xe^{-2ax}, x^2 e^{-ax}, x^2 e^{-2ax} \right)' \psi_0(t, z(t)),
\]

with

\[
\psi_0(t, z(t)) = (1, 0, 0, 0, 0, 0, 0, 0)'.
\]

Now we see the pair \( \{M, G\} \) is consistent. 

Note that only \( z_1 \) has a diffusion term. All other state variables don’t have diffusion terms and thus don’t have risk premium when we change the probability measures, from risk-neutral to physical.

**Remark 5.5** The lower-order exponential terms in \( G(z, x) \) can be generated from the higher-order ones. For instance, the term containing \( e^{-ax} \) can be generated from term with \( xe^{-ax} \) through \( G_x'(z, x) \) in the consistent conditions. It seems that we might save some terms in \( G(z, x) \) and lower the dimension of the state variables. But this is not the case since the consistent condition should be valid for all \( z \). It can be shown, for example, that if we leave out the terms of \( z_4 e^{-ax} \) and \( z_6 xe^{-ax} \), since they can be generated by \( \sigma_2 e^{-ax} z_1 \) and \( \sigma_3 xe^{-ax} z_1 \), then \( z_1 \) and \( z_9 \) should satisfy some constraint, which is in conflict with the requirement that consistent conditions should be valid for all \( z \).

**Remark 5.6** A note on the initial forward rate curve. In bond pricing, the initial forward rate curve is not required to be fitted perfectly. However, when the model is used to price
interest rate derivatives, the initial forward rate curve needs to be fitted perfectly. As we start with a forward rate model, this requirement can be easily met since the initial forward rate curve is included by construction. To make this point more explicit, we can rewrite the forward rate curve as

\[ r(t, x) = r(0, x) + G(z, x), \]

with \( r(0, x) \) being the initial forward rate curve, or, \( r(0, x) = G(z_0) \) from consistency. But with this transformation, the state variables should have initial value \( z_0 = 0 \). It is easy to confirm that the expression of the forward rate curve in equation (5.7) is equivalent to the forward rate curve in equation (5.9) due to the affine form of \( G(z, x) \) in \( z \). In Wang (2014a), the forward rate process in equation (5.9) is used.

**Remark 5.7** A further look at the consistent drift condition shows that \( (a\sigma^2 + \sigma_3) \) can be interpreted as the long-run mean of the state variable \( z_1 \). This interpretation is very useful in our case of unspanned stochastic volatility where the stochastic volatility component serves as the kernel of the long-run mean of the state variable. It might help explain the finding in Andersen and Benzoni (2010) that bond yield volatility is largely unrelated to the bond yields. This is because the long-run mean of the forward rate or yield is unspanned with bond yields.

By comparing the two parameterizations consistent with the volatility structure defined in equations (5.3) and (5.2), we see that the effect on the consistent parameterizations of adding an extra constant in the volatility structure. Proposition 5.8 gives a parameterization consistent with the volatility in equation (5.3). It can be seen that the single factor contributes to the level, slope, and curvature of the forward curve, or the cross-sectional relation of the forward curve. To be more specific, the level is \( \sigma_1 z_1 \), the slope is \( \sigma_2 z_1 \), and the curvature is \( \sigma_3 xe^{-ax} z_1 \).

As shown above, the shape of the forward rate volatility is transmitted to the shape of the forward rate curve. The one-factor model, however, is limited because only one factor is used to capture the time series movement of the yield curve. In our empirical analysis, the one-factor model is always outperformed by the models with more factors. Another point worth mentioning is that auxiliary state variables also contribute to the shape of the yield curve in terms of slope and curvature. This contribution rules out the arbitrage
opportunity among the cross-sectional yields. But as mentioned earlier and shown above, the auxiliary state variables do not carry contemporaneous randomness, which means they cannot bring randomness to the system. From this point of view, these variables are one-period-ahead predictable and only bring carry-over effects from last period.

**Proposition 5.8** An interest rate model $\mathcal{M}$ defined by the SC volatility, $(\sigma_1 + \sigma_2 x) e^{-a_1 x}$, as in equation (5.3), is consistent with a family $\mathcal{G}$ of forward curves generated by the parameterization

$$G(z, x) = z_1 (\sigma_1 + \sigma_2 x) e^{-a_1 x} + z_2 e^{-a_1 x} + z_3 e^{-2a_1 x} + z_4 x e^{-2a_1 x} + z_5 x^2 e^{-2a_1 x}$$

(5.10)

with the dynamics of state variables as

$$dz(t) = \begin{bmatrix}
\frac{\sigma_1 a_1 + \sigma_2}{a_1^2} - a_1 z_1 \\
\sigma_2 z_1 - a_1 z_2 \\
z_4 - \frac{\sigma_1(a_1 z_1 + \sigma_2)}{a_1^2} - 2a_1 z_3 \\
2z_5 - \frac{\sigma_2(2a_1 z_1 + \sigma_2)}{a_1^2} - 2a_1 z_4 \\
\frac{\sigma_2}{a} - 2a z_5
\end{bmatrix} dt + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} dW(t),$$

(5.11)

and the forward rate curve as

$$r(t, x) = G(z, x),$$

with the initial forward curve $G(z_0, x) = r(0, x)$ and $z = (z_1, \ldots, z_5)'$.

Since, for illustration purposes, only a one-factor model is used, we expect that this factor corresponds to the level factor in the traditional term structure models. In equation (5.5), we have a term $[\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}]z_1$, in which $\sigma_1 z_1$ represents this level factor. Of course, $z_1$ also contributes to the slope and curvature of the forward curve. This is because we are using a general form of volatility. In Section 5.3, where we discuss the consistent Augmented Nelson-Siegel model and specify the volatility structure as a special form of equation (5.5), we will have each factor represents one of the three components, the level, slope and curvature.

By comparison, equation (5.10), however, only has $z_1 (\sigma_1 + \sigma_2 x) e^{-a_1 x}$, in which no term represents the level factor in the forward curve, even though $z_1$ corresponds to the
level factor. It means that, for the one-factor model in Proposition 5.8, this factor is supposedly to represent the level factor in the forward curve but it doesn’t appear as level term in the forward rate formula. This is why we choose the volatility structure in equation (5.5) as our most general volatility form. It also reflects a conflict in tradition term structure models, where the factor representing a certain geometric meaning of the forward curve does not appear in that geometric form in the forward rate formula.

Now we can rewrite the consistent parameterization in Proposition 5.4 in the traditional affine form. If a pair, \( \{ \mathcal{M}, G \} \), is consistent as defined in Proposition 5.4, we have

\[
 r(t, x) = G(z, x), \tag{5.12}
\]

for \( t > 0 \), and the initial forward curve, \( G(Z_0, x) = r_0 = r(0, x) \). The dynamics of \( dZ(t) \) are defined in equation (5.6). By stacking them up, we have

\[
 r(t, x) = \delta' z(t), \tag{5.13}
\]

\[
 dz(t) = \gamma_0(t, z(t)) dt + \psi_0(t, z(t)) dW(t), \tag{5.14}
\]

where,

\[
 \delta = \left( \left[ \sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax} \right], 1, x, e^{-ax}, e^{-2ax}, xe^{-ax}, xe^{-2ax}, xe^{-2ax}, xe^{-2ax}, xe^{-2ax}, xe^{-2ax}, xe^{-2ax} \right)' \tag{5.15}
\]

and \( z = (z_1, z_2, \ldots, z_9)' \).

With the proper choice of \( \sigma_1, \sigma_2, \) and \( a \), \( G(z, x) \) in equation (5.5) is consistent with a one-factor Ho-Lee and Hull-White model. However, equation (5.5) is not the minimum parameterization for one-factor Ho-Lee and Hull-White model. As in Bjork and Christensen (1999), we have the following two propositions.

**Proposition 5.9** For a one-factor Ho-Lee model defined by \( \sigma_0(t, x) = \sigma_1 \), L volatility, we have the following consistent parameterization as

\[
 G(z, x) = z_1 + z_2 x, \tag{5.16}
\]
with

\[
dz(t) = \begin{pmatrix} z_2 \\ \frac{\sigma_2^2}{\sigma_1^2} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} dW(t)
\] (5.17)

and \( z = (z_1, z_2)' \).

**Proposition 5.10** For a one-factor Hull-White model defined by \( \sigma_0(t, x) = \sigma_1 e^{-a_1 x} \), \( S \) volatility, we have a consistent parameterization as

\[
G(z, x) = z_1 e^{-ax} + z_2 e^{-2ax},
\] (5.18)

with

\[
dz(t) = \begin{pmatrix} -az_1 + \frac{\sigma_1^2}{a} \\ -2a z_2 - \frac{\sigma_2^2}{a} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} dW(t)
\] (5.19)

and \( z = (z_1, z_2)' \).

In Proposition 5.9, the consistent parameterization \( G(z, x) \) has one state variable, \( z_1 \), which corresponds to the level of the yield curve. The other state variable, \( z_2 \), is deterministic and can be interpreted as the slope of yield curve, which is a straight slope, instead of the exponential decaying slope. It looks controversial since it is monotonic in time to maturity. But to rule out arbitrage opportunity, we have to leave it there. Similarly, in Proposition 5.10, the consistent parameterization has one stochastic and one deterministic state variable. Different from the consistent Ho-Le model, the two state variables in consistent Hull-White model can all be interpreted as slopes of the yield curve, which are exponentially decaying slopes due to \( e^{-ax} \) and \( e^{-2ax} \). Unfortunately, it cannot generate a level factor in the Hull-White model. In both models, only \( z_1 \) has a contemporaneous Brownian motion.

From the curve fitting point of view, we do not need higher-order exponentials than those in the Nelson-Siegel and Svensson models. Only when we need a more flexible volatility structure will we choose the minimum parameterization for a concrete volatility specification. Another point worth mentioning is that, in the dynamics of state variable equation (5.6), only \( z_1 \) has a diffusion term. All other state variables only have drift terms. It is also true in equations (5.17) and (5.19). This property makes easier the change of the probability measures since here we only have one source of randomness in the market. Therefore, only one state variable has risk premium after being changed
under the physical probability measure. It will be addressed in details in Section 8 when we discuss the market price of risk.

As this paper mainly focuses on the cases with stochastic volatility, we will leave out the Gaussian multifactor cases. However, multifactor Gaussian models with \(LSC\) volatility structure and \(SC\) volatility structure, respectively, as

\[
\sigma_{i,0}(t,x) = \sigma_{i,1} + \sigma_{i,2}e^{-a_{i}x} + \sigma_{i,3}xe^{-a_{i}x}, \text{ for } i = 1, 2, \ldots, n,
\]

and

\[
\sigma_{i,0}(t,x) = (\sigma_{i,2} + \sigma_{i,3}x)e^{-a_{i}x}, \text{ for } i = 1, 2, \ldots, n,
\]

can be easily obtained with Proposition 5.4 and Proposition 5.8 by stacking up \(G(z_{i}, x)\) and \(dz_{i}\). These models will be studied in other projects.

### 5.2 Affine HJM Models with Stochastic Volatility

There are two ways to add stochastic volatility to interest rate models. One way does not involve adding additional sources of randomness. This is also the most common and straightforward way. As all the factors bearing randomness appear in the forward rate formula, all risks can be hedged or spanned with a position only containing bonds. This kind of stochastic volatility is what we call spanned stochastic volatility.

The other way to add stochastic volatility, in a sense of Heston (1993), involves adding additional randomness for the stochastic volatility components. It means that some stochastic volatility component may have its own source of randomness, or Brownian motion. If a factor bearing some randomness does not appear in the forward rate formula, this randomness, or risk, cannot be hedged or spanned with a bond-only position. This kind of stochastic volatility is dubbed as unspanned stochastic volatility (USV) in Collin-Dufresne and Goldstein (2002), Li and Zhao (2006), Heidari and Wu (2009), Trolle and Schwartz (2009), and Andersen and Benzoni (2010) among others. With unspanned stochastic volatility, the bond market itself becomes incomplete because positions containing only bonds cannot hedge all the risks in the bond market. Only when the interest rate derivative market is included can the two markets together, bond market and derivative market, become complete again.

Whether the stochastic volatility is spanned or unspanned, it totally depends on the
stochastic volatility factor. When it bears its own randomness and doesn’t appear in the forward rate formula, or equivalently the yield or bond price formula, this randomness becomes unspanned. This section first discusses the spanned stochastic volatility then moves to the unspanned stochastic volatility. The following shows that, in our model framework, adding stochastic volatility, spanned or unspanned, will not affect the forward rate parameterization. This property facilitates us to examine the geometric information of the level, slope, curvature of the forward rate curve. It also shows that the unspanned stochastic volatility can be interpreted as the kernel of the long-run mean of yield factor. Of course, in the case of unspanned stochastic volatility, one state variable is added for each unspanned source of risk.

For illustration purposes, the stochastic volatility \( h_t \) is assumed to follow a square-root process and solves the stochastic differential equation

\[
dh_t = \kappa (\theta - h_t) dt + \sqrt{h_t} dB_t, \tag{5.20}
\]

where \( \kappa \) is the mean-reversion speed and \( \theta \) is the long-run mean of \( h_t \). When \( \text{corr}(dB_t, dW_t) = \pm 1 \), we have spanned stochastic volatility. Otherwise, when \( \text{corr}(dB_t, dW_t) \neq \pm 1 \), we have unspanned stochastic volatility. Furthermore, under spanned stochastic volatility, \( \text{corr}(dB_t, dW_t) = \pm 1 \), we don’t need an explicit expression of \( dh_t \) since it can be spanned by \( z_t \).

Proper constraints on the parameters will be discussed when the specific model is introduced. The following discussion shows that the volatility specification in our model can be very flexible. Starting from Cox, Ingersoll, and Ross (1985a), affine term structure models usually depend on the square-root process to introduce stochastic volatility. The square-root process can keep the model within the affine family under both risk-neutral and physical measures, with a proper specification of market price of risk. The square-root process, with proper constraints on the parameters, will stay non-negative or strictly positive without attaining zero. We also specify our stochastic volatility as a square-root process.

This section shows some flexible choices of the stochastic volatility for our model while keeping the forward rate process unchanged. Thus, the geometric meaning of each factor maintains.
5.2.1 Spanned Stochastic volatility

Most affine term structure models with stochastic volatility have spanned stochastic volatility, where no additional randomness is introduced in the model. The volatility variables also enter into the bond price, or equivalently the yields and forward rates. Within our model framework, one natural choice of spanned volatility is to choose $h_{i,t} = z_{i,1,t}$. But to keep the forward rate formula unchanged, we have to manipulate the state variable process to satisfy the constraints required by the square-root process. Before getting into details, we first introduce as examples some possible choices of the spanned stochastic volatility.

Let $r_t$ be the short rate, a short-rate-dependent stochastic volatility can be defined as

$$\sqrt{h(t)} = \sqrt{r_t}.$$ (5.21)

With $\tau_i$ being some fixed time-to-maturity, a forward-rate-$\tau_i$-dependent stochastic volatility can be defined as

$$\sqrt{h(t)} = \sqrt{r(t, \tau_i)}.$$ (5.22)

Amin and Morton (1994) studied a one-factor model with the stochastic component of volatility in the form of $r(t, x)^\gamma$. Their model is not Markovian any more since the volatility is not separable in the way of equation (2.16).

The specification of $h(t)$ can also be the sum of a number of fixed forward rates as

$$\sqrt{h(t)} = \sqrt{\sum_{i=1}^{n} m_i r(t, \tau_i)}.$$ (5.23)

Of course, $h(t)$ can also depend on an affine function of $z(t)s$. For example,

$$\sqrt{h(t)} = \sqrt{g(z(t))},$$ (5.24)

where $g(z(t))$ is a non-negative affine function in $z(t)s$.

The volatility specifications listed above just show the flexibility of our model framework. The concrete model with $h_t$ above can be derived based on the following proposition. For different specifications of $h(t)$, some restrictions on the parameters are usually required to achieve admissibility as in Duffie and Kan (1996) and Dai and Singleton (2000). Here,
we only illustrate the flexibility of the model framework. We leave the details for future research since the general volatility specification we choose in this paper can achieve admissibility as specified.

Now with our consistent conditions for the affine $G(z, x)$, we can get an affine HJM model with spanned stochastic volatility $h_t$. In Proposition 5.11, $h_t$ still takes the general form as in equation (5.20). Later, we will show the special case when $h_t = z_{1,t}$. Due to the constraints on parameters in the square-root process, we need extra manipulations of the state variable process.

**Proposition 5.11** An interest rate model $M$ defined by the volatility

\[ \sigma_0(t, x) = [\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}]\sqrt{h_t} \]  

(5.25)

is consistent with the affine parameterization $G(z, x)$,

\[ G(z, x) = [\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}]z_1 + z_2 + z_3 x 
+ z_4 e^{-ax} + z_5 e^{-2ax} + z_6 x e^{-ax} + z_7 x e^{-2ax} + z_8 x^2 e^{-ax} + z_9 x^2 e^{-2ax}, \]  

(5.26)

with

\[
\begin{pmatrix}
\frac{(a \sigma_2 + \sigma_3)}{a^2} h_t - a z_1 \\
a \sigma_1 z_1 + z_3 \\
\sigma_1^2 h_t \\
-\frac{\sigma_1 (a \sigma_2 + \sigma_3)}{a^2} h_t + \sigma_3 z_1 - a z_4 + z_6 \\
-\frac{\sigma_2 (a \sigma_2 + \sigma_3)}{a^2} h_t - 2 a z_5 + z_7 \\
\frac{\sigma_1 (a \sigma_2 - \sigma_3)}{a} h_t - a z_6 + 2 z_8 \\
-\frac{\sigma_3 (2a \sigma_2 + \sigma_3)}{a^2} h_t - 2 a z_7 + 2 z_9 \\
\sigma_1 \sigma_3 h_t - a z_8 \\
-\frac{\sigma_3^2}{a} h_t - 2 a z_9
\end{pmatrix}
\begin{pmatrix}
dt \\
0 \\
dW(t)
\end{pmatrix}
\]

(5.27)

and the initial forward curve $G(z_0, x) = r(0, x)$. Then the forward rate model becomes

\[ r(t, x) = G(z, x) \]  

(5.28)
with the dynamics of state variables, \( dz(t) \), following equation (5.27) and \( z = (z_1, \ldots, z_9)' \).

It is easy to see that the parameterization with stochastic volatility takes the same form as the corresponding deterministic volatility model. The difference is that the dynamics of state variables have a stochastic volatility component now. This can be seen by comparing equations (5.27) and (5.6). By the consistent drift condition for affine \( G(z, x) \) in equation (4.3), we have the integral term

\[
\sigma_0(t, x) \int_0^x \sigma_0(t, u) du.
\]

This is the only term involving the volatility structure \( \sigma_0(t, x) \). With the specification of

\[
\sigma_0(t, x) = [\sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax}] \sqrt{h_t},
\]

we have

\[
\sigma_0(t, x) \int_0^x \sigma_0(t, u) du = [\sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax}] \sqrt{h_t} \int_0^x [\sigma_1 + (\sigma_2 + \sigma_3 u) e^{-au}] \sqrt{h_t} du
\]

\[
= [\sigma_1 + (\sigma_2 + \sigma_3 x) e^{-ax}] h_t \int_0^x [\sigma_1 + (\sigma_2 + \sigma_3 u) e^{-au}] du.
\]

By collecting the corresponding terms in the consistent conditions, we find that the volatility term only shows in the state dynamics and not in \( G(z, x) \). Detailed proof can be found in Appendix A.2.

As in the Gaussian case, the whole model is still within the family of affine models. In the dynamics of factor \( z_1 \), the long-run mean of \( z_1 \) is \( (a\sigma_2 + \sigma_3) h_t \), which is also stochastic. Now the stochastic volatility not only plays a part in the forward rate volatility but also serves as the long-run mean of factor \( z_1 \). Because \( h_t \) is stochastic not a constant, the long-run mean has different meaning here. We call it shadow mean in this paper. Meanwhile, it does not directly enter into the forward rate formula, or equivalently the bond yield formula. This observation is very useful in case of unspanned stochastic volatility. Details will be discussed later.

A natural choice of spanned stochastic volatility is to set \( h_t = z_{1,t} \). This choice immediately makes the model look like an ordinary affine term structure model. However, the difference lies in fact that factors in our model have clear geometric meanings. Since
now \(z_{1,t}\) acts as the stochastic volatility, even though \(G(z, x)\) stays unchanged, the state variable process needs some extra care. This is due to the parameter constraints required by the square-root process. If, in equation (5.27), we set \(h_t = z_1\), the long-run mean of \(z_1\) becomes 0. It can not guarantee that \(z_1\) stays positive. Then the model becomes inadmissible. However, it turns out that, by a slight manipulation, some constraints can be added to make the model admissible again.

**Proposition 5.12** An interest rate model \(M\) defined by the volatility

\[
\sigma_0(t, x) = [\sigma_1 + (\sigma_2 + \sigma_3x)e^{-ax}]\sqrt{z_{1,t}}
\]

is consistent with the affine parameterization \(G(z, x)\),

\[
G(z, x) = [\sigma_1 + (\sigma_2 + \sigma_3x)e^{-ax}]z_1 + z_2 + z_3x
+ z_4e^{-ax} + z_5e^{-2ax} + z_6xe^{-ax} + z_7xe^{-2ax} + z_8x^2e^{-ax} + z_9x^2e^{-2ax},
\]

with

\[
dz(t) =
\begin{pmatrix}
\xi - \frac{a^3 - (a\sigma_2 + \sigma_3)a}{a^2}z_1 \\
-\xi\sigma_1 - a\sigma_1z_1 + z_3 \\
\sigma_1^2z_1 \\
-\xi\sigma_2 + \frac{\sigma_3a^2 - \sigma_1(a\sigma_2 + \sigma_3)}{a^2}z_1 - az_4 + z_6 \\
-\frac{\sigma_2(\sigma_2 + \sigma_3)}{a^2}z_1 - 2az_5 + z_7 \\
-\xi\sigma_3 + \frac{\sigma_3a(\sigma_2 - \sigma_3)}{a}z_1 - az_6 + 2z_8 \\
-\frac{\sigma_3(2a\sigma_2 + \sigma_3)}{a^2}z_1 - 2az_7 + 2z_9 \\
\sigma_1\sigma_3z_1 - az_8 \\
-\frac{\sigma_3^2}{a}z_1 - 2az_9
\end{pmatrix}
\begin{pmatrix}
\sqrt{z_1} \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}
\begin{pmatrix}
dt \\
dW(t)
\end{pmatrix}
\]

and the initial forward curve \(G(z_0, x) = r(0, x)\). Then the forward rate model becomes

\[
r(t, x) = G(z, x)
\]

with the dynamics of state variable following equation (5.27) and \(z = (z_1, \ldots, z_9)'\). \(\xi \geq \frac{1}{2}\) makes \(z_1\) stay strictly positive.
The proof can be found in Wang (2014b) which covers the spanned stochastic volatility models. According to Cheridito, Filipovic, and Kimmel (2007), the boundary condition of non-attainment should be satisfied under both the risk-neutral probability measure $Q$ and the physical probability measure $P$. Otherwise, equivalence between the $Q$-measure and $P$-measure cannot be guaranteed. Without the new state variable process in equation (5.31), if we only substitute $h_t$ for $z_t$ in equation (5.27), the non-attainment condition is not satisfied in Proposition 5.12. Even though, with the market price of risk, we can set up a square-root process for $z_1$ with the proper constraints on the parameters to maintain the non-attainment condition under the $P$-measure, there is still no guarantee that the state process under the $Q$-measure can meet the non-attainment condition. Or, there is no guarantee that the $Q$-measure and $P$-measure are equivalent.

This matter can easily be overlooked as the dynamics of state variables under the $Q$-measure is usually not of interest in pricing bonds. As is to see later, when the unspanned stochastic volatility is introduced, this extra care is not required because the dynamics of the unspanned stochastic volatility factor can be prescribed as desire. Thus, there is no need to manipulate $z_s$ there. This will be further addressed when the market price of risk is discussed.

Proposition 5.12 is for a one-factor model. Multifactor models can be obtained by the sum of $G_i(z_i, x)$ and stacking up $dz_i(t)$, with each $G_i(z_i, x)$ and $dz_i(t)$ following the same affine forms as in equations (5.30) and (5.31) with the corresponding parameters, respectively. As this paper focuses on unspanned stochastic volatility, we only introduce the spanned stochastic volatility for completion purposes.

Similar to the Gaussian model, the factor $z_1$ contributes to the level $\sigma_1 z_1$, slope $\sigma_2 e^{-ax} z_1$, and curvature $\sigma_3 e^{-ax} z_1$ of the forward rate curve. As mentioned before, even though $z_1$ can generate $LSC$ for the cross-section of forward rates, it is limited in capturing the time series evolution of the forward rate curve. On the other hand, in multifactor case, each factor contributes to the $LSC$ of the forward rate curve and all factors capture the time evolution of the forward curve.

5.2.2 Unspanned Stochastic Volatility

Empirical studies have shown that the interest rate derivative market has its own risk factors different from those in the bond market. It means that the factors driving the
bond market can not fully explain the derivative market. Thus, a position only containing bonds can not hedge all the interest rate risks. Collin-Dufresne and Goldstein (2003) used interest rate straddles to provide the empirical evidence of unspanned stochastic volatility. Heidari and Wu (2003) used principal component analysis and showed that the volatility surface of swaptions has three orthogonal movements, independent of the principal movements of the yield curve. Han (2007) analyzed the swaptions and caps data and reported that separate volatility for the derivative market is useful in explaining the cross-sectional and time series variations of swaption implied volatilities.

In this paper we will also address these issues by including separate stochastic volatility factors. As mentioned in Bjork, Landen, and Svensson (2004), including an additional stochastic volatility factor $v(t)$ needs the expansion of the manifold $\mathcal{G}$ of parameterized forward curves as $\mathcal{G} : IM\{G(z, x), v(t)\}$. Then we can still follow the same consistent conditions to build a consistent parameterization $G(z, x)$.

Several points need to be made. First, as Bjork and Christensen (1999) stated, the parameterization is not unique. This can be seen from the consistent drift and volatility condition in equations (4.2) and (4.3). Clearly, the solution is not unique, which means that we may construct a different state space system $\{G(z, x), dz(t)\}$. Which one to choose should depend on the specific modeling purpose. Based on the empirical success of the static and dynamic Nelson-Siegel and Svensson family models, we add some structure to $G(z, x)$, such that we can find a consistent system $\{G(z, x), dz(t)\}$ to meet our demand. Empirically, which realization would be better is still an open question. This paper can be a trial on it.

In finding finite dimensional transformations of a Markovian HJM model, there is another line of research, such as Carverhill (1994), Inui and Kijima (1998), Ritchken and Sankarasubramanian (1995), and Chiarella and Kwon (2003). But none of them came up with a model as general as the one proposed here. Following this line, Trolle and Schwartz (2009) came up with a form of volatility that did not have level dependence. Furthermore, all models along this line of research don’t have state variables with the geometric meanings of level, slope, and curvature.

A further point about the choice of $G(z, x)$: models in Diebold and Li (2006), de Pooter, Ravazzolo, and Dijk (2010), and Koopman, Mallee, and Wel (2010) with GARCH are all theoretically inconsistent. Pooter (2007) claimed the use of some consistent forms of
Nelson-Siegel and Svensson family models. But those they claimed are only consistent with a one-factor Hull-White model and they still treated them as multifactor model when dynamizing the models. For this reason, the consistency there is only partial. Moreover, their model could not be used directly in pricing bond derivatives due to their inconsistency. But our model here is consistent and can be directly used to price interest rate derivatives.

Principal components studies suggest that three factors are sufficient to capture the dynamics of the term structure. We will keep our focuses on $n = 3$. In the cases of unspanned stochastic volatility, $n = 3$ means that the model has $n = 3$ forward rate factors. This paper focuses on the general form of stochastic volatility as

$$\sigma_i(t, x) = [\sigma_{i,1} + (\sigma_{i,2} + \sigma_{i,3}x)e^{-a_i x}]\sqrt{v_i(t)}.$$ (5.32)

The corresponding consistent parameterization can be derived from Proposition 5.4.

**Proposition 5.13** Assume a $n$-factor HJM model is defined by equation (5.32) for $i = 1, \ldots, n$. With the consistent conditions for the affine parameterization, we have instantaneous forward rate as

$$f(t, x) = G(z, x)$$

$$= \sum_{i=1}^{n} \begin{pmatrix}
z_{i,1}[\sigma_{i,1} + (\sigma_{i,2} + \sigma_{i,3}x)e^{-a_i x}] + z_{i,2} + z_{i,3}x \\
+z_{i,4}e^{-a_{i,1}x} + z_{i,5}e^{-2a_{i,1}x} + z_{i,6}e^{-a_{i,1}x} + z_{i,7}xe^{-a_{i,1}x} \\
+z_{i,8}xe^{-a_{i,1}x} + z_{i,9}x^2e^{-2a_{i,1}x}
\end{pmatrix},$$ (5.33)

with the dynamics of state variables as

$$\begin{pmatrix}
dz_i(t) \\
dv_i(t)
\end{pmatrix}$$
\[
\begin{pmatrix}
\frac{(a_i \sigma_i^2 + \sigma_i^3)}{a_i^2} v_i(t) - a_i z_{i,1} \\
\sigma_i \sigma_{i,1} z_{i,1} + z_{i,3} \\
\sigma_{i,1}^2 v_i(t) \\
- \frac{\sigma_i (a_i \sigma_i^2 + \sigma_i^3)}{a_i^2} v_i(t) + \sigma_{i,3} z_{i,1} - a_i z_{i,4} + z_{i,6} \\
- \frac{\sigma_i (a_i \sigma_i^2 + \sigma_i^3)}{a_i^2} v_i(t) - 2a_i z_{i,5} + z_{i,7} \\
\sigma_{i,1} (a_i \sigma_i^2 - \sigma_i^3) v_i(t) - a_i z_{i,6} + 2z_{i,8} \\
- \frac{\sigma_i (2a_i \sigma_i^2 + \sigma_i^3)}{a_i^2} v_i(t) - 2a_i z_{i,7} + 2z_{i,9} \\
\sigma_{i,1} \sigma_{i,3} v_i(t) - a_i z_{i,8} \\
- \frac{\sigma_{i,3}^2}{a_i} v_i(t) - 2a_i z_{i,9} \\
\kappa_i (\theta_i - v_i(t))
\end{pmatrix}
\] 
\[dt \]
\[
+ \begin{pmatrix}
\sqrt{v_i(t)} & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\sigma_{i,v} \sqrt{v_i(t)} \rho \sigma_{i,v} \sqrt{v_i(t)} (1 - \rho^2)
\end{pmatrix}
\begin{pmatrix}
dW_i(t) \\
dB_i(t)
\end{pmatrix}
\]

\[5.34\]

and \(x = (T - t), \ z = (z'_1, \ldots, z'_n)\), \(z_i = (z_{i,1}, \ldots, z_{i,9})\). The initial forward curve is \(G(z_0, x) = f(0, x)\) and \(dW_i(t) dB_{j,t} = 0\), for any \(i, j = 1, \ldots, n\).

For \(i = 1\), we have a two-factor model with one forward rate factor and one unspanned stochastic volatility factor. It is trivial that at least two factors are required to obtain a model exhibiting unspanned stochastic volatility. This is also mentioned in Casassus, Collin-Dufresne, and Goldstein (2005). Of course, various extended forms of volatility can be adopted for \(v_i(t)\).

With the multifactor model and under the risk-neutral probability measure, \(Q\), the forward rate curve now has a composite \(LSC\) contributed by each of the forward rate factors, \(z_{i,1}\). With the level being \(\sigma_1 z_1\), slope \(\sigma_2 e^{-ax} z_1\), and curvature \(\sigma_3 x e^{-ax} z_1\), we
have

\[ L^c = \sum_{i=1}^{n} \sigma_i z_{i,1}, \quad (5.35) \]

\[ S^c = \sum_{i=1}^{n} \sigma_i z_{i,1}, \quad (5.36) \]

and

\[ C^c = \sum_{i=1}^{n} \sigma_i z_{i,1}, \quad (5.37) \]

where, \( L^c, S^c, C^c \) denotes the composite \( LSC \) since each factor contributes to all the three components of level, slope, and curvature.

After the change of probability measure from risk-neutral measure, \( Q \), to the physical measure, \( P \), in Section 8, the composite \( LSC \), or \( L^c, S^c, C^c \), will be analyzed under \( P \)-measure.

A further look at the process of the state variables in equation (5.34) shows that the unspanned stochastic volatility, \( v_i(t) \), serves as the kernel of the long-run mean of the forward rate factor, \( z_{i,1} \). This long-run mean is called shadow mean because \( v_i(t) \) is stochastic instead of being a constant like the usual mean of a random variable. Details can be found in Section 9. Specifically, under the risk-neutral probability measure, \( Q \), the shadow mean for \( z_{i,1} \), for \( i = 1, \ldots, n \), is

\[ \text{Mean}_{z_{i,1}} = \frac{(a_i \sigma_i + \sigma_{i,3})}{a_i^2} v_i(t). \]

This is different from traditional affine term structure models, where long-run means of the yield factors are non-stochastic. Furthermore, the unspanned stochastic volatility \( v_i(t) \) in equation (5.34) only has a general specification. Except the constraints for the non-attainment condition, no other requirement is needed. This enables a very flexible specification for it. For example, macroeconomic variables can be added to it as long as they can keep \( v_i(t) \) positive and the whole model is still an admissible affine model.

There is another, maybe more important, meaning for adding macroeconomic variables in the volatility \( v_i(t) \) since the volatility also serves as the long-run means of the yield factors. In this way, the macroeconomic variables play as the fundamentals which decide the long-run means of factors driving the yield curve while keeping themselves out of the yield equation. This extension may contribute to the macro-finance literature in another
angle. We will leave this to further research agenda.

As to the auxiliary state variables, since they are one-period-ahead predictable and don’t bring randomness into the system except helping rule out the arbitrage opportunity, we will not get into them.

The multifactor model derived in Proposition 5.13 brings us the composite level, slope and curvature of the forward curve due to the specification of the general volatility structure. A natural next stop is to propose a model with independent level, slope, and curvature. The Augmented Nelson-Siegel model, as a special case of our general model framework, suggested in the following section, meets this need.

5.3 Consistent Augmented Nelson-Siegel Model

As discussed above, there is no stochastic process which is consistent with the original Nelson-Siegel model. The original Nelson-Siegel model is a linear exponential polynomial function. It is linear in some parameters, which we call state variables after adding dynamics, and it is exponential in time to maturity. Due to its inconsistency, the Nelson-Siegel model cannot be used directly in pricing the interest rate derivatives, even though the dynamic variations of the Nelson-Siegel model outperform the common affine term structure models. This hinders its use in pricing bonds and interest rate derivatives simultaneously.

To overcome this shortcoming, Trolle and Schwartz (2009) developed a three-factor model with unspanned stochastic volatility. Our model is different from Trolle and Schwartz (2009) in several aspects. First, we develop our model based on the consistent conditions in Bjork and Christensen (1999). We simplify their consistent conditions to the affine case as it will be our focus. As shown above, the model construction in this way is more intuitive and simpler than Trolle and Schwartz (2009). For example, the dynamics of state variables in equation (5.34) shows that

\[
\text{Mean}_{\text{shadow}} = \frac{(a_i \sigma_{i,2} + \sigma_{i,3})}{a_i^3} v_i(t)
\]  

(5.38)

can be viewed as the long-term mean of state variable \(z_{i,1}\). We call it shadow mean because \(v_i(t)\) is stochastic instead of being a constant like the usual meaning of mean. This is further discussed in Section 9.
Second, the factors in our model have clear geometric meanings as the level, slope, curvature, while Trolle and Schwartz (2009) model don’t have this property. Our forward rate volatility takes a more general form than Trolle and Schwartz (2009). We include a level component in the volatility while Trolle and Schwartz (2009) only have a slope and curvature component. It is well-known that the volatility of the term structure is stochastic but it is hard to be viewed as completely unspanned, as in Trolle and Schwartz (2009). Third, as shown in the general multifactor model in equation (5.33), the obtained forward rate process is a linear exponential polynomial. We can recast it into an Augmented Nelson-Siegel model by choosing the appropriate volatility structure, which can be found in this part.

Based on the comparison of the Nelson-Siegel model and other popular interest rate models, it can be shown that the original Nelson-Siegel model can be augmented, to the minimum extent, by combining a Ho-Lee model, a Hull-White model, and a Mercurio-Moraleda model, with each model carrying one factor. Here we will develop a three-factor model called Augmented Nelson-Siegel model. The augmented term can be viewed as the compensation of term premium to that of the original Nelson-Siegel model.

5.3.1 Consistent Augmented Nelson-Siegel Model with Gaussian Volatility

As usual, under risk-neutral probability measure $Q$, we define a three-factor HJM model as follows

$$df(t, x) = \mu(t, x)dt + \sum_{i=1}^{3}\sigma_i(t, x)dW_i(t),$$  \hspace{1cm} (5.39)

with volatility structures

$$\sigma_1(t, x) = \sigma_1,$$  \hspace{1cm} (5.40)

$$\sigma_2(t, x) = \sigma_2e^{-a_1x},$$  \hspace{1cm} (5.41)

and

$$\sigma_3(t, x) = \sigma_3xe^{-a_2x},$$  \hspace{1cm} (5.42)

where, $x = T - t$.

As shown in section 5.1, the three volatility components in equations (5.40) through (5.42) are nothing but a Ho-Lee model, a Hull-White model, and a Mercurio-Moraleda
model, respectively. Since Mercurio-Moraledo model (Mercurio and Moraleda (2000)) is a
generalization of Ho-Lee model and Hull-White model, the volatility structure suggested
here can also be seen as a three-factor extension of Mercurio-Moraledo model (Mercurio
and Moraleda (2000)).

Based on Proposition 5.4, we can have a (minimum) consistent parameterization for
each volatility component as follows, respectively,

\[
G(z_1, x) = z_{1,1} + z_{1,2}x,
\]

\[ (5.43) \]

\[
G(z_2, x) = z_{2,1}e^{-a_1x} + z_{2,2}e^{-2a_1x},
\]

\[ (5.44) \]

and

\[
G(z_3, x) = z_{3,1}xe^{-a_2x} + z_{3,2}xe^{-2a_2x} + z_{3,3}xe^{-2a_2x} + z_{3,4}e^{-a_2x} + z_{3,5}e^{-2a_2x}.
\]

\[ (5.45) \]

**Proposition 5.14** For a three-factor HJM model with forward rate process and volatility
structure defined in equations (5.39) through (5.42), the consistent parametrization corre-
sponding to each volatility specification are obtained in equations (5.43) through (5.45).
Accordingly, the forward rate process has the form

\[
f(t, x) = G(z_1, x) + G(z_2, x) + G(z_3, x)
\]

\[ (5.46) \]

\[
= z_{1,1} + z_{1,2}x + z_{2,1}e^{-a_1x} + z_{2,2}e^{-2a_1x} + z_{3,1}xe^{-a_2x} + z_{3,2}xe^{-2a_2x} + z_{3,3}xe^{-2a_2x} + z_{3,4}e^{-a_2x} + z_{3,5}e^{-2a_2x}
\]

\[ (5.47) \]

\[
= (z_{1,1} + z_{2,1}e^{-a_1x} + z_{3,1}xe^{-a_2x}) + (z_{1,2}x + z_{2,2}e^{-2a_1x} + z_{3,2}xe^{-2a_2x} + z_{3,3}xe^{-2a_2x})
\]

\[ (5.48) \]

\[
= NS + AugTerm,
\]

\[ (5.49) \]

with

**Nelson-Siegel model:** \( NS = z_{1,1} + z_{2,1}e^{-a_1x} + z_{3,1}xe^{-a_2x} \)

\[ (5.50) \]

**Augmented Term:** \( AugTerm = z_{1,2}x + z_{2,2}e^{-2a_1x} + z_{3,2}xe^{-2a_2x} + z_{3,3}xe^{-2a_2x} + z_{3,4}e^{-a_2x} + z_{3,5}e^{-2a_2x} \)

\[ (5.51) \]
and the dynamics of state variables as

\[ dz(t) = \begin{pmatrix} dz_1(t) \\ dz_2(t) \\ dz_3(t) \end{pmatrix}, \]

and

\[ dz_1(t) = \begin{pmatrix} dz_{1,1}(t) \\ dz_{1,2}(t) \end{pmatrix} = \begin{pmatrix} z_{1,2}(t) \\ \sigma_1^2 \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} dW_1(t), \tag{5.52} \]

\[ dz_2(t) = \begin{pmatrix} dz_{2,1}(t) \\ dz_{2,2}(t) \end{pmatrix} = \begin{pmatrix} -a_1z_{2,1}(t) + \frac{\sigma_2^2}{a_1} \\ -2a_1z_{2,2}(t) - \frac{\sigma_2^2}{a_1} \end{pmatrix} dt + \begin{pmatrix} \sigma_2 \\ 0 \end{pmatrix} dW_2(t), \tag{5.53} \]

\[ dz_3(t) = \begin{pmatrix} dz_{3,1}(t) \\ dz_{3,2}(t) \\ dz_{3,3}(t) \\ dz_{3,4}(t) \\ dz_{3,5}(t) \end{pmatrix} = \begin{pmatrix} -a_2z_{3,1}(t) + \frac{\sigma_3^2}{a_2} \\ -2a_2z_{3,2}(t) + 2z_{3,3}(t) - \frac{\sigma_3^2}{a_2} \\ -2a_2z_{3,3}(t) - \frac{\sigma_3^2}{a_2} \\ z_{3,1}(t) - a_2z_{3,4}(t) \\ z_{3,2}(t) - 2a_2z_{3,5}(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_3 \\ 0 \\ 0 \\ 0 \end{pmatrix} dW_3(t), \tag{5.54} \]

and \( x = (T - t) \), the initial forward curve \( G(z_0, x) = f(0, x) \), and \( dW_i(t)dW_j(t) = 0 \), for any \( i, j = 1, 2, 3 \).

Following Theorem 4.1, by applying the consistent conditions for the affine parameterization, it is straightforward to verify that the parameterizations in equations (5.43) through (5.45) are consistent with the volatility structures in equations (5.40) through (5.45).
Equations (5.50) and (5.51) show that our three-factor HJM model has one part the same as the original Nelson-Siegel model, (NS), and one augmented term, (AugTerm), which compensates the original Nelson-Siegel model. The augmented term (AugTerm) still remains within the linear exponential polynomial family. The dynamics of state variables, $d z_{1,1}(t)$, $d z_{2,1}(t)$, and $d z_{3,1}(t)$, constitute the state variables of the original Nelson-Siegel model. They are the only state variables which bear instantaneous randomness. The augmented term (AugTerm) has no instantaneous randomness and are one-period ahead predictable. The model derived here still has an affine form in the vector of the state variable $Z(t)$. The usual affine term structure models have the same number of state variables as that of risk factors. In contrast to them, the model here has more state variables than risk factors. The state variables constitute the augmented term (AugTerm) don’t bear randomness and thus no risk premium on them.

As a special case to our consistent Augmented Nelson-Siegel model in Proposition (5.14), the volatility structure in equations (5.41) and (5.42) can be set at $a_1 = a_2$. This will reduce the dimension of the state variables by two. Similar specifications to this special case has been studied in Diebold and Li (2006) and Christensen, Diebold, and Rudebush (2010). However, Diebold and Li (2006) keep the original Nelson-Siegel model while adding an arbitrary $AR(1)$ process for state variables. Christensen, Diebold, and Rudebush (2010) followed the common short rate framework derived a three-factor augmented Nelson-Siegel model with the augmented term being time-homogeneous. Unlike Christensen, Diebold, and Rudebush (2010), the AugTerm in our HJM framework is time-varying.

Under the risk-neutral probability measure, we specify a three-factor HJM model with the forward rate process in equation (5.39) and the volatility structure as

$$
\sigma_0(t, x) = \sigma_1 + \sigma_2 e^{-ax} + \sigma_3 x e^{-ax},
$$

where $x = T - t$.

Here the exponential decaying speed is the same for both the slope and curvature components. In a similar way as above, we can find a (minimum) consistent parameterization
for the volatility structure in equation (5.55) as

\[
G(z, x) = z_{1,1} + z_{1,2}x + z_{2,1}e^{-ax} + z_{2,2}e^{-2ax} \\
+ z_{3,1}xe^{-ar} + z_{3,2}xe^{-2ar} + z_{3,3}x^2 e^{-2ar}.
\] (5.56)

Comparing the parameterizations in equation (5.56) with equations (5.43) through (5.45), we find the parameterization in equation (5.56) has two state variables less. This is because, when the decaying speed is set the same for the slope and curvature components, two terms in equation (5.45), \(z_{3,4}e^{-ax} + z_{3,5}e^{-2ax}\), now can be covered by equation (5.44).

**Proposition 5.15** For a three-factor HJM model with the forward rate process defined in equation (5.39) and the volatility structure in equation (5.55), a consistent parametrization can be derived as equation (5.56). Then the forward rate process has the form

\[
f(t, x) = z_{1,1} + z_{1,2}x + z_{2,1}e^{-ax} + z_{2,2}e^{-2ax} \\
+ z_{3,1}xe^{-ar} + z_{3,2}xe^{-2ar} + z_{3,3}x^2 e^{-2ar} \\
= (z_{1,1} + z_{2,1}e^{-ax} + z_{3,1}xe^{-ar}) \\
+ (z_{1,2}x + z_{2,2}e^{-2ax} + z_{3,2}xe^{-2ar} + z_{3,3}x^2 e^{-2ar}) \\
= NS + AugTerm,
\] (5.57)

with

\[
NS = z_{1,1} + z_{2,1}e^{-ax} + z_{3,1}xe^{-ax}
\]

and

\[
AugTerm = z_{1,2}x + z_{2,2}e^{-2ax} + z_{3,2}xe^{-2ax} + z_{3,3}x^2 e^{-2ax}.
\] (5.58)

The state variable process is

\[
dz(t) = \begin{pmatrix} dz_1(t) \\ dz_2(t) \\ dz_3(t) \end{pmatrix}
\]

with

\[
dz_1(t) = \begin{pmatrix} dz_{1,1}(t) \\ dz_{1,2}(t) \end{pmatrix} = \begin{pmatrix} z_{1,2}(t) \\ \sigma_1^2 \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \\ 0 \end{pmatrix} dW_1(t),
\] (5.59)
\[
dz_2(t) = \begin{pmatrix}
dz_{2,1}(t) \\
dz_{2,2}(t)
\end{pmatrix} = \begin{pmatrix}
-a z_{2,1}(t) + z_{3,1}(t) + \frac{\sigma_2^2}{a} \\
-2 a z_{2,2}(t) + z_{3,2}(t) - \frac{\sigma_2^2}{a}
\end{pmatrix} dt + \begin{pmatrix}
\sigma_2 \\
0
\end{pmatrix} dW_2(t),
\]

and

\[
dz_3(t) = \begin{pmatrix}
dz_{3,1}(t) \\
dz_{3,2}(t) \\
dz_{3,3}(t)
\end{pmatrix} = \begin{pmatrix}
-a z_{3,1}(t) + \frac{\sigma_3^2}{a^2} \\
-2 a z_{3,2}(t) + 2 z_{3,3}(t) - \frac{\sigma_3^2}{a^2} \\
-2 a z_{3,3}(t) - \frac{\sigma_3^2}{a}
\end{pmatrix} dt + \begin{pmatrix}
\sigma_3 \\
0 \\
0
\end{pmatrix} dW_3(t),
\]

where, \( x = (T - t) \), the initial forward curve is \( f(0, x) \), and \( dW_i(t)dW_j(t) = 0 \), for any \( i, j = 1, 2, 3 \).

Compared with Proposition 5.14, the state space dimension in Proposition 5.15 is two less and the state variable \( dz_2(t) \) has a different form since here it covers two extra terms from \( dz_3(t) \).

### 5.3.2 Consistent Augmented Nelson-Siegel Model with Unspanned Stochastic Volatility

As shown in previous section, we are ready to extend the consistent Augmented Nelson-Siegel model to incorporate unspanned stochastic volatility. Under the risk-neutral probability \( Q \), we define a three-factor forward rate curve as equation (5.39),

\[
df(t, x) = \mu(t, T) dt + \sum_{i=1}^{3} \sigma_i(t, x) dW_i(t).
\]

The volatility structure is extended to introduce the unspanned stochastic volatility as

\[
\sigma_1(t, x) = \sigma_1 \sqrt{v_1(t)},
\]

\[
\sigma_2(t, x) = \sigma_2 e^{-a_1 x} \sqrt{v_2(t)},
\]
and

$$\sigma_3(t, x) = \sigma_3 x e^{-a_2 x} \sqrt{v_3(t)}, \quad (5.65)$$

with

$$dv_i(t) = \kappa_i(\theta_i - v_i(t))dt + \sigma_{v_i} \sqrt{v_i(t)}(\rho_i dW_i(t) + \sqrt{1 - \rho_i^2} dB_i(t)), \quad (5.66)$$

where $dW_i(t)dB_j(t) = 0$, for any $i, j = 1, 2, 3$.

As noticed, it is straightforward to extend the unspanned stochastic volatility to more flexible specifications as shown in the previous sections like $h_i = \lambda_{i,1} f(t, T_x) + \lambda_{i,2} v_i(t)$. Later on in the empirical analysis, we will focus on the volatility form in equation (5.66).

Based on the previous analysis, the parameterizations in equations (5.43) through (5.45) are the (minimum) consistent parameterizations with volatility structures in equations (5.63) through (5.65).

**Proposition 5.16** A three-factor HJM model is defined with the forward rate process in equation (5.62) and the volatility structures in equations (5.63) through (5.66). we can have a consistent parametrization as

$$f(t, x) = (z_{1,1} + z_{2,1} e^{-a_1 x} + z_{3,1} x e^{-a_2 x})$$

$$+ (z_{1,2} x + z_{2,2} e^{-2a_1 x} + z_{3,2} x e^{-a_2 x} + z_{3,3} x^2 e^{-2a_2 x} + z_{3,4} e^{-a_2 x} + z_{3,5} e^{-2a_2 x})$$

$$= NS + AugTerm, \quad (5.67)$$

with $NS$ and $AugTerm$ defined in equations (5.50) and (5.51), respectively. The dynamics of state variables follow

$$dz_1(t) = \begin{pmatrix} dz_{1,1}(t) \\ dz_{1,2}(t) \end{pmatrix} = \begin{pmatrix} z_{1,2}(t) \\ \sigma_1^2 v_1(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \sqrt{v_1(t)} \\ 0 \end{pmatrix} dW_1(t), \quad (5.68)$$

$$dz_2(t) = \begin{pmatrix} dz_{2,1}(t) \\ dz_{2,2}(t) \end{pmatrix}$$

$$= \begin{pmatrix} -a_1 z_{2,1}(t) + \frac{a_2^2}{a_1} v_2(t) \\ -2a_1 z_{2,2}(t) - \frac{a_2^2}{a_1} v_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_2 \sqrt{v_2(t)} \\ 0 \end{pmatrix} dW_2(t), \quad (5.69)$$
\[
dz_3(t) = \begin{pmatrix}
dz_{3,1}(t) \\
dz_{3,2}(t) \\
dz_{3,3}(t) \\
dz_{3,4}(t) \\
dz_{3,5}(t)
d\end{pmatrix}
\begin{pmatrix}
-a_2 z_{3,1}(t) + \frac{\sigma_3^2}{a_2} v_3(t) \\
-2a_2 z_{3,2}(t) + 2z_{3,3}(t) - \frac{\sigma_3^2}{a_2} v_3(t) \\
-2a_2 z_{3,3}(t) - \frac{\sigma_3^2}{a_2} v_3(t) \\
z_{3,1}(t) - a_2 z_{3,4}(t) \\
z_{3,2}(t) - 2a_2 z_{3,5}(t)
\end{pmatrix}
dt + \sigma_3 \sqrt{v_3(t)}dW_3(t),
\]

(5.70)

with \( x = (T - t) \), the initial forward curve \( f(0, x) \), and \( dW_i(t)dB_j(t) = 0 \), for any \( i, j = 1, 2, 3 \). The stochastic volatility dynamics, \( v_i(t) \), is defined in equation (5.66).

Just like in Proposition 5.15, where we obtain an even more parsimonious Gaussian model when set the decaying speed \( a_1 = a_2 \), we can have a even more parsimonious model with unspanned stochastic volatility. Similarly, the dimension of the state variables \( dZ_3(t) \) can be reduced by two. Refer to Appendix A.1 for details.

Here the three independent forward rate factors, \( z_{1,1}, z_{2,1}, \) and \( z_{3,1} \) contribute to the level, slope and curvature, respectively and independently. To be more specific, we have

\[
L^i = z_{1,1},
\]

(5.71)

\[
S^i = z_{2,1},
\]

(5.72)

and

\[
C^i = z_{3,1}.
\]

(5.73)

Compared with the composite level, slope and curvature in equations (5.35) through (5.37), the independent level, slope and curvature above only have one dedicated factor. The state variable processes in equations (5.68) through (5.70) are different from those in the model with general volatility structure in Proposition 5.13. We will look through them one by one. The level factor \( z_{1,1} \) in equation (5.68) is not a usual mean reversion
process. After the change of probability measure from the risk-neutral $Q$ to the physical $P$ measure, with the introduction of the market price of risk, $z_{1,1}$ will become a mean reversion process and the long run mean under $P$-measure will be discussed in Section 8 and Appendix. For now we can show that the slope and curvature factors, under $Q$-measure, have their unspanned stochastic volatility factors as their long-run means. For the slope factor, $z_{2,1}$, it has a long-run mean, or shadow mean as

$$\text{Mean}_{z_{2,1}}^{\text{shadow}} = \sigma^2_2 \frac{a_2}{a_1^2} v_2(t), \quad (5.74)$$

while the curvature factor, $z_{3,1}$, has a long-run mean, or shadow mean as

$$\text{Mean}_{z_{3,1}}^{\text{shadow}} = \sigma^2_3 \frac{a_3}{a_2^2} v_3(t). \quad (5.75)$$

As is shown above, the stochastic volatilities also play their roles as the kernels of the shadow means for the slope and curvature factors and enter in the conditional mean of the level factor.

## 6 Unspannedness of Risk Factors: a Further Note

For the interest of unspanned factors, there are two lines of research. One is along Collin-Dufresne and Goldstein (2003), Heidari and Wu (2009), and Troll and Schwartz (2009), where the unobservable volatility factors are unobservable. The bond market is incomplete in itself, but the combination of the bond market and derivative market forms a complete market. In such a manner, the risks unspanned by bond market can be spanned with derivatives. The unspannedness is incorporated into the modeling scheme through the specifications of the underlying process, the dynamics of the forward rate process or state variables (similar as short rate process).

Another line of research is along Joslin, Priebsch, and Singleton (2014), they tried to explain the unspanned macroeconomic factors in bond market. Through the specification of market price of risk, they made state variables, those entering the pricing formula, dependent on the unspanned macroeconomic factors under the physical measure. But, under the risk-neutral measure, the state variables in the pricing formula don’t have macroeconomic factors involved. It means that the bond pricing formula is exactly the
same as without the unspanned macro factors. We can infer that, if we extend the model to price derivatives, the pricing formula of derivatives will be the same as that without those unspanned macroeconomic factors since the unspanned factors are only included in the dynamics, under the physical measure, of the full-set of the state variables. Of course, they can also be used to analyze the effects of macroeconomic factors on the derivative market. Since there is no way to hedge the macro risks, the economy-wide market in Joslin, Priebsch, and Singleton (2014) is incomplete.

Even though the above discussion on the spanned and unspanned stochastic volatility factors are commonly thought to be unobserved state variables, the specification of the volatility structure, $h(t)$, can be extended to include some observable variables such as macroeconomic variables. This extension can be achieved because the forward rate process doesn’t depend on the specification of the volatility factors. That is to say, the specification of the volatility factors is arbitrary. It facilitates the model framework discussed in this paper to be extended to macro-finance models. Ang and Piazzesi (2003), Rudebusch and Wu (2008), and Bikbov and Chernov (2010) covered spanned macroeconomic factors and Joslin, Priebsch, and Singleton (2014) suggested unspanned macro factors through the specification of the market price of risk. But none of them starts with the HJM framework.

This differs from the models with unspanned stochastic volatility in our paper, where the unspanned factors are introduced through the volatility structure of the forward rate. Joslin, Priebsch, and Singleton (2014) created an economy-wide incomplete market and no asset can span the macro factors. It seems that if we want to introduce unspanned macro factors into model, this way makes sense. It means that, even if the derivative market is included, the macro factors still remain unspanned. If the macro factors are introduced only as bond-unspanned stochastic volatility and spanned by the derivative market, it seems a little weird. But it still worth trying to see if the derivative market can span some macroeconomic risk factors that the bond market can not.

7 Pricing Zero-coupon Bonds

According to the identities discussed in Section 2.1, We can derive bond pricing function through the forward rate models we develop above.
7.1 Bond Price in n-factor HJM Model with USV

**Proposition 7.1** For a n-factor HJM model define in Proposition 5.13 with the state variables, \(dZ(t)\), and the stochastic volatility dynamics, \(dv_i(t)\), in equation (5.66), the time \(t\) price of a zero-coupon bond maturing at \(T\), \(P(t,T)\), is given as

\[
P(t,T) = \exp \left\{ \sum_{i=1}^{n} \sum_{j=1}^{9} b_{i,j}(T-t)z_{i,j}(t) \right\},
\]

where,

\[
b_{i,1}(x) = \frac{1}{a_i^2} \left\{[\sigma_{i,3} + a_i(\sigma_{i,2} + \sigma_{i,3}x)]e^{-a_ix} - a_i\sigma_{i,2} - \sigma_{i,3}\right\} - \sigma_{i,1}x,
\]

\[
b_{i,2}(x) = -x,
\]

\[
b_{i,3}(x) = -\frac{1}{2}x^2,
\]

\[
b_{i,4}(x) = \frac{1}{a_i}(e^{-a_ix} - 1),
\]

\[
b_{i,5}(x) = \frac{1}{2a_i}(e^{-2a_ix} - 1),
\]

\[
b_{i,6}(x) = \frac{1}{a_i^2}[(a_i x + 1)e^{-a_ix} - 1],
\]

\[
b_{i,7}(x) = \frac{1}{4a_i^2}[(2a_i x + 1)e^{-2a_ix} - 1],
\]

\[
b_{i,8}(x) = \frac{1}{a_i^3}[(a_i x + 1)^2e^{-a_ix} + e^{-a_ix} - 2],
\]

\[
b_{i,9}(x) = \frac{1}{4a_i^3}[(a_i x + 1)^2e^{-2a_ix} + a_i^2x^2e^{-2a_ix} - 1].
\]

**Proof.**

\[
P(t,T) = \exp \left( - \int_{t}^{T} f(t,s)ds \right)
\]

\[
= \exp(-\int_{0}^{T} G(z,u)du)
\]

\[
= \exp \left\{ \sum_{i=1}^{n} \sum_{j=1}^{9} b_{i,j}(T-t)z_{i,j}(t) \right\},
\]

with

\[
G(z,x) = [\sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax}]z_1 + z_2 + z_3 x + z_4 e^{-ax}
\]

53
\[ z_5 e^{-2ax} + z_6 xe^{-ax} + z_7 xe^{-2ax} + z_8 x^2 e^{-ax} + z_9 x^2 e^{-2ax}. \]

Take integral with respect to the coefficients of \( z_s \) in \( G(z, x) \), we obtain equations (7.2) through (7.10). ■ To facilitate pricing the interest rate derivatives, the bond price dynamics for a n-factor HJM model can be written as

\[
\frac{dP(t, T)}{P(t, T)} = r(t) dt + \sum_{i=1}^{n} b_{i,1}(x) \sqrt{v_i(t)} dW_i(t). \tag{7.11}
\]

Equivalently, the yields for the n-factor HJM model above can be found as

\[
y(t, x) = \sum_{i=1}^{n} \sum_{j=1}^{9} \frac{b_j(x)}{x} z_{i,j}(t) \tag{7.12}
\]

In the empirical analysis, the bond yields formula in equation (7.12) is used as the measurement equation when implementing the Kalman filter.

With this proposition, we can derive the bond prices for our consistent Augmented Nelson-Siegel models. As shown above, the difference between the consistent Augmented Nelson-Siegel model with Gaussian volatility and stochastic volatility lies only in the dynamics of state variables, or the volatility structure of the state variable dynamics, to be more specific. By comparing Proposition 5.4 and Proposition 5.11, we can see that three volatility structures share the same deterministic component and differ in the stochastic volatility components. All three models share the same specification of the forward rate process and almost-same state variable dynamics. This property, thus, actually gives us a very rich specification for the model framework we develop here. With different stochastic volatility structures, we can have different forward rate models, which only differ in the dynamics of the state variable process, but share the same forward rate specification. So, we can remain in the same consistent Augmented Nelson-Siegel model while change the specification of the stochastic volatility structures. This property can help relate the unobserved stochastic volatility to some observable variables, such as macroeconomic variables and macroeconomic news announcements.
7.2 Bond Price in Consistent Augmented Nelson-Siegel Model with USV

Here, we derive the bond price formula for the consistent Augmented Nelson-Siegel model with unspanned stochastic volatility.

**Proposition 7.2** For a three-factor consistent Augmented Nelson-Siegel model with unspanned stochastic volatility defined in Proposition 5.16, the time $t$ price of a zero-coupon bond maturing at $T$, $P(t,T)$, is given as

$$P(t,T) = \exp \left\{ \sum_{i=1}^{3} \sum_{j=1}^{5} b_{i,j} (T - t) z_{i,j}(t) \right\},$$

(7.13)

where,

- $b_{1,1}(x) = -x$, \hspace{1cm} (7.14)
- $b_{2,1}(x) = \frac{1}{a_1} (e^{-a_1 x} - 1)$, \hspace{1cm} (7.15)
- $b_{3,1} = \frac{1}{a_2} \left( (a_2 x + 1) e^{-a_2 x} - 1 \right)$, \hspace{1cm} (7.16)
- $b_{1,2} = -\frac{1}{2} x^2$, \hspace{1cm} (7.17)
- $b_{2,2} = \frac{1}{2a_1} (e^{-2a_1 x} - 1)$, \hspace{1cm} (7.18)
- $b_{3,2} = \frac{1}{4a_2^2} \left( (2a_2 x + 1) e^{-2a_2 x} - 1 \right)$, \hspace{1cm} (7.19)
- $b_{3,3} = \frac{1}{2a_2^2} \left( (a_2 x^2 + x + 1) e^{-2a_2 x} + 1 \right)$, \hspace{1cm} (7.20)
- $b_{3,4} = \frac{1}{a_2} (e^{-a_2 x} - 1)$, \hspace{1cm} (7.21)
- $b_{3,5} = \frac{1}{2a_2} (e^{-2a_2 x} - 1)$, \hspace{1cm} (7.22)

with $x = T - t$, $b_{i,j} = 0$, for $i = 1, 2$, and $j \geq 3$.

**Proof.** In a similar way to the proof of Proposition 7.1, This can be proved. □
Equivalent to the bond price in equation (7.13), for a three-factor consistent Aug-
mplemented Nelson-Siegel model, bond price dynamics can be written as
\[
\frac{dP(t, T)}{P(t, T)} = r(t)dt + \sum_{i=1}^{3} b_{i,1}(x)\sqrt{v_i(t)}dW(t). \tag{7.23}
\]

Then, the bond yield for the three-factor consistent Augmented Nelson-Siegel model
follows
\[
y(t, x) = \sum_{i=1}^{3} \sum_{j=1}^{5} \frac{b_j(x)}{x} z_{i,j}(t) \tag{7.24}
\]

The yield formula in equation (7.24) is used as the measurement equation in our
empirical analysis.

8 Market Price of Risk

Up until now, all the models we have discussed are under the risk-neutral probability
measure. For pricing purposes, the risk-neutral probability measure is adequate. However,
to implement econometric estimation, we have to put all processes under the equivalent
physical probability measure. The links between the two measures are the specification
of the market price of risks.

With the development of affine term structure models, the specifications of the market
price of risk also change. There are different ways to specify the risk premium function.
Recent literature suggests four main subclasses of the affine term structure models.
Depending on the specification of the market price of risk, there are completely affine,
semi-affine, essentially affine, and extended affine models.

For illustration purposes, we will briefly compare these specifications based on the
canonical representation of an affine model as in Dai and Singleton (2000). For the
extended form of the market price of risk, which is adopted in this paper, we will give
the form specific to our model. For an admissible affine process under the risk-neutral
probability measure,
\[
dz(t) = \kappa(\theta - z(t)) + \Sigma\sqrt{S(t)}dW(t), \tag{8.1}
\]
where, \( z(t) \in \mathbb{R}^n \), \( \kappa \) is a \( n \times n \) matrix, \( \theta \) is a \( n \) vector, \( \Sigma \) is a \( n \times n \) matrix, \( S(t) \) is a \( n \times n \)
diagonal matrix with \( S_{ii}(t) = \alpha_i + \beta_i z(t) \), and \( \alpha_i \) and \( \beta_i \) are \( n \)-vectors.
8.1 Completely Affine Model

In completely affine models, the market price of risk $\Lambda_t$ is specified proportional to the volatility $\sqrt{S(t)}$, or as

$$\Lambda_t = \sqrt{S(t)} \lambda,$$  \hspace{1cm} (8.2)

where $\lambda$ is a $n$-vector.

Historically, completely affine term structure models are the oldest and simplest kinds. This form of risk premia is well suited in order to nest many other models, e.g., multifactor versions of Vasicek (1977) and Cox, Ingersoll, and Ross (1985b) (CIR) models. In completely affine models, everything is affine: the SDE under the physical measure, the one under the risk-neutral probability measure, the short rate process, and the logarithm of the zero coupon bond price. The market price of risk in completely affine models is proportional to instantaneous volatility $\sqrt{S(t)}$. It increases with the volatility and when the volatility reaches zero, so does the market price of risk. This property is consistent with no-arbitrage theory. Besides this favorable property, the sign of the market price of risk never change since $S(t)$ is positive (or, at least, non-negative). Duffee (2002), however, suggested that the implied prices of risk switch signs over time. The name “completely affine” came from the fact that $\Lambda_t^t \Lambda_t$ is also affine in $X(t)$. The completely affine model is an oversimplification of real world.

8.2 Essentially Affine Model

Another family of affine models is the essentially affine models which are introduced by Duffee (2002). In these models, the market price of risk is specified as

$$\Lambda_t = \sqrt{S(t)} \lambda_1 + \sqrt{S(t)}^-\lambda_2 z(t),$$  \hspace{1cm} (8.3)

where $\lambda_1$ is a $n$-vector and $\lambda_2$ is a $n \times n$ matrix, and

$$\sqrt{S_{ii}(t)}^- = \begin{cases} \frac{1}{\sqrt{\alpha_i + \beta_i z(t)}} & \text{if } \inf(\alpha_i + \beta_i z(t)) > 0, \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (8.4)

This specification of the market price of risk in essentially affine models nests com-
pletely affine models. It extends the completely affine models by making the price of risk depend on the full set of the state variables instead of just instantaneous volatilities, $\sqrt{S(t)}$. The price of risk is now not completely proportional to the instantaneous volatility any more. The drift remains affine under both the physical and risk-neutral probability measures. The sign limitation is partially relaxed only when $\inf(\alpha_i + \beta_i z(t)) > 0$, otherwise $\sqrt{S_{ii}(t)} = 0$. Then, the essentially affine model boils down to a completely affine model. Since $\Lambda_t^\prime \Lambda_t$ is not affine in $z(t)$ any more, it is termed “essentially affine” model. It nests the completely affine model at $\lambda_2 = 0$.

8.3 Semi-affine Model

Duarte (2004) suggested a semi-affine market price of risk to overcome some limitations of the completely affine models. By adding a constant, Duarte (2004) extended the proportional specification of the market price of risk in completely affine models. It is defined as

$$\Lambda_t = \lambda_0 + \sqrt{S(t)}\lambda_1 + \sqrt{S(t)}\lambda_2 z(t),$$

(8.5)

where $\lambda_0$ and $\lambda_1$ are both $n$-vectors and $\lambda_2$ is a $n \times n$ matrix. It is easy to see that, when the market price of risk is defined in equation (8.5), the drift under the physical probability measure is not affine in $z(t)$ any more. This is why it is called a semi-affine model. Duarte (2004) found that the semi-affine model can well accommodate the sign switches of the risk premia. Unfortunately, even though the semi-affine model can show sign changes in the risk premia, it describes poorly the dynamics of risk premia. The signs are correct but the size is too small.

Seen from above, in completely affine models, the market price of risk of each state variable only depends on the state variables in its own volatility. For those state variables missing in the volatility, they won’t directly affect the market price of risk associated with those state variables. On the contrary, essentially affine and semi-affine models can have the market price of risk depend on the full set of the state variables, unless $\sqrt{S_{ii}(t)} = 0$.

8.4 Extended Affine Model

Essentially affine and semi-affine models have part of the market price of risk, $\sqrt{S(t)}$, inversely proportional to instantaneous volatility and force it to take zero when it attains
boundary zero from below. The unpopularity of the market price of risk inversely proportional to instantaneous volatility $\sqrt{S(t)}$ is that when the volatility reaches zero, the market price of risk becomes infinite. Cox, Ingersoll, and Ross (1985b) suggested that this lead to an arbitrage opportunity. Essentially and semi-affine models explicitly avoid this issue by setting zero the part inversely proportional to the instantaneous volatility.

Cheridito, Filipovic, and Kimmel (2007) proposed a general form of the market price of risk inversely proportional to instantaneous volatility. Their extended affine model nests the completely and essentially affine models. As a more general specification, we adopt their form of the market price of risk in this paper. The authors show that when the instantaneous volatility moves arbitrarily close to zero, the inversely-proportional market price of risk grows arbitrarily large. Thus, the arbitrage opportunity can still be ruled out by imposing some constraints on the parameters. For illustration purposes, we list here their model $A_1(2)$ as

$$
d\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = \begin{pmatrix} a_1^P \\ 0 \end{pmatrix} + \begin{pmatrix} b_{11}^P & 0 \\ b_{21}^P & b_{22}^P \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{z_1(t)} & 0 \\ 0 & \sqrt{\alpha_2 + \beta_2 z_1(t)} \end{pmatrix} \begin{pmatrix} dW_1^P(t) \\ dW_2^P(t) \end{pmatrix}, \tag{8.6}$$

$$
d\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = \begin{pmatrix} a_1^Q \\ 0 \end{pmatrix} + \begin{pmatrix} b_{11}^Q & 0 \\ b_{21}^Q & b_{22}^Q \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{z_1(t)} & 0 \\ 0 & \sqrt{\alpha_2 + \beta_2 z_1(t)} \end{pmatrix} \begin{pmatrix} dW_1^Q(t) \\ dW_2^Q(t) \end{pmatrix}. \tag{8.7}$$

These are the state variable processes under two probability measures, where $P$ and $Q$ correspond to the physical and risk-neutral probability measure, respectively. Their extended affine market price of risk is defined as

$$\Lambda_t = \begin{pmatrix} \lambda_{00} + \lambda_{11} z_1(t) \\ \sqrt{z_1(t)} \end{pmatrix} \begin{pmatrix} \lambda_{20} + \lambda_{21} z_1(t) + \lambda_{22} z_2(t) \\ \sqrt{\alpha_2 + \beta_2 z_1(t)} \end{pmatrix}, \tag{8.8}$$
with non-attainment condition
\[ \lambda_{10} \leq a_1^P - \frac{1}{2}. \] (8.9)

As stated in Cheridito, Filipovic, and Kimmel (2007), when \( \beta_{21}\lambda_{20} = \lambda_{21}\alpha_2 \), the extended affine model becomes a completely affine model. While \( \lambda_{10} = \lambda_{20} = 0 \) yields an essentially affine model.

### 8.5 Market Price of Risk for Our Models


Since we only focus on the bond market, which is incomplete with unspanned stochastic volatility, we cannot estimate the parameters in the volatility specification under \( Q \)-measure. Those parameters can only be estimated when the market is complete or when the interest rate derivatives are included. For now, we only specify the unspanned stochastic volatility under \( P \)-measure. With our model specification in previous section, following Cheridito, Filipovic, and Kimmel (2007) and Trolle and Schwartz (2009), we define our market price of risk \( \Lambda_{W_i}(t) \), for \( i = 1, \ldots, n \), as

\[
dW_i(t) = dW_i^P(t) + \Lambda_{W_i}(t)dt, \tag{8.10}
\]

where we suppress the superscript \( Q \) for the risk-neutral probability measure to keep consistent with the previous sections, and

\[
\Lambda_{W_i}(t) = \frac{\lambda_{W_i,0} + \lambda_{W_i,z}z_{i,1}(t) + \lambda_{W_i,v}v_i(t)}{\sqrt{v_i(t)}}. \tag{8.11}
\]

The state variable process \( z_{i,1} \) in equation (5.34), under the physical probability measure, becomes

\[
dz_{i,1}(t) = (\theta_{z_{i,1}}^P + \kappa_{z_{i,1}}^Pz_{i,1}(t) + \kappa_{z_{i,1,v}}^Pv_i(t))dt + \sqrt{v_i(t)}dW_i^P(t), \tag{8.12}
\]
θ_{z_{i,1}}^P = \lambda_{W_{i,0}}, \quad (8.13)

κ_{z_{i,1}}^P = \lambda_{W_{i,z}} - a_i, \quad (8.14)

and

κ_{z_{i,1,v}}^P = \frac{(a_i \sigma_{i,2} + \sigma_{i,3})}{a_i^2} + \lambda_{W_{i,v}}. \quad (8.15)

Since the bond market alone is incomplete, the volatility process under the risk-neutral probability measure cannot be estimated unless the derivative market is included. Therefore, the volatility dynamics under the physical probability measure is specified as

\[ dv_i(t) = \kappa_{\theta_i}^P (\theta_i - v_i(t)) dt + \sigma_{\theta_i} \sqrt{v_i(t)} \left( \rho_{i,v} dW_i^P(t) + \sqrt{1 - \rho_{i,v}^2} dB_i^P(t) \right), \quad (8.16) \]

In the estimation, \( \theta_{v_i}^P \) is normalized to 1 in order to keep the scaling term \( \sigma_{v_i} \) as a free parameter. \( \theta_{v_i}^P \) is the long-run mean of \( v_i(t) \) and \( \sigma_{v_i} \) is the scaling term for \( v_i(t) \). They cannot be identified due to the unobservability of \( v_i(t) \). In short rate models, \( \sigma_{v_i} \) is usually normalized to 1. Since the forward rate volatility is of our interest, we choose to normalize \( \theta_{v_i}^P \) to 1 and keep \( \sigma_{v_i} \) free. For details, please refer to Dai and Singleton (2000).

In the specifications of the market price of risk, equations (8.10) and (8.11), we introduce five parameters. Now, we transform them into five new parameters under the physical measure since we are more interested in the dynamics of state variables under the physical measure. The transformation relations are shown in equations (8.13) through (8.15). As the state process under the physical measure will be used in empirical analysis, we will focus on the transformed parameters instead of the parameters under the risk-neutral measure. The non-attainment restraints for \( v_i(t) \) not reaching zero, under physical measure is

\[ \kappa_{v_i}^P \geq \frac{\sigma_{v_i}^2}{2}. \quad (8.17) \]

Since \( z_{i,j}(t) \) for all \( i \) and \( j \geq 2 \) don’t have contemporaneous randomness, there is no risk premium associated with these state variables. As stated in previous section, this is one property in which our model differs from most of the traditional affine term structure models.

If we set constants \( \lambda_{W_{i,0}} \) and the cross-dependence parameters \( \lambda_{W_{i,z}} \) as zero, our
model boils down to a completely affine model. When we only set $\lambda_{W,i,0}$ as zero, the model becomes an essentially affine model.

Another point worth mentioning is that Trolle and Schwartz (2009) claimed their market price of risk is adopted from Cheridito, Filipovic, and Kimmel (2007). But their adoption, just like ours here, is a simplified version of Cheridito, Filipovic, and Kimmel (2007)’s market price of risk. For the $n$-factor model, the fully-adopted extended affine market price of risk should involve all $v_i(t)$ and $z_{i,1}(t)$ for each $\Lambda_{W}(t)$. What we did here is to treat each pair, $(dW_i^p(t), dB_i^p(t))$, as a group then stack all $n$ pairs. In this way, we exclude the interdependence between $(v_i(t), z_{j,1}(t))$ for $i \neq j$.

A further look at equation (8.12) shows us more meanings of the unspanned stochastic volatility. Now rearrange equation (8.12) as

$$dz_{i,1}(t) = \left( -\kappa_{z_{i,1}}^P \left( \frac{\theta_{z_{i,1}}^P + \kappa_{z_{i,1},v}^P v_i(t)}{-\kappa_{z_{i,1}}^P} - z_{i,1}(t) \right) \right) dt + \sqrt{v_i(t)}dW_i^P(t).$$

(8.18)

We can interpret that $z_{i,1}$ has a mean-reversion speed of $(-\kappa_{z_{i,1}}^P)$, a long-run central tendency of

$$Mean_{z_{i,1}}^{\text{shadow}} = \frac{\theta_{z_{i,1}}^P + \kappa_{z_{i,1},v}^P v_i(t)}{-\kappa_{z_{i,1}}^P},$$

(8.19)

and the volatility of $v_i(t)$. We call this long-run central tendency the shadow mean of $z_{i,1}$ because $v_i(t)$ is stochastic instead of being a constant. This interpretation tells that the unspanned stochastic volatility serves as the kernel of the central tendency of $z_{i,1}$.

Details is discussed in Section 9.

9 Empirical Analysis

In this section, various model specifications discussed previously are tested with two applications. The first application uses the U.S. Treasury yields, while the second application deals with USD LIBOR and swap rates. We first describe the data sets and then discuss how they are processed before estimation. The models used in both applications are cast into a state space form, which facilitates the use of Kalman filtration. The application to the U.S. Treasury data is estimated with the conventional Kalman filter. Since the LIBOR and swap rates are all nonlinear functions of the state variables. A brief introduction is provided about the extended Kalman filter which is widely used
in the estimation of nonlinear dynamic systems.

9.1 Discretization of State Process

All the models discussed so far are in continuous-time while the market data are observed at discrete-time intervals. To implement econometric analysis we need to work with the discrete-time dynamics of the state variables implied by its continuous-time counterpart. Take the Euler scheme for an instance. For the state variable processes under the physical measure in equation (8.12) and (8.16)

\[ dz_{i,1}(t) = (\theta^P_{z_{i,1}} + \kappa^P_{z_{i,1}} z_{i,1}(t) + \kappa^P_{z_{i,1},v_i}(t))dt + \sqrt{v_i(t)}dW^P_i(t), \tag{9.1} \]

and

\[ dv_i(t) = \kappa^P_{v_i}(\theta^P_{v_i} - v_i(t))dt + \sigma^P_{v_i}\sqrt{v_i(t)}dW^P_i(t) + (1 - \rho^2_i)dB^P_i(t), \tag{9.2} \]

their discrete-time counterparts are

\[ z_{i,1}(t + \Delta t) = z_{i,1}(t) + (\theta^P_{z_{i,1}} + \kappa^P_{z_{i,1}} z_{i,1}(t) + \kappa^P_{z_{i,1},v_i}(t))\Delta t + \sqrt{v_i(t)}\sqrt{\Delta t}u_{i,t+\Delta t}, \tag{9.3} \]

and

\[ v_i(t + \Delta t) = v_i(t) + \kappa^P_{v_i}(\theta^P_{v_i} - v_i(t))\Delta t + \sigma^P_{v_i}\sqrt{v_i(t)}\left(\rho^P_{v_i}\sqrt{\Delta t}u_{i,t+\Delta t} + \sqrt{1 - \rho^2_i}\sqrt{\Delta t}\epsilon_{i,t+\Delta t}\right), \tag{9.4} \]

where, \( u_{i,t} \sim iid.N(0,1) \) and \( \epsilon_{i,t} \sim iid.N(0,1) \) are independent. One immediate issue with the volatility state variables \( v_i(t + \Delta t) \) is that they can become negative. This is different from the continuous-time square-root process, where the volatility variables, \( v_i(t + \Delta t) \), will stay non-negative when \( 2\kappa^P_{v_i}\theta^P_{v_i} \geq \sigma^2_{v_i} \). Depending on the the size of time interval \( \Delta t \), it might introduce bias in the state variables.

On the contrary, if the SDEs followed by state variables can yield closed-form solutions, we can actually solve the SDEs and obtain an unbiased discretization of the state variables. This involves solving the square-root process. Since the kalman filter is essentially about handling the conditional mean and conditional variance of the state process (and measurement process), we will solve the SDEs to obtain them. Based on the properties
of the square-root process, we assume the state variable dynamics as

$$z_t = m_0 + m_1 z_{t-1} + \omega^{1/2} (z_{t-1}) u_{1,t}. \quad (9.5)$$

For notational simplicity, we will assume $z_t = (z_1(t), \ldots, z_d(t), v_1(t), \ldots, v_d(t))$, such that, the state variables dynamics $z_t$ contains all state variables, including all $z_i(t)$ and all volatility variables $v_i(t)$. The affine property of the square-root process makes it easy to see that conditional mean obeys

$$E_t[z_T] = m_0(T - t) + m_1(T - t) z_t \quad (9.6)$$

and the conditional variance obeys

$$Var_t[z_T] = \omega(z_t) \omega^T(z_t) = \omega_0(T - t) + \omega_1(T - t) z_t. \quad (9.7)$$

Previous sections have seen various state variable dynamics in different model specifications. Given the special structure of state variable process, it is easy to show that they all follow the form of

$$dz_t = (a + Az_t) dt + \sigma(z_t) dW_t, \quad (9.8)$$

where, $z_t, dW_t \in \mathbb{R}^d$, $a$ is a $d \times 1$ vector, $A$ and $\sigma(z_t)$ are $d \times d$ matrices. As stated above for $z_t = (z(t), v(t))$, $dW_t = (dW_t, dB_t)$. The square-root property implies that

$$\sigma(z_t) \sigma^T(z_t) = G^0 + G^z z_t,$$

where $G^z = (G^1, G^2, \ldots, G^d)$ and $G^i$ is a $d \times d$ matrix for $i = 0, 1, \ldots, d$. Based on Fisher and Gilles (1996), we use vectorization operation to derive the ODEs that the conditional mean and conditional variance should satisfy. Appendix A.4 shows in details the derivation. We obtain ODEs that coefficients of conditional mean should satisfy,

$$m'_0(t - t) = a + A m_0(T - t) \quad (9.9)$$

and

$$m'_1(t - t) = A m_1(T - t), \quad (9.10)$$

with initial conditions $m_0(0) = 0$ and $m_1(0) = I_d$.

Assume $G = vec(G^z) = (vec(G^1), \ldots, vec(G^d))$. The conditional variance should
satisfy

$$\omega_0'(T - t) = vec(G^0) + Gm_0(T - t) + (A \otimes I_d + I_d \otimes A)\omega_0(T - t)$$  \hspace{1cm} (9.11)

and

$$\omega_1'(T - t) = Gm_1(T - t) + (A \otimes I_d + I_d \otimes A)\omega_1(T - t)$$  \hspace{1cm} (9.12)

with initial conditions $\omega_0(0) = 0$ and $\omega_1(0) = 0$.

Equations (9.9), (9.10), (9.11), and (9.12) form a system of ODEs that the conditional mean and conditional variance should satisfy. In common dynamic affine term structure models, matrix $A$ is usually invertible and $\sigma(z_t)\sigma^T(z_t) = \Sigma S \Sigma$, with $S$ being diagonal. In that case, an explicit solution to the ODEs can be obtained. But here in our model specifications, $A$ is not invertible. Therefore, we turn to numerical methods to solve the ODEs.

The ODEs for the conditional mean and conditional variance in equations (9.9), (9.10), (9.11), and (9.12) are solved with the popular fourth-order Runge-Kutta method\(^5\). This method is based on the Taylor series and uses a weighted average of the function values to compute differentials. Two additional middle points are included to increase the approximation accuracy.

### 9.2 Extended Kalman Filter

Kalman filter is a set of equations used to update a linear system of two linear equations. One is a discrete-time linear difference equation, called state equation, the other is called the observation equation, which relates the observation variable to the state variables. The observation equation is also called measurement equation. It provides an efficient way to recursively estimate the state process.

The key point in Kalman filtration is that the predicted errors of the observations are used to update a prior prediction of the state process such that a posterior prediction of the state process is obtained and is used to calculate the prior prediction of the state process for the next period. The gain, called Kalman gain, from the filtration is the product of Kalman factor and the errors of the predicted observations. This is the feedback effect. The state process affects the observation variables through the

observation equation, which, in return, improves on the prior prediction on the state process based on the Kalman gain. How the state process is related to the observation variables is also how the errors of the predicted observations is fed back to update the prior predictions of the state variables.

Kalman filtration involves two steps: a prediction step and an updating step. At the beginning of each time period, only the posterior state process and the posterior covariance of the state prediction errors are needed. In the prediction step, a prior predicted state process is calculated through the state equation with the state process from the last period. Meanwhile, a prior covariance matrix of the prediction errors is also computed. In the updating step, the Kalman factor is calculated as a function of the prior covariance of the prediction errors, coefficient matrix of the state variables in the observation equation, and the covariance of the observation innovations. Then the predicted prior state process is updated as a posterior state process with the Kalman filter and the prediction errors of the observations. Finally, the covariance of the prior prediction errors of the state process is updated as a posterior covariance of the prediction errors of the state process. Now the posterior state process and the posterior covariance of the prediction errors for the state process are ready to be used in the next period.

When measurement equations are nonlinear, the extended Kalman filter is widely implemented. The extended Kalman filter uses a linearization to approximate the nonlinear measurement equation around the expected value of state variables. The extended Kalman filter differs in the calculation of the covariance and the Kalman factor using the Jacobian matrix of the nonlinear system. Similar to the Taylor series, higher order derivatives can be used in the linearization. But, usually, only first order derivatives, or the Jacobian matrix is used. In the application to U.S. Treasury yields, the measurement equation is a linear function of the state variables, so ordinary Kalman filter is used. Because the LIBOR and swap rates are nonlinear functions of the state variables, the extended Kalman filter is executed.

For nonlinear system

\[ Y_t = g(z_t) + \epsilon_t \quad (9.13) \]

\[ z_t = Az_{t-1} + u_t, \quad (9.14) \]

where, \( g(z_t) \) are nonlinear vector functions of order \( m \times 1 \). In the application of LIBOR
and swap rates, $Y_t$ is the vector of LIBOR and swap rates and $g(z_t)$ is the nonlinear functional form of the state variables. $u_t$ and $\epsilon_t$ are independent state innovations and observation innovations, respectively, and $u_t \sim N(0, Q_t(z_{t-1}))$, $\epsilon_t \sim N(0, R)$. Conditional variance $Q_t(z_{t-1})$ is a linear function of $z_{t-1}$. We assume that the information set at the end of $(t-1)$ is $\mathcal{F}_{t-1} = \{\hat{z}_{t-1}, \hat{P}_{t-1}, Y_{t-1}\}$, which is also the information set at the beginning of $t$, before $Y_t$ is observed. After $Y_t$ is observed, the information set becomes $\mathcal{F}_{t} = \{\mathcal{F}_{t-1}, Y_t\}$. At the end of time $t$, after the updating stage, the information set becomes $\mathcal{F}_t = \{\hat{z}_t, \hat{P}_t, Y_t\}$. The nonlinear system is linearized around $\hat{z}_t^-$ with following linear system

$$
\hat{z}_t^- = E[z_t|\mathcal{F}_{t-1}] = A\hat{z}_{t-1}
$$

(9.15)

and

$$
Y_t = g(z_t) + \epsilon_t = g(\hat{z}_t^-) + G(\hat{z}_t^-)(z_t - \hat{z}_t^-) + \epsilon_t,
$$

(9.16)

where the Jacobin matrix $G$ of $g(z_t)$ is

$$
G_{i,j}(\hat{z}_t^-) = \frac{\partial g_i}{\partial z_j}(\hat{z}_t^-).
$$

(9.17)

The extended Kalman filter can be implemented as follows,

1. Prediction step, at the beginning of time $t$, with $\hat{z}_{t-1}$ and $\hat{P}_{t-1}$, calculate

$$
\hat{z}_t^- = A(\hat{z}_{t-1}),
$$

(9.18)

$$
\hat{Y}_t^- = g(\hat{z}_t^-),
$$

(9.19)

and calculate Jacobian matrix

$$
G_{i,j}(\hat{z}_t^-) = \frac{\partial g_i}{\partial z_j}(\hat{z}_{t-1})
$$

(9.20)

and

$$
\hat{P}_t^- = A\hat{P}_{t-1}A^T + Q.
$$

(9.21)

2. Updating step, calculate

$$
K = \hat{P}_t^- G^T (G\hat{P}_t^- G^T + R)^{-1},
$$

(9.22)
\[ \hat{z}_t = \hat{z}^- + \hat{P}_t^- G^T (G \hat{P}_t^- G^T + R)^{-1} (Y_t - g(\hat{z}^-)), \] (9.23)

and

\[ \hat{P}_t = \hat{P}_t^- - KG\hat{P}_t^- . \] (9.24)

Thus, at the end of time \( t \), we have \( \hat{z}_t \) and \( \hat{P}_t \) ready for the next period \( t + 1 \). Due to the complexity of the LIBOR and swap rate functions, Jacobian \( G(z_t) \) is numerically calculated.

The Kalman filter is initialized on the unconditional mean and variance of the state variables when the system is stationary. In our case, the state variable process is non-stationary. Different ways have been developed to initialize the Kalman filter for a non-stationary state space model, such as Resenberg (1973), Jong (1991), and Koopman (1997), to name a few. The gist of all those proposed methods is to use the sample data to infer the initial values of state variables. Our Kalman filter is initialized by using what we call the mean-value method, which uses the whole data to estimate the initial values. We estimated all the \( z_s \) for each period by minimizing the sum of \( (f(t, x) - G(z, x))^2 \), for all \( x \). This is just like fitting \( G(z, x) \) to the market observed data. Then the average of each \( z \) over all sample periods is calculated. The average values of \( z_s \) are used as the initial values of the state variables. In this way, the whole sample information is used to estimate the initial values of the state variables and these mean values can be interpreted as investors prior information on the initial forward curve. To further smooth out the effects of the initial condition, the first ten observations are cut off when calculate the fitted and forecast errors.

We also try to partition the state variables into non-stationary ones and stationary ones. For the stationary state variables, their unconditional means are used as initial values. For those non-stationary state variables, their initial values are added as nuisance parameters to the optimization process. But the results turn out to be worse than the mean-value method. Judging by the fitted and predicted errors, our mean-value method is satisfactory.

9.3 Application to U.S. Treasury Yields

Different model specifications are first applied to the U.S. Treasury yields. Nine series of the Constant Maturity Treasury (CMT) yields are used. These yields are different
from, but highly correlated with, the market yields mainly due to the statistical method used to calculate the CMT yields. Constant Maturity Treasury yields have been used as the replacements of the on-the-run market yields in academic research and business practices, such as Duffee (1998) and Duffie and Singleton (1999). This is mainly due to its availability. After the brief introduction of the CMT yields, we will first convert the CMT yields into zero-coupon bond yields, then Kalman filter is used to conduct the maximum likelihood estimation. Model comparisons are conducted based on the fitted errors, one-period ahead forecast errors, and AIC and BIC. The geometric information of the extracted factors are also analyzed. A statistic called information utilization is proposed to study how investors respond to the changes in the market situation.

9.3.1 Data

All the yields data are obtained from the website of the Board of Governors of the Federal Reserve in its H.15 release. Nine series of weekly yields are retrieved, including 3-, 6-month, 1-, 2-, 3-, 5-, 7-, 10-, and 20-year yields, over the time period of Jan. 4, 2002 through Nov. 18, 2011. Figure-2 depicts the whole CMT yields. The data covers the 2008 financial crisis in U.S. and includes the time periods of extremely low interest rates. Even the website of the Department of Treasury states the possibility of zero and negative yields. The data is a bit noisier than usual. Constant Maturity Treasury yields are computed by the U.S. Treasury department based on the actively traded on-the-run treasury securities and published by the Federal Reserve Bank of New York. These yields are par yields and bond equivalent yields on the bonds which pay semiannual interest. Simple compounding is used. The Constant Maturity Treasury Yields are neither annual percentage yields nor zero-coupon bond yields. Constant Maturity means that the yields are read on some fixed maturities even though there is no outstanding Treasury securities with those maturities. Constant Maturity Treasury Yields may be different from the market-observed yields due to the statistical method used to fit the smooth yield curve. Therefore, we first convert the CMT yields to zero-coupon bond yields with

\[ y(t, T) = 2 \ln \left( 1 + \frac{y^{CMT}(t, T)}{2} \right) \]  

(9.25)

\[ ^{6}\] For details, please refer to the websites of U.S. Treasury and Federal Reserve
where \( y(t, T) \) is the zero-coupon bond yield at time \( t \) maturing at time \( T \) and \( y^{CMT}(t, T) \) is the CMT yield at time \( t \) with maturity time \( T \). Figure-3 shows the converted zero-coupon bond yields.

Descriptive statistics of the CMT yields, the zero-coupon yields converted with equation (9.25), and the difference between the CMT yields and their corresponding zero-coupon bond yields are reported in Table-1, Table-2, and Table-3. Not surprisingly, the yields with shorter maturity have larger standard deviations with the 6-month yields fluctuate the most. The data also shows that the economy experiences some extremely low interest rate periods with the minimum 3-month yield being 0.1 basis point. This characteristic in the data also explains that the fitted yields computed with the estimated parameters also become negative occasionally. By comparing the maximum and the minimum values for each yield, we find that the yields of 3 months through 2 years span a wide range: 0.001% to 5.16%. It is mainly because our data period covers the last financial crisis and the low interest rate periods. This obviously brings more noise into the analysis. Interestingly, we found the moving pattern of factors changed after the crisis. This will be addressed later.

### 9.3.2 Estimation Results

Unlike the application to LIBOR and swap rates, where the measurement equation is nonlinear in state variables, the application to the zero-coupon yields converted from the CMT yields uses an ordinary Kalman filter since the zero-coupon bond yields in the measurement equation are affine in state variables. The difference is that the linearization is unnecessary in the ordinary Kalman filter. Please refer to the extended Kalman filter in Section 9.2 for details. The conditional mean and conditional variance follow the system of ordinary differential equations formed by equations (9.9), (9.10), (9.11), and (9.12). The Fourth-order Runge-Kutta method is used to numerically solve them.

Duffee and Stanton (2012) compared the methods used in the estimation of the dynamic term structure models, including Efficient Method of Moments (EMM), Simulated Maximum Likelihood (SML) and Quasi-maximum Likelihood (QML). They found that QML estimation offers better finite sample properties. Together with the ordinary Kalman Filter, this paper uses the Quasi-maximum Likelihood method to estimate the models discussed in the previous sections.
Specifically, we estimated one-, two-, and three-factor general volatility models, and a three-factor no-arbitrage Augmented Nelson-Siegel model. Table-5 reports the estimation results for the three models with general volatility as in Proposition 5.13 and Table-6 for the no-arbitrage Augmented Nelson-Siegel model as in Proposition 5.16.

According to the values of the likelihood function, the model preference is given in the order of the three-factor general volatility (GV3F) model, no-arbitrage Augmented Nelson-Siegel (ANS) model, then two- and one-factor general volatility models. The fitting and forecasting performance also show the preference of the three-factor general volatility (GV3F) model and the Augmented Nelson-Siegel (ANS) model. Therefore, the analysis will mainly focus on these two models. Not only do these two models have different specifications, but, more importantly, they have totally different interpretation of the factors. Each factor in the GV3F model contributes all three components of the yield curve: the level, slope, and curvature, while each factor in the ANS model only dedicate to one of the three components of the yield curve. Please refer to equations (5.35) through (5.37) and equations (5.71) through (5.73) for details. This will be elaborated in following sections.

For the GV3F model, we define the composite level, slope, and curvature in the following equations

\[ L^c = \sum_{i=1}^{n} \sigma_{i,1} z_{i,1}, \]
\[ S^c = \sum_{i=1}^{n} \sigma_{i,2} z_{i,1}, \]
\[ C^c = \sum_{i=1}^{n} \sigma_{i,3} z_{i,1}. \]

In Table-5, \( \sigma_{i,1}, \sigma_{i,2}, \) and \( \sigma_{i,3} \) determine factor \( z_{i,1} \)’s contribution to the level, slope, and curvature components of the yield curve. \( z_{1,1} \) contributes more to the level component than to the slope and curvature, while \( z_{2,1} \) contributes almost equally to the level and curvature components. \( z_{3,1} \) contributes more to the slope and curvature than to the level. The total contribution effect of each factor also depends on its value according to the equations above.

\( \sigma_{i,v} \) decides how much of volatility factor \( v_i \) is reflected in the yield factor of \( z_{i,1} \). It can be seen that \( \sigma_{i,v} \) does not vary much, which means that each volatility factor \( v_i \)
enters into yield factor in a similar way. \( a_i \) is the exponential decaying coefficient and a large value means \( z_{i,1} \) decays faster. The estimated value of \( a_1 \) is the highest one, which says that the the contribution of \( z_{1,1} \) to the slope and curvature components is diminishing faster than the other two factors, \( z_{2,1} \) and \( z_{3,1} \). \( a_2 \), as a value in the middle, makes \( z_{3,1} \)'s contribution to the slope and curvature decay faster than \( z_{2,1} \), which is the most persistent factor among the three.

As mentioned in the previous section, when \( \rho_i = \pm 1 \), or when the Brownian motions for \( z_i \) and \( v_i \) are perfectly correlated, the unspannedness disappears and the volatility becomes spanned stochastic volatility. Therefore, \( 1 - |\rho_i| \) can be regarded as how much unspannedness exists in the model. For the first volatility factor \( v_1 \), the unspannedness is 0.6439, or 64\% of \( v_1 \) can not be spanned with bond position. Similarly, 65.84\% of \( v_2 \) and 59.73\% of \( v_3 \) are unspanned. In total, there are almost half of the volatility risk cannot be spanned by positions only containing bonds.

\[ \theta^P_{zi}, \kappa^P_{zi}, \text{and} \kappa^P_{zi,v_i} \text{control the dynamics of} \ z_i \text{under the physical measure. Together} \]
\[ \text{with} \ v_i, \text{they form a term,} \ Mean_{z_{i,1}}^{\text{shadow}} = \frac{\theta^P_{z_{i,1}} + \kappa^P_{z_{i,1},v}v(t)}{-\kappa^P_{z_{i,1}}}, \text{which can be interpreted as the central tendency of} \ z_i, \text{as in equation (8.18). But this interpretation is different from the usual meaning of the central tendency. For a usual Ornstein-Uhlenbeck process, the central tendency is constant. Here, however,} \ v_i \text{is stochastic and changes over time. Besides being the kernel of the so-called central tendency,} \ v_i \text{is also the stochastic volatility.} \]

The double-role played by \( v_i \) makes the so-called central tendency in equation (8.18) more like a shadow tendency or co-moving tendency. Therefore, \( z_i \) co-moves with its shadow mean, \( Mean_{z_{i,1}}^{\text{shadow}} \), and not necessarily goes around or tends to \( v_i \). Whether \( z_i \) centers around \( Mean_{z_{i,1}}^{\text{shadow}} \), or is just co-moving with \( Mean_{z_{i,1}}^{\text{shadow}} \), it depends on how the factors are extracted from the data. But no matter in which way, we can safely say that \( z_i \) co-moves with \( Mean_{z_{i,1}}^{\text{shadow}} \) after all the central tendency is a fixed target to approach, while the shadow tendency is a floating target to follow. This also makes our model different from Chen (1996), which studied a three factor model with stochastic mean and volatility. But his stochastic mean and volatility are played by two different variables instead of one. Here the stochastic volatility plays double-role. \( \kappa^P_{z_i} \) serves as the reversion speed of \( z_i \), it controls how fast \( z_i \) co-moves with \( Mean_{z_{i,1}}^{\text{shadow}} \).

Being the kernel of the shadow mean, \( Mean_{z_{i,1}}^{\text{shadow}}, v_i \) plays the central role in it. After all, \( \theta^P_{zi}, \kappa^P_{zi}, \text{and} \kappa^P_{zi,v_i} \text{are constants, which only shift and rescale} \ v_i \text{and do not} \]
change the moving pattern of $v_i$. From this point of view, the kernel, $v_i$, can also be loosely called shadow mean of $z_{i,1}$.

From Table-5, $z_{1,1}$ and $z_{2,1}$ are co-moving with their shadow means, $Mean_{z_{1,1}}^{\text{shadow}}$ and $Mean_{z_{2,1}}^{\text{shadow}}$, respectively, much faster than $z_{3,1}$ with $Mean_{z_{3,1}}^{\text{shadow}}$. It also means that $z_{3,1}$ co-moves with $Mean_{z_{3,1}}^{\text{shadow}}$ in a more persistent way. Compared with $z_i$, $v_i$ reverts faster than $z_i$ as $\kappa_{v_i}^p$ is larger than $\kappa_{z_i}^p$ for each $i$. Intuitively, this makes sense since the volatility responds quickly to the news than the yield factors and also can shakes off the effects faster than the yield factors. This property can also be found in all the plots of the factors.

Table-6 reports the estimates of the parameters for our no-arbitrage Augmented Nelson-Siegel (ANS) model. The parameters have similar meanings to those in the GV3F model. The ANS model also has three factors but each factor dedicates to one of the three components, level, slope and curvature, of the yield curve. There is no decaying coefficient for level factor $z_{1,1}$.

The decaying coefficient for the slope factor and curvature factor, $z_{2,1}$ and $z_{3,1}$, are smaller than their counterparts in the GV3F model. It means that the slope and curvature in the ANS model are more persistent. However, the $z_{1,1}$ and $z_{3,1}$ co-move at a higher speed with their shadow tendency, $v_1$ and $v_2$, respectively, than their counterparts in the GV3F model. While $v_2$ and $v_3$ have a similar mean-reverting speed to that in the GV3F model, $v_1$ in ANS model enjoys a faster reverting speed. The unspannedness for the level factor has a surprising 91.2%, while the 88.9% of slope volatility and 89.6% of curvature volatility are unspanned. The curvature volatility and curvature factor have a negative correlation of -0.1046.

### 9.3.3 Comparison of Model Performance

Figure-4 and Figure-5 show the whole set of fitted yields and the fitting errors for three-factor general volatility (GV3F) model. In Figure-5, it can be seen that there are some spikes at the short-term and mid-term yields during the financial crisis in 2008. Figure-7 shows that the one-week ahead forecast errors have similar spikes in the financial crisis period. Figure-6 shows the one-week ahead forecast of all yields. Similar patterns can also be found in the fitted errors and forecast errors with the Augmented Nelson-Siegel (ANS) model. Refer to Figure-23 and Figure-25 for the ANS model for details.
Root-mean-square errors for the fitted yields and the one-week ahead forecasts with various models are reported in Table-9 and Table-10. The AIC and BIC for various models are reported in Table-13. Figure-8 and Figure-9 plot the RMSEs for the fitted and forecast yields. The overall in-sample goodness-of-fit and out-of-sample forecast performance from the RMSEs indicate the preference of the GV3F and no-arbitrage ANS models to one- and two-factor general volatility models. These two models have highly comparable performance to other literature (Christensen, Diebold, and Rudebush (2010)) even though our sample data have fewer cross-sectional yields and cover the financial crisis time and our models exhibit unspanned stochastic volatility.

From Table-9, the GV3F model and the ANS model show comparable in-sample performance with the average RMSEs of 5.57 and 5.33, which are much lower than the other two model specifications. The GV3F and ANS models both have three yield factors which allow them to capture the time series movement of the yield curve much better than the models with fewer factors. The comparable in-sample performance also shows that the ANS model is capable of capturing the cross-sectional variations in the yield curve better than the GV3F model. The ANS model is more parsimonious than the GV3F model with 7 fewer parameters. The parsimony of the ANS model, however, doesn’t hinder its cross-sectional performance since it has factors cover the level, slope and curvature in the yield curve. On the other hand, for the GV3F model, each of its three factors contributes to all three components of the yield curve. It seems that there might be some redundancy in the overlapping coverage of the cross-section of the yields.

In Table-10, the GV3F model slightly outperforms the ANS model in one-week ahead forecasts. The advantage of the GV3F model is very narrow, with less than two basis points. Therefore, we can claim that the ANS model is more desirable than the GV3F model because it is more parsimonious and its three factors represent independent geometric meanings of the yield curve. The breakdowns of in-sample performance comparison shows that the ANS model has four yields, 1-, 5-, 10-, 20-year yields, fitted better than the GV3F model with fitted errors less than 5 basis points. The ANS model also fits better for the 3-month yields. But the GV3F model is superior in the one-week ahead forecast performance over mid-term and long-term yields, including 5-, 7-, 10-, 20- year yields. It also forecasts 3-month yields better than the ANS model.

The AIC and BIC reported in Table-13 show the preference of the GV3F and ANS
models over models with fewer factors. It favors the GV3F model over the ANS model mainly because the GV3F model has a higher likelihood function value. The comparisons above is just for the overall performance. The following section will look into the details of the GV3F and ANS models with the interpretation of factors extracted from each model.

9.3.4 Interpretation of Factors

Like most affine term structure models, the factors in our models are also unobservable. Assigning some meaning to them helps to understand the roles they play. That is also one of the initiatives of our modeling attempts. With the help of the Kalman filter, the unobserved factors are extracted using the maximum likelihood estimates of the parameters.

Table-15 shows the descriptive statistics of the extracted factors and other state variables from the three-factor general volatility (GV3F) model. Table-19 reports the descriptive statistics of the extracted factors and other state variables from the no-arbitrage Augmented Nelson-Siegel (ANS) model. Even though the theoretical value for the volatility factors should be positive, due to the numerical calculation, the extracted values of volatility factors have several negative values. Unlike Duffee (1999), which truncated the negative values to zero in the estimation process, we leave it to the optimizer to choose a nearby point and pass through the negative points.

As mentioned in previous sections, the model framework in this paper is proposed on the basis of the geometric information of the yield curve, which is reflected by the geometric functions that factors play in the model. This was briefly mentioned in Section 9.3.2. In this section we will analyze the time series movement of the factors and the roles they play in shaping the yield curve.

Unlike other stochastic volatility models, such as Chen (1996), in our models, the stochastic volatility factor $v_i$ plays double roles: the volatility and the kernel of the "long-run" mean of the yield factor $z_{i,1}$. This "long-run" mean is called the shadow mean. $v_i$ is the kernel of the shadow mean of $z_{i,1}$. The shadow mean is the co-moving target of the yield factor. The co-movement pattern, however, can either be approaching a constant (long-run mean), or just following shadow mean. The reason that we separate these two patterns is because, the stochastic volatility factor $v_i$ plays as the volatility
of the yield factor, \( z_{i,1} \), and also as the kernel of the shadow mean of the yield factor. Being a volatility factor, \( v_i \) can fluctuate more than the yield factor \( z_{i,1} \), while, being the kernel of the shadow mean, \( v_i \) can stay either higher or lower than \( z_{i,1} \) and not necessarily is centered around by \( z_{i,1} \). That is why \( \text{Mean}^{\text{shadow}}_{z_{i,1}} = \frac{\theta_{P}^{P} Z_{i,1} + \kappa_{P}^{P} v_i(t)}{-\kappa_{P}^{P} Z_{i,1}} \) is called shadow mean. However, this definition doesn’t change the way yield factors \( z_{i,1} \) move with its shadow mean, \( \text{Mean}^{\text{shadow}}_{z_{i,1}} \).

Three-factor General Volatility (GV3F) Model  

We first look at the GV3F model, Figure-10 through Figure-12 show the time series of the yield factors \( z_{i,1} \) plotted against the volatility factor \( v_i \) as the yield factor and volatility factor are pairwise correlated. It can be seen that the fluctuation of the three volatility factors, \( v_i \), are more frequent and short-lived compared with the yield factors, \( z_{i,1} \). This is because the reverting parameters satisfy \( \kappa_{v_i} > |\kappa_{P}^{P} Z_{i,1}| \). If we split each figure into two parts at 2007, all factors, yield and volatility, fluctuate more often after the economy falls into the last crisis. The second yield factor \( z_{2,1} \) and its correlated volatility factor \( v_2 \) have clear co-moving pattern during the crisis from about 2007 through 2009.

This pattern appears even stronger for the third yield factor \( z_{3,1} \) and the third volatility factor \( v_3 \) and lasts even longer through the end of the sample period. On the other hand, the co-moving pattern is not very clear before the crisis. The co-moving pattern is not much visible for the first yield factor \( z_{1,1} \) and volatility factor \( v_1 \). Two of the three yield factors tell that the yield factors co-move with their correlated volatility factors, especially for \( z_{3,1} \), which shows stronger and longer co-moving pattern. Figure-13 through Figure-15 show the movements of yield factors and their corresponding shadow means which are calculated with \( \text{Mean}^{\text{shadow}}_{z_{i,1}} = \frac{\theta_{P}^{P} Z_{i,1} + \kappa_{P}^{P} v_i(t)}{-\kappa_{P}^{P} Z_{i,1}} \). The calculation is more like a rescaling and repositioning of the volatility factors since the shadow mean is a linear function of the volatility factor. Now the first yield factor \( z_{1,1} \) shows some co-moving pattern with its shadow mean, \( \text{Mean}^{\text{shadow}}_{z_{1,1}} \), from the beginning of the financial crisis through 2009. The third yield factor \( z_{3,1} \) still maintains the strongest and longest co-moving pattern with its shadow mean, \( \text{Mean}^{\text{shadow}}_{z_{3,1}} \). All these figures indicate that there is a clear structural break after the last financial crisis.

Figure-16 through Figure-18 depict the co-movement of the composite level, slope and curvature, \( L^c, S^c, C^c \), of the yield curve and their corresponding shadow means. All
values are calculated according to equations (5.35) through (5.37). Since each factor in the GV3F model contributes to all of the level, slope, and curvature, the composition of the three factors in $L^c, S^c, C^c$ might blur their co-movement patterns with their composite shadow means. This explains that the patterns for three figures are not as clear as those in Figure-13 through Figure-15. But the composite slope, $S^c$, and composite curvature, $C^c$, show a very similar moving pattern after the economy falls into the last financial crisis. This is mainly because some factors may become more dominant than others. As is to show in the ANS model, the pattern is revealed more clearly since each factor in the ANS model dedicates to one of the level, slope, and curvature of the yield curve.

**Augmented Nelson-Siegel (ANS) Model**

Even though both the ANS model and the GV3F model have the same dimension of factor space, the main difference lies in that the ANS model has separate factors for the level, slope, and curvature of yield curve. Without composing the factors to form the shape of the yield curve, the ANS model has clear advantage of identifying the co-moving pattern of the yield factors.

Figure-28 through Figure-30 depict the co-movement of the yield factor $z_{i,1}$ and volatility factor $v_i$. The first two yield factors, $z_{1,1}$ and $z_{2,1}$, or the level and slope factors, show a clear co-moving pattern with their correlated volatility factors, $v_1$ and $v_2$, when the economy enters into the 2008 financial crisis. On the other hand, the third yield factor, $z_{3,1}$, does not show a clear co-movement pattern with its correlated volatility factor $v_3$. By comparing with the three factors in the GV3F model, the three yield factors in the ANS model exhibit much clearer co-movement pattern after the financial crisis.

Figure-31 through Figure-33 plot the three yield factors and their shadow means. The co-movement patterns appear clear with the level and slope factors, or the first two factors. All these figures indicate that, the three factors, level, slope, and curvature, all become volatile after the financial crisis and they begin to move in a similar pattern. Before the financial crisis, they don’t show any clear sign of co-movement. This characteristic of factor dynamics indicates that, after the financial crisis, the yield curve evolves in a different way and a structural break happens after the financial crisis.

The advantage of the GV3F model over the ANS model is that the former bears a more general form of forward rate volatility than the latter. The GV3F model nests the ANS model. It brings more flexibility to model but the in-depth analysis of the extracted
factors show blurring pictures due to the composition of the factors. On the contrary, being a special case of the GV3F model, the ANS model is more parsimonious and the independence of yield factors brings clearer pictures of the time series movement of the factors. The empirical performance of the two models are comparable with the GV3F model having a narrow advantage, which comes at the cost of a larger parameter space. As to fitting the cross-sectional yields, the three factors in the GV3F model bring a redundancy to the model because each factor contributes to all of the level, slope, and curvature of the yield curve. As shown above, the ANS model is also capable of capturing the cross-sectional relations among the yields because each of its three factors captures one of the level, slope, and curvature.

9.3.5 Information Utilization

The discussion of the RMSEs in Section 9.3.3 only gives a macro-picture of the model performance. This section will look into the details about how the model reacts to the information evolution. If the model is viewed as the investor’s decision-making rule, then this discussion can be regarded as how investors react to the changes in the market. It can reveal more information than the simple overall comparison of model performance through the RMSEs of the fitted and forecast yields.

A statistic called information utilization is defined as the ratio of the fitted error to the predicted error of the yields. The forecast error indicates how an investor miscalculates the yields with the information available from the last period or at the beginning of this period, while the fitted error shows how an investor reevaluates the market situation and updates his perception of the market. Theoretically and ideally, a rational investor is supposed to narrow the error since more information is used in reevaluating the market situation. It means that, under normal situation, the fitted error is smaller than the forecast error, or the information utilization should have an absolute value smaller than 1. If the absolute value is higher than 1, it means investors overreact to the information that just becomes available. Moreover, if the information utilization is lower than -1, it indicates that investor not only overreacts the the new information but also reverts their initial judgment about the market situation based on the information from the last period.
Information utilization, $info_{utl}$, is defined as

$$info_{utl} = \frac{y_t(\tau) - \hat{y}_t^{fit}(\tau)}{y_t(\tau) - \hat{y}_{t-1}^{forecast}(\tau)} = \begin{cases} 
\text{overreaction} & info_{utl} > 1 \\
\text{normal reaction} & |info_{utl}| < 1 \\
\text{overreaction and reversion} & info_{utl} < -1
\end{cases}$$

(9.26)

where $y_t(\tau)$ is the time-$t$ yield with the time-to-maturity $\tau$, $\hat{y}_t^{fit}(\tau)$ is the fitted yield, and $\hat{y}_{t-1}^{forecast}(\tau)$ is the one-period-ahead forecast yield.

**Three-factor General Volatility (GV3F) Model**  
Figure-19 through Figure-21 show how investors react to the market information when the GV3F model is applied. The $info_{utl}$ is truncated to $\pm 5$ for illustration purposes. The actual values can be much larger than 5. It provides far more information than the simple comparison of the RMSEs. Most yields experience two main volatile periods: one is around 2002 to 2004, the other is around the 2008 financial crisis. The 3-month, 1-, 2-, 3-, 10-, 20-year yields experience long quiet periods within the band of $\pm 1$ between these two volatile periods. When the financial crisis hits the economy, they exhibit large spikes. The 3-month short rate and 10-year benchmark yield begin to respond to the crisis almost simultaneously with more fluctuations in the $info_{utl}$ of the benchmark yield. Interestingly, the mid-term yields, such as 2-, 3-, 5-year yields, all show lags of response. This is mainly because the short rate and benchmark yield are closely watched by investors and more informative decisions are being made. The $info_{utl}$ of the 3-month yield and 10-year benchmark yield clearly show the change of market situation and it would be interesting to compare it with business cycle information.

**Augmented Nelson-Siegel (ANS) Model**  
Figure-34 through Figure-36 show $info_{utl}$ when ANS is applied. A truncation point of $\pm 5$ is also used. It seems that investors have a good grip of the 3-month yield, which experiences a long quiet period within the band of $\pm 1$. Interestingly, when the financial crisis hits the economy, it also exhibits large spikes, and after the peak time of crisis, it goes back to the normal level. Intuitively, the 3-month yield is widely used as the short rate and it is closely watched. It is not unusual that investors have put more efforts in monitoring it in order to make more informative decisions. On the contrary, the 6-month and 1-year yields show more fluctuations than
their corresponding ones with the GV3F model.

With the maturity getting longer, info_utl becomes quieter too. The 20-year yield, which has the longest term in the data, is the most quiet one mainly because it is the least liquid yield. Similar to the 3-month yield, the 10-year benchmark yield also has two volatile periods: one is from late 2002 to early 2004 and the other is during the 2008 financial crisis, especially after mid-2009. But it also experiences a long quiet period before the crisis. Due to the popularity of 10-year benchmark yield, investors seem to make better decisions about it. The pictures painted by the 3-month short rate and the 10-year benchmark yield can be useful in monitoring the changes in the market situation. The next natural step is to compare this response indication with business cycle information and see if they can help with business cycle detection.

### 9.4 Application on LIBOR and Swap Rates

This section will apply various model specifications to the LIBOR and swap rates. We first establish the relation between the zero-coupon bond price and the LIBOR and swap rates and then implement the extended Kalman filter in the maximum likelihood estimation. Model comparison is conducted based on the fitted errors and predicted errors of the LIBOR and swap rates. AIC and BIC for various models are also reported. Geometric information of factors and information utilization are analyzed too.

#### 9.4.1 Data

We use weekly LIBOR and swap rates retrieved from Bloomberg. The data range is from Jan. 4, 2002 to Nov. 18, 2011. All weekly data are Friday closed rates. The LIBOR data have three series, 3-, 6-, and 9-month rates. The swap rates consist of eight series with swap tenors of 1-, 2-, 3-, 4-, 5-, 7-, 10-, and 15-year. The LIBOR and swap rates together bring us a total of 5676 (=516×11) rates. Figure-37 depicts the time series of the whole data set. Table-4 reports the descriptive statistics of all 11 series of LIBOR/Swap rates.

#### Zero-coupon Bond Price, LIBOR and Swap Rates

We first define the relationship between the zero-coupon bond price and the LIBOR and swap rates. Assume a tenor

---

7 In this section, we still use the terms such as zero-coupon bond yield and yield curve. But, in this section, the zero-coupon bond is based on LIBOR and swap rates instead of Treasury bonds.

8 I thank Professor Hany A. Shawky and Professor David M. Smith for allowing me to use the Bloomberg Terminal in the Department of Finance.
structure \( \mathcal{T} = \{ T_n : n = 1, 2, \ldots, N \} \) from \( t = T_0 \) to \( T_N \) with \( N \) consecutive interest rate reset dates \( T_i = t + \delta i, i = 0, 1, 2, \ldots, N - 1 \). The reference rate for the time interval \( \delta = T_n - T_{n-1} \) is the Eurodollar deposit rate with tenor \( \delta \), for \( n = 1, 2, \ldots, N - 1 \). For simplicity, we assume the time interval is evenly distributed. Then the time-\( t \) LIBOR rate, \( \text{LIB}(t, T_n) \) is defined as

\[
\text{LIB}(t, T_n) = \frac{1}{\delta_n} \left( \frac{1 - P(t, T_n)}{P(t, T_n)} \right),
\]

(9.27)

where \( P(t, T_n) \) is the zero-coupon bond price. Following the market practice of using the LIBOR forward rate, we can define the time-\( t \) forward LIBOR rate, \( \text{LIB}(t, T_m, T_n) \), over time interval \( \delta (n - m) = T_n - T_m \), as

\[
\text{LIB}(t, T_m, T_n) = \frac{1}{\delta (n - m)} \left( \frac{P(t, T_m) - P(t, T_n)}{P(t, T_n)} \right).
\]

(9.28)

By the definition of forward LIBOR rate in equation (9.28), if the zero-coupon bond \( P(t, T_n) \) is chosen as numeraire, the forward LIBOR rate, \( \text{LIB}(t, T_m, T_n) \), is a martingale under the associated equivalent probability measure, known as forward-risk adjusted measure.

Thus, the swap rate at time-\( t \) for time interval \( t \) to \( T_n \) (spot-starting swap) with payments at dates \( T_i, i = 1, 2, \ldots, n \), can be expressed as

\[
\text{Swp}(t, T_n) = \frac{1 - P(t, T_n)}{\sum_{i=1}^{n} \delta P(t, T_i)}.
\]

(9.29)

Note that the swap rate \( \text{Swp}(t, T_n) \) is for the time period \( t \) to \( T_n \) and the first payment date is \( T_1 \). In this case, time \( t \) is actually the first date when the reference rate is set and the reset date sequence becomes \( t, T_1, \ldots, T_{n-1} \). This is different from the forward swap rate where the first reference rate reset date is the forward-starting date and the first payment date is one period after that. The time-\( t \) forward swap rate for the period \( T_m \) to \( T_n \) is

\[
\text{Swp}(t, T_m, T_n) = \frac{P(t, T_m) - P(t, T_n)}{\sum_{i=m+1}^{n} \delta P(t, T_i)}.
\]

(9.30)

Here the first payment date is \( T_{m+1} \). Obviously, if let \( T_m = t \), we have \( m = 0 \) and equation (9.30) becomes equation (9.29). As swap contract is a sequence of forward rate agreements, the forward swap rate in equation (9.30) is more intuitive than equation
9.4.2 Estimation Results

As shown in Section 9.4.1, the LIBOR and Swap rates are nonlinear in state variables, which requires the use of the extended Kalman filter in Section 9.2. A step of linearization is added to the ordinary Kalman filter. The conditional mean and conditional variance also follow a system of ordinary differential equations formed with equations (9.9), (9.10), (9.11), and (9.12). The fourth-order Runge-Kutta method is used to numerically solve them.

Like the application to yields, we use the Quasi-maximum Likelihood method to estimate various model specifications, including one-, two-, and three-factor general volatility (GV3F) models and a three-factor consistent Augmented Nelson-Siegel (ANS) model. Table-7 reports the estimation results for the GV3F model as in Proposition 5.13 and Table-8 for the ANS model as in Proposition 5.16. Based on the values of the likelihood function, the model preference is given in the order of the Augmented Nelson-Siegel (ANS) model, three-factor general volatility GV3F) model, then two- and one-factor general volatility models. The fitting and forecasting performance also show the preference of the ANS model and the GV3F model. Just like the application to yields, the analysis will mainly focus on these two models. More importantly, these two models have different factor meanings. The GV3F model forms the composite level, slope, and curvature of the yield curve, while the ANS model has independent ones. Please refer to equations (5.35) through (5.37) and equations (5.71) through (5.73) for their definitions.

In Table-7, according to the values of \( \sigma_{i,1}, \sigma_{i,2}, \) and \( \sigma_{i,3}, z_{2,1} \) and \( z_{3,1} \) contribute more to the level component than \( z_{1,1} \). \( z_{2,1} \) has more contributions to the slope than other factors. The net effect of each factor also depends on its value according to their definitions. \( \sigma_{i,v} \) decides how much of the volatility factor \( v_i \) is reflected in the yield factor of \( z_{i,1} \). It shows that \( \sigma_{i,v} \) does not vary much, which means each volatility factor \( v_i \) enters into the yield factor in a similar way. The estimated value of \( a_1 \) and \( a_3 \) are close and higher than \( a_2 \), which means the contribution of \( z_{1,1} \) and \( z_{3,1} \) to the slope and curvature components is diminishing faster than the factor \( z_{2,1} \). \( a_2 \), as the lowest value among the three, makes \( z_{2,1} \)'s contribution to the slope and curvature the most persistent one.

The amount of unspannedness in the GV3F model is 65.56%, 64.49%, and 71.24% for
the three volatility factors, $v_1$, $v_2$, and $v_3$. It means that the bonds can only span about one-third of the total volatility risk.

Just like the application to the Treasury yields, $\theta_{z_1}$, $\kappa_{z_1}$, and $\kappa_{z_i,v_i}$ control the dynamics of $z_i$ under physical measure. Together with $v_i$, they form the shadow mean, $Mean_{z_i}^{\text{shadow}} = \theta_{z_i} + \kappa_{z_i,v_i}(t)$, as in equation (8.18). Table-7 shows that the three LIBOR/swap rates factors, $z_{1,1}$, $z_{2,1}$, and $z_{3,1}$ are co-moving with their shadow mean at a similar speed, respectively. Compared with $z_i$, $v_i$ reverts faster since $\kappa_{z_i,v_i}$ is larger than $\kappa_{z_i}$ for each $i$. Intuitively, this makes sense since the volatility responds quickly to the news than the yield factors. They also can brush off the effects faster than the yield factors. This property can be seen from all the plotting of the factors.

The estimates of parameters for the ANS model are reported in Table-8. The parameters have similar meanings to those in the GV3F model. The model also has three factors but each factor dedicates to one component of the level, slope and curvature of the yield curve. There is no decaying coefficient for the level factor $z_{1,1}$. The decaying coefficient for slope factor, $z_{2,1}$, is larger than its counterpart in the GV3F model while the curvature factor, $z_{3,1}$, has a smaller value than its counterpart in the GV3F model. It means that the slope in the ANS model is diminishing faster and, thus, is less persistent.

But the curvature factor, $z_{3,1}$, is more persistent and decays slower in the ANS model. Except the slope factor, $z_{2,1}$, the level factor, $z_{1,1}$, and the curvature factor, $z_{3,1}$, co-move at a slower speed with their shadow means, $Mean_{z_{1,1}}^{\text{shadow}}$ and $Mean_{z_{3,1}}^{\text{shadow}}$, respectively, than their counterparts in the GV3F model. The three volatility factors revert slower in the ANS model compared with those in the GV3F model. 84.91% of the stochastic volatility correlated with the level factor can not be spanned with LIBOR/Swap position. That rate goes up to 86.57% for the volatility associated with the slope factor, while it falls to 57.15% for the curvature factor.

### 9.4.3 Comparison of Model Performance

Root-mean-square errors for the fitted and one-period ahead forecasts LIBOR/Swap rates are reported in Table-11 and Table-12. The AIC and BIC for various models are reported in Table-14. Figure-42 and Figure-43 plot RMSEs for fitted and forecast LIBOR/swap rates. The overall in-sample goodness-of-fit and out-of-sample forecast performance from RMSEs show the preference for the ANS model and the GV3F model over one- and
two-factor general volatility models. Figure-38 and Figure-5 plot the fitted LIBOR/Swap rates and the fitting errors for the GV3F model. In Figure-39, it can be seen that there are some spikes at the short-end and mid-term of the LIBOR/swap rate surface in the financial crisis around 2008. Figure-41 which shows the one-week ahead forecast errors has similar spikes during the financial crisis period. Figure-40 shows the one-week ahead forecast of LIBOR/Swap rates. Similar patterns can also be found in the fitted errors and forecast errors with the Augmented Nelson-Siegel (ANS) model. See Figure-58 and Figure-60 for the performance of the ANS model.

From Table-11, the GV3F model and ANS model show comparable in-sample performance with the average RMSEs of 8.95 and 8.05, which are lower than the other two model specifications, with a slight advantage (0.9 basis point) to the ANS model. The GV3F model and ANS model both have three yield factors which allow them to capture the time series movement of the LIBOR/swap rates better than the models with fewer factors. The comparable in-sample performance of them also shows that the ANS model can well capture the cross-sectional variations in the LIBOR/swap rates compared with the GV3F model. The ANS model is more parsimonious than GV3F with 7 parameters less. The parsimony of the ANS model does not hurt its cross-sectional performance because its factors are capable of covering the level, slope and curvature of the cross-section LIBOR/swap rates. The GV3F model, on the other hand, has each of its three factors contribute to all three components of the cross-sectional rates. It seems that there might be redundancy in the overlapping coverage of the cross-sectional rates. Similar results also come with the applications to the Treasury yields.

In Table-12, the ANS model outperforms the GV3F model by 6 basis points in one-week ahead forecast. The advantage of the ANS model is noticeable with about 30% improvement compared with the GV3F model. It is safe to state that the ANS model is more desirable than the GV3F model because of its superior in-sample and out-of-sample forecast performance, not to mention it is more parsimonious. The breakdowns of the in-sample performance comparison shows that the ANS model has 6 out of 11 rates fitted better than the GV3F model. The forecast performance is unanimously leaning toward the ANS model for all rates. This further confirms the superiority of the ANS model.

The AIC an BIC reported in Table-14 show the preference of the ANS and GV3F models over those with fewer factors, and favors the ANS model over the GV3F model,
not surprisingly.

As noticed in the application to the Treasury yields, the overall comparisons above is limited as it lacks the time series details. The following section will look into the details of the GV3F and ANS models with the interpretation of the factors extracted from each model.

9.4.4 Interpretation of Factors

Similar to the analysis of the Treasury yields, the unobservable factors are extracted using the maximum likelihood estimates of the parameters. Table-17 shows the descriptive statistics of the extracted factors and other state variables from the GV3F model. Table-20 reports the descriptive statistics of the extracted factors and other state variables from the ANS model. Like in the last application, the paper does not truncate the negative values to zero, as in Duffee (1999), in the estimation process. Instead, it leaves to the optimizer to choose a nearby point and pass through the negative values.

In this section we analyze the time series movement of the factors and the roles they play in forming the geometric shape of the yield curve inferred from the LIBOR and swap rates. The interpretation of the factors in Section 9.3.4 also applies here, refer to Section 9.3.4 for details.

Three-factor General Volatility (GV3F) Model  For the GV3F model, Figure-44 through Figure-46 depict the time series movement of each yield factor, $z_{i,1}$, plotted against its correlated volatility factor, $v_1$. As can be seen, the fluctuation of the three volatility factors are short-lived compared with the yield factors. This is due to the reverting parameters of the volatility factors satisfy the condition of $\kappa_{v_1}^p > |\kappa_{z_1}|$. Similarly, each figure can be split into two parts around 2007, all the yield and volatility factors fluctuate more often after the economy falls into the last crisis. The first factor, $z_{1,1}$, shows a clear co-moving pattern with its correlated volatility factor $v_1$ from the beginning of the financial crisis through mid-2009, while the second factor, $z_{2,1}$, and $v_2$ co-move for even longer time through the end of the sample period. The third factor $z_{3,1}$ shows even clearer and stronger co-movement pattern with $v_2$ after the financial crisis. All three figures indicate closer co-movement patterns between the yield factors and their correlated volatility factors after the financial crisis.
Figure-47 through Figure-49 show the evolutions of the yield factors and their corresponding shadow means calculated with $\text{Mean}_{z_{i,1}}^{\text{shadow}} = \frac{\theta^{Z_{i,1}} + \kappa^{Z_{i,1}} v_i(t)}{-\kappa^{Z_{i,1}}}$. The pairwise co-movement patterns between the yield factors and their correlated volatility factors become even more noticeable. In the analysis of the Treasury yields in Section 9.3.4, a similar pattern is also found. This further confirms that our model is capable of capturing the structural break after the financial crisis. It also means our models can help identify dramatic changes in the market.

Figure-50 through Figure-52 depict the co-movement of the composite level, slope and curvature, $L^c, S^c, C^c$, of the LIBOR-swap-rates-based yield curve and their corresponding shadow means. In the GV3F model, each factor contributes to all three components: the level, slope, and curvature. The composition of the three factors in $L^c, S^c, and C^c$ may blur their co-movement patterns with their composite shadow means. This explains that the patterns in the three figures are not as clear as those in Figure-13 through Figure-15. But the composite slope, $S^c$, and composite curvature, $C^c$, show very similar moving pattern after the economy falls into the 2008 financial crisis. This is mainly because some factors dominate others in the composition. As is shown, the ANS model reveals more clearly the co-movement patterns since each factor in the ANS model dedicates to one component in the shape of the yield curve.

Augmented Nelson-Siegel (ANS) Model  As mentioned before, the main difference between the ANS and GV3F models lies in the fact that the former has independent factors for each of the level, slope, and curvature of the yield curve. Without composing the factors to form the shape of the yield curve, the ANS model has a clear advantage of identifying the co-moving pattern of the yield factors. Surprisingly, the co-movement patterns of the yield factors and volatility factors in the two applications are similar. In both cases, the level and slope factors and their correlated volatility factors show much more clear co-movement patterns than the curvature factor and its correlated volatility factor.

Figure-63 through Figure-65 picture the co-movement of the yield factor $z_{i,1}$ and volatility factor $v_i$. The first two yield factors, $z_{1,1}$ and $z_{2,1}$, or the level and slope factors, show clear co-moving pattern with their correlated volatility factors, $v_1$ and $v_2$, when economy drops into the 2008 financial crisis. On the other hand, the third yield factor,
$z_{3,1}$, does not show a clear co-movement pattern with its correlated volatility factor $v_3$.

Figure-66 through Figure-68 plot the three yield factors and their shadow means. The co-movement pattern appears clear for the slope and level factor. All these figures indicate that, the three factors, level, slope, and curvature, all become volatile during the financial crisis and they begin to move in a similar pattern. Before the financial crisis, they do not show any clear sign of co-movement. This characteristic of factor dynamics indicates that, after the financial crisis, the yield curve evolves in a different way. It also means a structural break happens after the financial crisis begins.

9.4.5 Information Utilization

As shown in the analysis of the Treasury yields, RMSEs are limited in revealing the information flows or changes in the market. In this section, we will use the information utilization defined in Section 9.3.5 to see how investors behave in the LIBOR/swap rates market.

**Three-factor General Volatility (GV3F) Model** Figure-53 through Figure-56 show how investors react to the flow of market information when the GV3F model is used. The $info_{utl}$ is truncated to $\pm 5$. Without the truncation, the absolute values of $info_{utl}$ can be much larger than 5. It provides far more information than the simple comparison of RMSEs.

Most rates, except the 4-, 5-, 7-year swap rates, experience two main volatile periods: one is around 2002 to 2004, the other is around the last financial crisis. The 3-, 6-, 9-month LIBOR and the 10-, 15-year swap rates experience some long quiet periods within the band of $\pm 1$ between these two volatile periods. When the economy alls into the 2008 financial crisis, $info_{utl}$ exhibits large spikes. The sudden spikes for the 3-month LIBOR last only through early 2009, while those for the 15-year swap rate last through 2009. Meanwhile, the spikes for the 10-year rate still occasionally appear. The 6-month and 9-month LIBOR do not dial down the effects from the financial crisis as fast as the 3-month rate. The spikes for the 9-month rate last through the end of the sample period. The 1-, 2-, 3-year swap rates seem quieter than the 4-, 5-, 7-year rates with fewer and shorter spikes. But most of the rates, 8 out of 11, still show the eruption of the financial crisis. The responses of the 3-month and 6-month LIBOR appear leading the way when
the financial crisis begins. The 9-month, 1-, 2-, 10-, 15-year rates respond later. Most figures show the potential use of info utl in identifying dramatic changes in the market and economy.

**Augmented Nelson-Siegel (ANS) Model** Figure-69 through Figure-72 show the time series of info utl when the ANS model is used. Truncation point of ±5 is also applied. All figures show more fluctuations than those with the GV3F model. Only the 3-month LIBOR shows sudden spikes at the beginning of the financial crisis. For other rates, they do not show sudden spikes when dramatic change happens. According to the model comparison based on the RSMEs, AIC and BIC, the ANS model is more desirable than the GV3F model. The possible explanation is that investors make small but frequent, instead of big and sudden, adjustments when the ANS model is used as decision rule. Thus, in the application to LIBOR/swap rates, the ANS model performs better than the GV3F model but cannot help identify sudden changes in the market, which is carried out better by the GV3F model.

10 Conclusion and Possible Extensions

Based on the geometric information of the yield curve (or equivalently, the forward rate curve), this paper proposes three affine Heath-Jarrow-Morton forward rate models. One model is a general model framework in the sense of Heath, Jarrow, and Morton (1992), another one is a multifactor affine HJM model with unspanned stochastic volatility, and the third is an Augmented Nelson-Siegel (ANS) model with unspanned stochastic volatility. The geometric information of the yield curve, such as the level, slope, and curvature, are represented with either correlated or independent factors. While maintaining the form of the yield curve unchanged, the model framework is flexible enough to accommodate various forms of forward rate volatility ranging from Gaussian to stochastic ones. Depending on whether or not the stochastic volatility factors bear their own randomness, the model framework can introduce either spanned or unspanned stochastic volatility. The model framework can also be applied to different combinations of factors representing the level, slope, and curvature.

When the volatility risk is unspanned, the volatility factor plays double roles in the model. One, of course, is to serve as the volatility factor and the other is to act as the
kernel of the shadow mean, which is a target for the yield factor to move along with. It is called shadow mean because it appears in the position of the long-run mean of the yield factor and it is also stochastic. This feature helps explain an empirical finding in the literature that the interest rate volatility is correlated with the level of interest rate and, meanwhile, cannot be spanned with a position containing only bonds.

Various model specifications are applied to two datasets, including zero-coupon bond yields calculated from the Treasury Constant Maturity yields and the LIBOR/swap rates. One-, two-, and three-factor general volatility (GV3F) models and a consistent Augmented Nelson-Siegel (ANS) model are tested with the two datasets. In both applications, the GV3F and ANS models outperform one- and two-factor general volatility models. The GV3F model has three factors with each one contributing to all three geometric components of the yield curve, including the level, slope and curvature. On the other hand, the ANS model has each factor only dedicate to one of the three components of the yield curve. RMSEs of the fitted and one-period-ahead forecast values from both models are highly comparable with other literature, such as Christensen, Diebold, and Rudebush (2010), which is a three-factor Gaussian short-rate model. But our analysis has the clear advantage of incorporating unspanned stochastic volatility and geometric information into the model.

The yield factors and volatility factors are extracted with the Kalman filter. Further analysis of those factors provides in-depth knowledge of the factor dynamics. With two models and two datasets, yield factors and volatility factors, yield factors and their shadow means, level, slope and curvature and their shadow means are analyzed in pairs. In most cases, the pairwise analysis shows that the co-movement pattern shifts after the economy falls into the 2008 financial crisis. It means that our model can help identify dramatic changes in the market and economy. This is useful in the business cycle analysis as well.

The paper also proposes a statistic called information utilization, \( \text{info\_utl} \), which is used to analyze how investors react to the information evolution in the market. The time series characteristic of \( \text{info\_utl} \) is studied for each yield and LIBOR/swap rate. It shows that, in most cases, the way that investors react to the market changes after the 2008 financial crisis. This further confirms that a structural change happens after the financial crisis. Another finding of the paper is that some rates show a response lag
compared to others. For example, the info_utl of the mid-term rate begin to spike later than the short rate when the financial crisis begins. Just like the pairwise analysis of the extracted factors, this finding is worth further investigation and is useful in identifying the business cycle.

Since the model framework proposed in the paper is theoretical consistent, it is ready to price interest rate derivatives as in Wang (2014a), where interest rate Caps are used. This surely expands the use of the model framework. As mentioned before, it is possible to join our models with the business cycle studies using the analysis of the extracted factors and information utilization. Also, it would be interesting to compare various combinations of the geometric components of the yield curve as noted in Section 5.1. We leave these extensions for future research.
References


A Appendix

A.1 Short Note on Stratonovich calculus

Assume \( F(t, z_t) \) with \( dz_t = \mu(z_t)dt + \sigma(z_t)dW_t \), then we have

\[
F(t, z_t) = F'_t(t, z_t)dt + F'_z(t, z_t)dz_t + \frac{1}{2}F''_{zz}(t, z_t)(dz_t)^2
\]

Under Stratonovich calculus, the chain rule for ordinary calculus can be extended to stochastic calculus, thus,

\[
F(t, z_t) = F'_t(t, z_t)dt + F'_z(t, z_t)\circ dz_t
\]

The corresponding terms between Ito’s lemma and Stratonovich calculus is

\[
F'_z(t, z_t)\circ dz_t = F'_z(t, z_t)dz_t + \frac{1}{2}F''_{zz}(t, z_t)(dz_t)^2
= F'_z(t, z_t)dZ_t + \frac{1}{2}d(F'_z(t, z_t))(dz_t)
\]

For equation (3.1), let \( \sigma_0(z_t, x) = F'_z(t, z_t) \) and

\[
\begin{align*}
\sigma_0(z_t, x)\circ dW_t &= \sigma_0(z_t, x)dW_t + \frac{1}{2}d(\sigma_0(z_t, x))(dW_t) \\
&= \sigma_0(z_t, x)dW_t + \frac{1}{2}d < \sigma_0(z_t, x), W_t > \\
\text{or} \quad \sigma_0(z_t, x)dW_t &= -\frac{1}{2}d < \sigma_0(z_t, x), W_t > + \sigma_0(z_t, x)\circ dW_t
\end{align*}
\]

A.2 Proof of Proposition 5.11

For volatility structure in equation (5.25)

\[
\begin{align*}
\sigma_0(t, x) &\int_0^x \sigma_0(t, u)du \\
&= h_t \left[ \sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax} \right] \left( a\sigma_2 + \sigma_3 \right) \\
&\quad + h_t \left[ \sigma_0^2 x^2 - \frac{a\sigma_1\sigma_2 + \sigma_3 e^{-ax}}{a^2} - \frac{a\sigma_0^2 e^{-2ax}}{2} + \frac{a\sigma_1 - \sigma_1 x e^{-ax}}{a} - \frac{1}{2} \sigma_0^2 x e^{-2ax} \right] \\
\end{align*}
\]

(A.1)

With \( G(z, x) \) following equation (5.26), we have

\[
\begin{align*}
G_x(x, z) &= \left[ \sigma_1 + (\sigma_2 + \sigma_3 x)e^{-ax} \right] (-a)z_1 + a\sigma_1 z_1 + \sigma_3 z_1 e^{-ax} + z_3 + z_4(-a)e^{-ax} \\
&\quad + (-2a)z_5 e^{-2ax} + \sigma_0^2 e^{-2ax} + z_6(-a) x e^{-ax} + z_7 e^{-2ax} + z_7(-2a) e^{-2ax} \\
&\quad + z_8 2 x e^{-ax} + z_8(-a) x^2 e^{-ax} + z_9 2 x e^{-2ax} + z_0(-2a) x^2 e^{-2ax} \\
\end{align*}
\]

(A.2)

By Theorem 4.1, in a similar way to the proof of Proposition 5.4, we have Proposition 5.11

95
A.3 A Parsimonious ANS Model with USV

An even more parsimonious Augmented Nelson-Siegel model can be obtained by setting $a_1 = a_2$ in Proposition 5.16. Under the risk-neutral probability measure, we specify a three-factor HJM model with forward rate volatility structure as

$$\sigma_1(t, x) = \sigma_1 \sqrt{v_1(t)}$$  \hspace{1cm} (A.3)

$$\sigma_2(t, x) = \sigma_2 e^{-ax} \sqrt{v_2(t)}$$ \hspace{1cm} (A.4)

$$\sigma_3(t, x) = \sigma_3 e^{-ax} \sqrt{v_3(t)}$$ \hspace{1cm} (A.5)

and

$$dv_i(t) = \kappa_i(\theta_i - v_i(t))dt + \sigma_{vi}(\sqrt{v_i(t)}(\rho_i dW_i(t) + \sqrt{1 - \rho_i^2} dB_i(t)))$$ \hspace{1cm} (A.6)

Here the exponential decaying speed is the same for slope and curvature component.

**Proposition A.1** For a three-factor HJM model with forward rate volatility in equations (A.3) through (A.5), with consistent the condition of affine parameterization, we have the instantaneous forward rate process as

$$f(t, x) = G(z, x)$$

$$= z_{1,1} + z_{1,2}x + z_{2,1}e^{-ax} + z_{2,2}e^{-2ax} + z_{3,1}xe^{-ax} + z_{3,2}xe^{-2ax} + z_{3,3}x^2e^{-2ax}$$

$$= (z_{1,1} + z_{2,1}e^{-ax} + z_{3,1}xe^{-ax}) + (z_{1,2}x + z_{2,2}e^{-2ax} + z_{3,2}xe^{-2ax} + z_{3,3}x^2e^{-2ax}$$

$$= NS + AugTerm$$

with $NS = z_{1,1} + z_{2,1}e^{-ax} + z_{3,1}xe^{-ax}$ and $AugTerm = z_{1,2}x + z_{2,2}e^{-2ax} + z_{3,2}xe^{-2ax} + z_{3,3}x^2e^{-2ax}$, and state variable process as

$$dz_1(t) = \begin{pmatrix} dz_{1,1}(t) \\ dz_{1,2}(t) \end{pmatrix} = \begin{pmatrix} z_{1,2}(t) \\ \sigma_1 \sqrt{v_1(t)} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 \sqrt{v_1(t)} \\ 0 \end{pmatrix} dW_1(t)$$ \hspace{1cm} (A.7)

$$dz_2(t) = \begin{pmatrix} dz_{2,1}(t) \\ dz_{2,2}(t) \end{pmatrix}$$

$$= \begin{pmatrix} -a z_{2,1}(t) + z_{3,1}(t) + \frac{\sigma_2^2}{a} v_2(t) \\ -2 a z_{2,2}(t) + z_{3,2}(t) - \frac{\sigma_2^2}{a} v_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_2 \sqrt{v_2(t)} \\ 0 \end{pmatrix} dW_2(t)$$ \hspace{1cm} (A.8)

$$dz_3(t) = \begin{pmatrix} dz_{3,1}(t) \\ dz_{3,2}(t) \\ dz_{3,3}(t) \end{pmatrix}$$

$$= \begin{pmatrix} -a z_{3,1}(t) + \frac{\sigma_3^2}{a^2} v_3(t) \\ -2 a z_{3,2}(t) + 2 z_{3,3}(t) - \frac{\sigma_3^2}{a} v_3(t) \\ -2 a z_{3,3}(t) - \frac{\sigma_3^2}{a} v_3(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_3 \sqrt{v_3(t)} \\ 0 \end{pmatrix} dW_3(t)$$ \hspace{1cm} (A.9)

and $x = (T - t)$, initial forward curve $G(z_0, x) = f(0, x)$, and, $dW_i(t)dW_j(t) = 0$, for any $i$ and $j$.

Compared with Proposition 5.16, in proposition A.1, the state space dimension is two less and the state variable $dZ_2(t)$ has different form since here it covers two extra terms.
from \(dZ_3(t)\).

### A.4 Conditional Mean and Variance

SDE of \(z_t\) implies

\[
    z_T = z_t + \int_t^T (a + Az_s)ds + \int_t^T \sigma(z_s)dW_s
\]

(A.10)

then

\[
    E_t[z_T] = z_t + \int_t^T (a + AE_T[z_s])ds
\]

(A.11)

or, by assuming \(E_t[z_T] = \hat{z}_{T|t}\), we have

\[
    \hat{z}_{T|t} = z_t + \int_t^T (a + A\hat{z}_{T|s})ds
\]

(A.12)

Differentiating equation (A.12) with respect to \(T\) yields

\[
    \hat{z}'_{T|t} = a + A\hat{z}_{T|t}
\]

(A.13)

that is

\[
    \hat{z}'_{T|t} = a + A\hat{z}_{T|t} = a + Am_0(T - t) + Am_1(T - t)z_t
\]

(A.14)

Differentiating equation (9.6) with respect to \(T\) yields

\[
    \hat{z}'_{T|t} = m'_0(T - t) + m'_1(T - t)z_t
\]

(A.15)

where, \(m'_0(T - t)\) and \(m'_1(T - t)\) are derivatives with respect to \((T - t)\) or \(T\). By comparing equations (A.14) and (9.9), we obtain ODEs that coefficients of conditional mean should satisfy,

\[
    m'_0(T - t) = a + Am_0(T - t)
\]

(A.16)

and

\[
    m'_1(T - t) = Am_1(T - t)
\]

(A.17)

with initial conditions \(m_0(0) = 0\) and \(m_1(0) = I_d\). They have solutions as

\[
    m_1(T - t) = e^{A(T - t)}
\]

(A.18)

and

\[
    m_0(T - t) = \left( \int_t^T e^{A(s-t)}ds \right) a
\]

(A.19)

Differentiate equation (A.19) reveals

\[
    m'_0(T - t) = m_1(T - t)a
\]

(A.20)

Equation (A.20) can also be shown from the fact that equation (9.6) is a martingale. The derivation of conditional variance is a bit involved. We have to use some matrix operations. Applying Ito’s lemma to equation (9.6) brings us

\[
    d\hat{z}_{T|t} = - (m'_0(T - t) + m'_1(T - t)z_t)dt + m_1(T - t)dz_t
\]

\[
    = - ((a + Am_0(T - t)) - Am_1(T - t)z_t + m_1(T - t)(a + Az_t))dt
\]

\[
    + m_1(T - t)\sigma(z_t)dW_t
\]

\[
    = m_1(T - t)\sigma(z_t)dW_t
\]

(A.21)
with $A m_1(T - t) = m_1(T - t)A$ owing to property of $e^{A(T-t)}A = A e^{A(T-t)}$. Then we have
\[
\dot{z}_{T|t} = \dot{z}_{T|t} + \int_t^T m_1(T - s)\sigma(z_s)dW_s
\] (A.22)
and the conditional variance becomes
\[
Var_t[z_T] = E_t \left[ \int_t^T m_1(T - s)\sigma(z_s)\sigma^T(z_s)m_1^T(T - s)ds \right]
= \int_t^T m_1(T - s)E_t \left[ \sigma(z_s)\sigma^T(z_s) \right] m_1^T(T - s)ds
\] (A.23)
By $\text{vec}$ operation,
\[
\text{vec}(Var_t[z_T]) = \int_t^T \text{vec} \left[ m_1(T - s)E_t \left[ \sigma(z_s)\sigma^T(z_s) \right] m_1^T(T - s) \right] ds
= \int_t^T m_1(T - s) \sigma(z_s) \sigma^T(z_s) \text{vec} \left[ G^0 + G^*\dot{z}_{s|t} \right] ds
\] (A.24)
With $G = \text{vec}(G^*) = (\text{vec}(G^1), \ldots, \text{vec}(G^d))$ and $\text{vec}(Var_t[z_T]) = \omega_0(T - t) + \omega_1(T - t)z_t$
\[
\omega_0(T - t) = \int_t^T m_1(T - s) \sigma(z_s) \sigma^T(z_s) \text{vec}(G^0) + G m_0(s - t) ds
\] (A.25)
\[
\omega_1(T - t) = \int_t^T m_1(T - s) \sigma(z_s) \sigma^T(z_s) m_1 G_m(s - t) ds
\] (A.26)
Differentiate equation (A.25) with respect to $T$,
\[
\omega_0'(T - t) = m_1(0) \sigma(z_s) \sigma^T(z_s) \text{vec}(G^0) + G m_0(T - t)
+ \int_t^T \left[ m_1'(T - s) \sigma(z_s) \sigma^T(z_s) \right] \text{vec}(G^0) + G m_0(s - t) ds
\]
\[
= m_1(0) \sigma(z_s) \sigma^T(z_s) \text{vec}(G^0) + G m_0(T - t)
+ \int_t^T \left[ A m_1(T - s) \sigma(z_s) \sigma^T(z_s) \right] \text{vec}(G^0) + G m_0(s - t) ds
\]
with Kronecker product $AB \otimes CD = (A \otimes C)(B \otimes D)$ and $m_1(0) = I_d$. Thus,
\[
\omega_0'(T - t) = \text{vec}(G^0) + G m_0(T - t) + (A \otimes I_d + I_d \otimes A)\omega_0(T - t)
\] (A.27)
Similarly, for equation (A.26), we can have
\[
\omega_1'(T - t) = G m_1(T - t) + (A \otimes I_d + I_d \otimes A)\omega_1(T - t)
\] (A.28)
Equations (A.16), (A.17), (A.27), and (A.28) form the system of ODEs that conditional mean and conditional variance should satisfy.

### A.5 State Variable Process under P-measure for ANS model

With the market price of risk specified in Section 8.5 in equations (8.10) and (8.11)
\[
dW_{i}(t) = dW_{i}^{P}(t) + \Lambda_{W_{i}}(t)dt
\]
\[ \Lambda_{W_i}(t) = \frac{\lambda_{W_i,a} + \lambda_{W_i,z_{i,1}}(t) + \lambda_{W_i,v_{i}}(t)}{\sqrt{v_i(t)}} \]

and consistent augmented Nelson-Siegel model with unspanned stochastic volatility in Proposition 5.16, we can derive the state process under physical measure. As noted before, change of measure only affects the state variables with diffusion terms, which include \( z_{1,1}(t), z_{2,1}(t) \) and \( z_{3,1}(t) \). Then, we have

\[ dz_{1,1}(t) = \left( \theta_{z_{1,1}}^P + \kappa_{z_{1,1}}^P z_{1,1}(t) + \kappa_{z_{1,1,1}}^P v_{1}(t) + z_{1,2}(t) \right) dt + \sqrt{v_{1}(t)} dW_{1}^P(t) \quad (A.29) \]

with

\[ \theta_{z_{1,1}}^P = \sigma_1 \lambda_{W_1,0} \quad (A.30) \]
\[ \kappa_{z_{1,1}}^P = \sigma_1 \lambda_{W_1,z_{1,1}} \quad (A.31) \]
\[ \kappa_{z_{1,1,1}}^P = \sigma_1 \lambda_{W_1,v_{1}} \quad (A.32) \]

and, for \( i = 2, 3 \),

\[ dz_{i,1}(t) = \left( \theta_{z_{i,1}}^P + \kappa_{z_{i,1}}^P z_{i,1}(t) + \kappa_{z_{i,1,1}}^P v_{i}(t) \right) dt + \sqrt{v_{i}(t)} dW_{i}^P(t) \quad (A.33) \]

and

\[ \theta_{z_{2,1}}^P = \sigma_2 \lambda_{W_2,0} \quad (A.34) \]
\[ \kappa_{z_{2,1}}^P = \sigma_2 \lambda_{W_2,z_{2,1}} - a_1 \quad (A.35) \]
\[ \kappa_{z_{2,1,1}}^P = \frac{\sigma_2^2}{a_1} + \sigma_2 \lambda_{W_2,v_{2}} \quad (A.36) \]
\[ \theta_{z_{3,1}}^P = \sigma_3 \lambda_{W_3,0} \quad (A.37) \]
\[ \kappa_{z_{3,1}}^P = \sigma_3 \lambda_{W_3,z_{3,1}} - a_2 \quad (A.38) \]
\[ \kappa_{z_{3,1,1}}^P = \frac{\sigma_3^2}{a_2} + \sigma_3 \lambda_{W_3,v_{3}} \quad (A.39) \]

The volatility dynamics \( dv_{i}(t) \) under physical measure is the same as those in Section 8.5 since it is not affected by \( dz_{i,1}(t) \).
Table 1: Descriptive Statistics of Zero-coupon Treasury Yields

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>1.84</td>
<td>1.32</td>
<td>1.70</td>
<td>5.12</td>
<td>0.001</td>
</tr>
<tr>
<td>6 month</td>
<td>1.98</td>
<td>1.58</td>
<td>1.73</td>
<td>5.24</td>
<td>0.03</td>
</tr>
<tr>
<td>1 year</td>
<td>2.11</td>
<td>1.74</td>
<td>1.65</td>
<td>5.20</td>
<td>0.01</td>
</tr>
<tr>
<td>2 year</td>
<td>2.39</td>
<td>2.04</td>
<td>1.49</td>
<td>5.16</td>
<td>0.20</td>
</tr>
<tr>
<td>3 year</td>
<td>2.66</td>
<td>2.45</td>
<td>1.35</td>
<td>5.14</td>
<td>0.33</td>
</tr>
<tr>
<td>5 year</td>
<td>3.19</td>
<td>3.20</td>
<td>1.09</td>
<td>5.11</td>
<td>0.85</td>
</tr>
<tr>
<td>7 year</td>
<td>3.58</td>
<td>3.65</td>
<td>0.90</td>
<td>5.19</td>
<td>1.35</td>
</tr>
<tr>
<td>10 year</td>
<td>3.94</td>
<td>4.02</td>
<td>0.73</td>
<td>5.31</td>
<td>1.86</td>
</tr>
<tr>
<td>20 year</td>
<td>4.58</td>
<td>4.65</td>
<td>0.60</td>
<td>5.91</td>
<td>2.61</td>
</tr>
</tbody>
</table>

The zero-coupon Treasury yields are calculated with the Constant Maturity Treasury yields. The values are in percentage.

Table 2: Descriptive Statistics of Constant Maturity Treasury Yields

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>1.85</td>
<td>1.33</td>
<td>1.72</td>
<td>5.19</td>
<td>0.01</td>
</tr>
<tr>
<td>6 month</td>
<td>2.00</td>
<td>1.59</td>
<td>1.75</td>
<td>5.31</td>
<td>0.03</td>
</tr>
<tr>
<td>1 year</td>
<td>2.12</td>
<td>1.75</td>
<td>1.67</td>
<td>5.27</td>
<td>0.10</td>
</tr>
<tr>
<td>2 year</td>
<td>2.41</td>
<td>2.06</td>
<td>1.51</td>
<td>5.23</td>
<td>0.20</td>
</tr>
<tr>
<td>3 year</td>
<td>2.68</td>
<td>2.47</td>
<td>1.37</td>
<td>5.21</td>
<td>0.33</td>
</tr>
<tr>
<td>5 year</td>
<td>3.21</td>
<td>3.23</td>
<td>1.10</td>
<td>5.18</td>
<td>0.85</td>
</tr>
<tr>
<td>7 year</td>
<td>3.61</td>
<td>3.68</td>
<td>0.91</td>
<td>5.26</td>
<td>1.35</td>
</tr>
<tr>
<td>10 year</td>
<td>3.98</td>
<td>4.06</td>
<td>0.74</td>
<td>5.38</td>
<td>1.87</td>
</tr>
<tr>
<td>20 year</td>
<td>4.63</td>
<td>4.70</td>
<td>0.62</td>
<td>6.00</td>
<td>2.63</td>
</tr>
</tbody>
</table>

The values are in percentage.
Table 3: Descriptive Statistics of Differences between Constant Maturity Treasury Yields and Zero-coupon Treasury Yields

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>1.577</td>
<td>0.437</td>
<td>2.16</td>
<td>6.62</td>
<td>2.5e-05</td>
</tr>
<tr>
<td>6 month</td>
<td>1.733</td>
<td>0.625</td>
<td>2.28</td>
<td>6.93</td>
<td>2.2e-04</td>
</tr>
<tr>
<td>1 year</td>
<td>1.799</td>
<td>0.757</td>
<td>2.19</td>
<td>6.82</td>
<td>2.5e-03</td>
</tr>
<tr>
<td>2 year</td>
<td>1.989</td>
<td>1.05</td>
<td>2.03</td>
<td>6.72</td>
<td>1.0e-02</td>
</tr>
<tr>
<td>3 year</td>
<td>2.235</td>
<td>1.51</td>
<td>1.93</td>
<td>6.67</td>
<td>2.7e-02</td>
</tr>
<tr>
<td>5 year</td>
<td>2.850</td>
<td>2.57</td>
<td>1.73</td>
<td>6.59</td>
<td>0.18</td>
</tr>
<tr>
<td>7 year</td>
<td>3.422</td>
<td>3.34</td>
<td>1.56</td>
<td>6.79</td>
<td>0.45</td>
</tr>
<tr>
<td>10 year</td>
<td>4.034</td>
<td>4.07</td>
<td>1.38</td>
<td>7.11</td>
<td>0.87</td>
</tr>
<tr>
<td>20 year</td>
<td>5.374</td>
<td>5.44</td>
<td>1.34</td>
<td>8.82</td>
<td>1.71</td>
</tr>
</tbody>
</table>

The difference is calculated from the Constant Maturity Treasury yield minus its corresponding zero-coupon bond yield. The values are expressed in basis point.

Table 4: Descriptive Statistics of LIBOR/Swap Rates

<table>
<thead>
<tr>
<th>Tenor</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>2.3175</td>
<td>1.8149</td>
<td>5.7250</td>
<td>0.2458</td>
<td>1.8031</td>
</tr>
<tr>
<td>6 month</td>
<td>2.4587</td>
<td>1.7561</td>
<td>5.6300</td>
<td>0.3844</td>
<td>1.9053</td>
</tr>
<tr>
<td>9 month</td>
<td>2.5626</td>
<td>1.6910</td>
<td>5.7000</td>
<td>0.5568</td>
<td>2.0969</td>
</tr>
<tr>
<td>1 year</td>
<td>2.4990</td>
<td>1.7447</td>
<td>5.7460</td>
<td>0.3590</td>
<td>2.2010</td>
</tr>
<tr>
<td>2 year</td>
<td>2.8270</td>
<td>1.5624</td>
<td>5.7280</td>
<td>0.4440</td>
<td>2.6640</td>
</tr>
<tr>
<td>3 year</td>
<td>3.1597</td>
<td>1.4080</td>
<td>5.7230</td>
<td>0.5965</td>
<td>3.0873</td>
</tr>
<tr>
<td>4 year</td>
<td>3.4432</td>
<td>1.2789</td>
<td>5.7330</td>
<td>0.8660</td>
<td>3.4855</td>
</tr>
<tr>
<td>5 year</td>
<td>3.6806</td>
<td>1.1705</td>
<td>5.7465</td>
<td>1.1330</td>
<td>3.7623</td>
</tr>
<tr>
<td>7 year</td>
<td>4.0359</td>
<td>1.0164</td>
<td>5.7800</td>
<td>1.5960</td>
<td>4.1903</td>
</tr>
<tr>
<td>10 year</td>
<td>4.3665</td>
<td>0.9015</td>
<td>6.0700</td>
<td>2.0180</td>
<td>4.5440</td>
</tr>
<tr>
<td>15 year</td>
<td>4.6722</td>
<td>0.8322</td>
<td>6.3200</td>
<td>2.4070</td>
<td>4.8993</td>
</tr>
</tbody>
</table>

The values are in percentage.
Table 5: Estimates of Parameters in General Volatility Models with Yields

<table>
<thead>
<tr>
<th></th>
<th>GV1F</th>
<th>GV2F</th>
<th>GV3F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{i,1}$</td>
<td>0.3615</td>
<td>0.0918</td>
<td>0.0274</td>
</tr>
<tr>
<td>$\sigma_{i,2}$</td>
<td>0.1138</td>
<td>0.0188</td>
<td>0.0467</td>
</tr>
<tr>
<td>$\sigma_{i,3}$</td>
<td>0.0231</td>
<td>0.0604</td>
<td>0.0174</td>
</tr>
<tr>
<td>$\sigma_{i,v}$</td>
<td>1.1661</td>
<td>0.6032</td>
<td>0.9403</td>
</tr>
<tr>
<td>$\kappa_{p,z}$</td>
<td>-1.1941</td>
<td>-1.9737</td>
<td>-1.6811</td>
</tr>
<tr>
<td>$\kappa_{p,v}$</td>
<td>1.4988</td>
<td>1.9687</td>
<td>1.3609</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>0.4761</td>
<td>0.8109</td>
<td>0.5347</td>
</tr>
<tr>
<td>$\phi_{z}$</td>
<td>0.3299</td>
<td>1.1433</td>
<td>0.6748</td>
</tr>
<tr>
<td>$\lambda_{t}$</td>
<td>0.4431</td>
<td>0.3271</td>
<td>0.5352</td>
</tr>
<tr>
<td>$\tau_r$</td>
<td>0.0036</td>
<td>0.0253</td>
<td>0.2440</td>
</tr>
<tr>
<td>loglikelihood</td>
<td>1.8366e+04</td>
<td>2.1593e+04</td>
<td>2.4185e+04</td>
</tr>
</tbody>
</table>

The values below the estimates of the parameters are the standard errors calculated from
the information matrix of the likelihood function.
Table 6: Estimates of Parameters in Augmented Nelson-Siegel Model with Yields

<table>
<thead>
<tr>
<th></th>
<th>Augmented Nelson-Siegel</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$i=1$</td>
<td>$i=2$</td>
<td>$i=3$</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1451</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0806</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0493</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td></td>
<td>0.1262</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{i,v}$</td>
<td>1.0530</td>
<td>0.3863</td>
<td>1.4242</td>
</tr>
<tr>
<td></td>
<td>0.0807</td>
<td>0.0870</td>
<td>0.0855</td>
</tr>
<tr>
<td>$a_i$</td>
<td>0.6646</td>
<td>0.4003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0836</td>
<td>0.0429</td>
<td></td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>0.0818</td>
<td>0.1111</td>
<td>-0.1046</td>
</tr>
<tr>
<td></td>
<td>0.0790</td>
<td>0.0925</td>
<td>0.0852</td>
</tr>
<tr>
<td>$\theta_{z_i}^p$</td>
<td>0.4452</td>
<td>0.4531</td>
<td>0.0853</td>
</tr>
<tr>
<td></td>
<td>0.0905</td>
<td>0.0871</td>
<td>0.0871</td>
</tr>
<tr>
<td>$\kappa_{z_i}^p$</td>
<td>-1.2613</td>
<td>-0.2641</td>
<td>-1.6352</td>
</tr>
<tr>
<td></td>
<td>0.0872</td>
<td>0.0921</td>
<td>0.0965</td>
</tr>
<tr>
<td>$\kappa_{z_i,v_i}^p$</td>
<td>-0.6984</td>
<td>-1.0882</td>
<td>-0.6183</td>
</tr>
<tr>
<td></td>
<td>0.0820</td>
<td>0.0875</td>
<td>0.0813</td>
</tr>
<tr>
<td>$\kappa_{v_i}^p$</td>
<td>5.1106</td>
<td>2.0291</td>
<td>2.8159</td>
</tr>
<tr>
<td></td>
<td>0.0904</td>
<td>0.0881</td>
<td>0.0825</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>8.228e-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.325e-05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$log$ $likelihood$ 2.2655e+04

The values below the estimates of the parameters are the standard errors calculated from the information matrix of the likelihood function.
Table 7: Estimates of parameters in General Volatility Models with LIBOR/Swap Rates

<table>
<thead>
<tr>
<th></th>
<th>GV1F i=1</th>
<th>GV1F i=2</th>
<th>GV2F i=1</th>
<th>GV2F i=2</th>
<th>GV2F i=3</th>
<th>GV3F i=1</th>
<th>GV3F i=2</th>
<th>GV3F i=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{i,1})</td>
<td>0.0919</td>
<td>0.0615</td>
<td>0.1194</td>
<td>0.0751</td>
<td>0.1086</td>
<td>0.1044</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0023</td>
<td>0.0078</td>
<td>0.0063</td>
<td>0.0061</td>
<td>0.0071</td>
<td>0.0067</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_{i,2})</td>
<td>0.0042</td>
<td>0.0686</td>
<td>0.0552</td>
<td>0.0637</td>
<td>0.0550</td>
<td>0.1010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0047</td>
<td>0.0081</td>
<td>0.0106</td>
<td>0.0093</td>
<td>0.0100</td>
<td>0.0126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_{i,3})</td>
<td>0.0686</td>
<td>0.0427</td>
<td>0.0573</td>
<td>0.0388</td>
<td>0.0581</td>
<td>0.0782</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0212</td>
<td>0.0057</td>
<td>0.0102</td>
<td>0.0074</td>
<td>0.0103</td>
<td>0.0097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_{i,v})</td>
<td>0.4356</td>
<td>1.2168</td>
<td>0.8446</td>
<td>1.0838</td>
<td>0.7865</td>
<td>0.6704</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0731</td>
<td>0.0580</td>
<td>0.0415</td>
<td>0.0739</td>
<td>0.0400</td>
<td>0.0300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a_i)</td>
<td>2.3940</td>
<td>0.9354</td>
<td>0.9097</td>
<td>1.0420</td>
<td>0.8046</td>
<td>0.9971</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2309</td>
<td>0.0170</td>
<td>0.0201</td>
<td>0.0629</td>
<td>0.0291</td>
<td>0.0418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho_i)</td>
<td>0.2891</td>
<td>0.2710</td>
<td>0.4074</td>
<td>0.3444</td>
<td>0.3551</td>
<td>0.2876</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4864</td>
<td>0.0583</td>
<td>0.0618</td>
<td>0.0560</td>
<td>0.0512</td>
<td>0.0579</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\theta_{zi}^p)</td>
<td>0.6320</td>
<td>0.0644</td>
<td>0.0495</td>
<td>0.0594</td>
<td>0.0517</td>
<td>0.0377</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2044</td>
<td>0.0662</td>
<td>0.0571</td>
<td>0.1430</td>
<td>0.0844</td>
<td>0.1087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa_{zi}^p)</td>
<td>-2.1845</td>
<td>-1.5215</td>
<td>-1.2671</td>
<td>-1.498</td>
<td>-1.2778</td>
<td>-1.1272</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5372</td>
<td>0.0537</td>
<td>0.0565</td>
<td>0.1257</td>
<td>0.1233</td>
<td>0.0786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa_{zi,v}^p)</td>
<td>-0.6338</td>
<td>0.3423</td>
<td>1.1990</td>
<td>0.3465</td>
<td>1.0398</td>
<td>0.3510</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2068</td>
<td>0.0483</td>
<td>0.0549</td>
<td>0.0959</td>
<td>0.1173</td>
<td>0.1161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa_{v,i}^p)</td>
<td>1.7625</td>
<td>4.5534</td>
<td>2.3845</td>
<td>4.0151</td>
<td>2.4748</td>
<td>2.8411</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1745</td>
<td>0.0580</td>
<td>0.0883</td>
<td>0.1241</td>
<td>0.0784</td>
<td>0.0792</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\sigma_r)</td>
<td>1.526e-05</td>
<td>3.086e-06</td>
<td>8.910e-07</td>
<td>2.329e-07</td>
<td>6.609e-08</td>
<td>9.515e-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loglikelihood</td>
<td>2.3058e+04</td>
<td>2.6574e+04</td>
<td>2.8248e+04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The values below the estimates of the parameters are the standard errors calculated from the information matrix of the likelihood function.
Table 8: Estimates of Parameters in Augmented Nelson-Siegel Model with LIBOR/Swap Rates

<table>
<thead>
<tr>
<th>Augmented Nelson-Siegel</th>
<th>$i=1$</th>
<th>$i=2$</th>
<th>$i=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{i,1}$</td>
<td>0.0742</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{i,2}$</td>
<td>0.0138</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{i,3}$</td>
<td>0.0853</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{i,v}$</td>
<td>1.2036</td>
<td>0.4888</td>
<td>0.6084</td>
</tr>
<tr>
<td></td>
<td>0.3809</td>
<td>0.2436</td>
<td>0.04631</td>
</tr>
<tr>
<td>$a_i$</td>
<td>1.0432</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0804</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>0.1509</td>
<td>0.1343</td>
<td>0.4285</td>
</tr>
<tr>
<td></td>
<td>0.2784</td>
<td>0.3117</td>
<td>0.1146</td>
</tr>
<tr>
<td>$\rho_{z_i}$</td>
<td>0.2092</td>
<td>0.4675</td>
<td>0.1824</td>
</tr>
<tr>
<td></td>
<td>0.0319</td>
<td>0.2484</td>
<td>0.0315</td>
</tr>
<tr>
<td>$\kappa_{z_i}$</td>
<td>-0.7196</td>
<td>-1.9289</td>
<td>-1.0061</td>
</tr>
<tr>
<td></td>
<td>0.4819</td>
<td>0.3433</td>
<td>0.3315</td>
</tr>
<tr>
<td>$\kappa_{z_{i,v}}$</td>
<td>-0.7927</td>
<td>-0.4784</td>
<td>-0.5523</td>
</tr>
<tr>
<td></td>
<td>0.2249</td>
<td>0.2324</td>
<td>0.1350</td>
</tr>
<tr>
<td>$\kappa_{v_i}$</td>
<td>0.7243</td>
<td>1.7987</td>
<td>2.8017</td>
</tr>
<tr>
<td></td>
<td>0.2232</td>
<td>0.4783</td>
<td>0.3124</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>8.052e-07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.14e-08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$loglikelihood$ 2.8755e+04

The values below the estimates of the parameters are the standard errors calculated from the information matrix of the likelihood function.
Table 9: Root-mean-square Errors of Fitted Yields

<table>
<thead>
<tr>
<th></th>
<th>GV1F</th>
<th>GV2F</th>
<th>GV3F</th>
<th>ANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>50.68</td>
<td>11.75</td>
<td>6.76</td>
<td>5.47</td>
</tr>
<tr>
<td>6 month</td>
<td>42.04</td>
<td>7.54</td>
<td>6.44</td>
<td>8.02</td>
</tr>
<tr>
<td>1 year</td>
<td>24.07</td>
<td>8.59</td>
<td>4.88</td>
<td>3.57</td>
</tr>
<tr>
<td>2 year</td>
<td>20.05</td>
<td>9.68</td>
<td>4.77</td>
<td>5.41</td>
</tr>
<tr>
<td>3 year</td>
<td>38.65</td>
<td>9.89</td>
<td>5.28</td>
<td>6.07</td>
</tr>
<tr>
<td>5 year</td>
<td>55.90</td>
<td>7.76</td>
<td>3.44</td>
<td>3.79</td>
</tr>
<tr>
<td>7 year</td>
<td>54.98</td>
<td>10.07</td>
<td>6.06</td>
<td>6.15</td>
</tr>
<tr>
<td>10 year</td>
<td>32.07</td>
<td>10.25</td>
<td>4.71</td>
<td>3.98</td>
</tr>
<tr>
<td>20 year</td>
<td>71.95</td>
<td>13.35</td>
<td>6.82</td>
<td>3.79</td>
</tr>
<tr>
<td>Average</td>
<td>46.14</td>
<td>10.03</td>
<td>5.57</td>
<td>5.33</td>
</tr>
</tbody>
</table>

Root-mean-square errors of fitted yields are calculated from the difference between actual yields and the model fitted values of the yields. Each row, except the last one, is the root-mean-square error for the yield with its corresponding time-to-maturity. The average is the root-mean-square error of the whole sample. All values are expressed as basis point.

Table 10: Root-mean-square Errors of Forecast Yields

<table>
<thead>
<tr>
<th></th>
<th>GV1F</th>
<th>GV2F</th>
<th>GV3F</th>
<th>ANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>70.81</td>
<td>51.56</td>
<td>24.48</td>
<td>26.40</td>
</tr>
<tr>
<td>6 month</td>
<td>59.77</td>
<td>43.66</td>
<td>18.00</td>
<td>11.93</td>
</tr>
<tr>
<td>1 year</td>
<td>45.62</td>
<td>45.41</td>
<td>20.32</td>
<td>10.49</td>
</tr>
<tr>
<td>2 year</td>
<td>35.39</td>
<td>48.67</td>
<td>24.30</td>
<td>17.13</td>
</tr>
<tr>
<td>3 year</td>
<td>42.16</td>
<td>50.19</td>
<td>25.26</td>
<td>22.20</td>
</tr>
<tr>
<td>5 year</td>
<td>49.40</td>
<td>43.51</td>
<td>21.20</td>
<td>22.50</td>
</tr>
<tr>
<td>7 year</td>
<td>46.65</td>
<td>37.92</td>
<td>19.57</td>
<td>22.51</td>
</tr>
<tr>
<td>10 year</td>
<td>35.02</td>
<td>37.22</td>
<td>18.94</td>
<td>28.38</td>
</tr>
<tr>
<td>20 year</td>
<td>107.95</td>
<td>48.82</td>
<td>29.81</td>
<td>40.65</td>
</tr>
<tr>
<td>Average</td>
<td>59.31</td>
<td>45.48</td>
<td>22.72</td>
<td>24.05</td>
</tr>
</tbody>
</table>

Root-mean-square errors of forecast yields are calculated from the difference between the actual yields and the one-period ahead forecast values of the yields. Each row, except the last one, is the root-mean-square error for the yield with its corresponding time-to-maturity. The average is the root-mean-square error of the whole sample. All values are expressed as basis point.
Table 11: Root-mean-square Errors of Fitted LIBOR/Swap Rates

<table>
<thead>
<tr>
<th></th>
<th>GV1F</th>
<th>GV2F</th>
<th>GV3F</th>
<th>ANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>33.68</td>
<td>14.37</td>
<td>11.31</td>
<td>9.20</td>
</tr>
<tr>
<td>6 month</td>
<td>28.18</td>
<td>9.93</td>
<td>8.00</td>
<td>8.97</td>
</tr>
<tr>
<td>9 month</td>
<td>25.32</td>
<td>13.98</td>
<td>15.38</td>
<td>13.19</td>
</tr>
<tr>
<td>1 year</td>
<td>24.96</td>
<td>16.76</td>
<td>10.99</td>
<td>12.09</td>
</tr>
<tr>
<td>2 year</td>
<td>21.51</td>
<td>15.65</td>
<td>6.86</td>
<td>7.43</td>
</tr>
<tr>
<td>3 year</td>
<td>26.45</td>
<td>7.89</td>
<td>3.97</td>
<td>3.96</td>
</tr>
<tr>
<td>4 year</td>
<td>31.20</td>
<td>6.41</td>
<td>4.15</td>
<td>3.56</td>
</tr>
<tr>
<td>5 year</td>
<td>33.02</td>
<td>10.64</td>
<td>4.64</td>
<td>4.16</td>
</tr>
<tr>
<td>7 year</td>
<td>28.60</td>
<td>14.82</td>
<td>4.98</td>
<td>6.21</td>
</tr>
<tr>
<td>10 year</td>
<td>24.01</td>
<td>7.97</td>
<td>4.13</td>
<td>4.55</td>
</tr>
<tr>
<td>15 year</td>
<td>68.80</td>
<td>24.86</td>
<td>13.75</td>
<td>8.21</td>
</tr>
<tr>
<td>Average</td>
<td>33.77</td>
<td>13.96</td>
<td>8.95</td>
<td>8.05</td>
</tr>
</tbody>
</table>

Root-mean-square errors of the fitted LIBOR and swap rates are calculated from the difference between the actual rates and the model fitted values of the LIBOR and swap rates. Each row, except the last one, is the root-mean-square error for the rate with its corresponding tenor. The average is the root-mean-square error of the whole sample. All values are expressed as basis point.

Table 12: Root-mean-square Errors of Forecast LIBOR/Swap Rates

<table>
<thead>
<tr>
<th></th>
<th>GV1F</th>
<th>GV2F</th>
<th>GV3F</th>
<th>ANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 month</td>
<td>34.46</td>
<td>22.09</td>
<td>24.15</td>
<td>19.24</td>
</tr>
<tr>
<td>6 month</td>
<td>27.60</td>
<td>16.97</td>
<td>15.08</td>
<td>12.01</td>
</tr>
<tr>
<td>9 month</td>
<td>24.75</td>
<td>20.68</td>
<td>16.98</td>
<td>18.09</td>
</tr>
<tr>
<td>1 year</td>
<td>30.86</td>
<td>34.56</td>
<td>24.73</td>
<td>19.30</td>
</tr>
<tr>
<td>2 year</td>
<td>28.84</td>
<td>34.44</td>
<td>24.31</td>
<td>19.35</td>
</tr>
<tr>
<td>3 year</td>
<td>30.81</td>
<td>27.03</td>
<td>22.29</td>
<td>17.56</td>
</tr>
<tr>
<td>4 year</td>
<td>33.53</td>
<td>21.23</td>
<td>21.23</td>
<td>16.44</td>
</tr>
<tr>
<td>5 year</td>
<td>34.61</td>
<td>18.24</td>
<td>20.65</td>
<td>16.00</td>
</tr>
<tr>
<td>7 year</td>
<td>31.92</td>
<td>17.56</td>
<td>21.02</td>
<td>16.20</td>
</tr>
<tr>
<td>10 year</td>
<td>35.71</td>
<td>22.11</td>
<td>25.29</td>
<td>16.81</td>
</tr>
<tr>
<td>15 year</td>
<td>81.10</td>
<td>46.88</td>
<td>37.34</td>
<td>21.38</td>
</tr>
<tr>
<td>Average</td>
<td>38.71</td>
<td>27.12</td>
<td>23.65</td>
<td>17.65</td>
</tr>
</tbody>
</table>

Root-mean-square errors of the forecast LIBOR and swap rates are calculated with the difference between the actual rates and the one-period ahead forecast values of the LIBOR and swap rates. Each row, except the last one, is the root-mean-square error for the rate with its corresponding tenor. The average is the root-mean-square error of the whole sample. All values are expressed as basis point.
Table 13: AIC and BIC for Model Comparison with Yields

<table>
<thead>
<tr>
<th>Model</th>
<th>Loglikelihood</th>
<th>Number of Parameters</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GV1F</td>
<td>18366</td>
<td>11</td>
<td>-36710</td>
<td>-36692</td>
</tr>
<tr>
<td>GV2F</td>
<td>21593</td>
<td>21</td>
<td>-43164</td>
<td>-43109</td>
</tr>
<tr>
<td>GV3F</td>
<td>24185</td>
<td>31</td>
<td>-48348</td>
<td>-48256</td>
</tr>
<tr>
<td>ANS</td>
<td>22655</td>
<td>24</td>
<td>-45288</td>
<td>-45222</td>
</tr>
</tbody>
</table>

AIC is calculated with formula $-2\text{Loglikelihood} + 2d$ and BIC is calculated with $-2\text{Loglikelihood} + \log(n) \ast d$, where $\text{Loglikelihood}$ is the value of likelihood function with estimates of parameters, $d$ is the number of parameters in the model.

Table 14: AIC and BIC for Model Comparison with LIBOR/Swap Rates

<table>
<thead>
<tr>
<th>Model</th>
<th>Loglikelihood</th>
<th>Number of Parameters</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GV1F</td>
<td>23058</td>
<td>11</td>
<td>-46094</td>
<td>-46075</td>
</tr>
<tr>
<td>GV2F</td>
<td>26574</td>
<td>21</td>
<td>-53126</td>
<td>-53069</td>
</tr>
<tr>
<td>GV3F</td>
<td>28248</td>
<td>31</td>
<td>-56474</td>
<td>-56380</td>
</tr>
<tr>
<td>ANS</td>
<td>28755</td>
<td>24</td>
<td>-57488</td>
<td>-57420</td>
</tr>
</tbody>
</table>

AIC is calculated with formula $-2\text{Loglikelihood} + 2d$ and BIC is calculated with $-2\text{Loglikelihood} + \log(n) \ast d$, where $\text{Loglikelihood}$ is the value of likelihood function with estimates of parameters, $d$ is the number of parameters in the model.
Table 15: Fitted State Variables from 3-factor General Volatility Model for Yields

<table>
<thead>
<tr>
<th>$Z_1$</th>
<th>$z_{1,1}$</th>
<th>$z_{1,2}$</th>
<th>$z_{1,3}$</th>
<th>$z_{1,4}$</th>
<th>$z_{1,5}$</th>
<th>$z_{1,6}$</th>
<th>$z_{1,7}$</th>
<th>$z_{1,8}$</th>
<th>$z_{1,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0555</td>
<td>0.1183</td>
<td>0.0025</td>
<td>0.0100</td>
<td>-0.0001</td>
<td>0.0006</td>
<td>-0.0002</td>
<td>0.0002</td>
<td>-8.28e-05</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>0.2332</td>
<td>0.0526</td>
<td>0.0005</td>
<td>0.0077</td>
<td>0.0003</td>
<td>0.0007</td>
<td>0.0002</td>
<td>0.0001</td>
<td>5.89e-05</td>
</tr>
<tr>
<td>Max</td>
<td>0.5344</td>
<td>0.2279</td>
<td>0.0049</td>
<td>0.0209</td>
<td>0.0014</td>
<td>0.0031</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0003</td>
</tr>
<tr>
<td>Min</td>
<td>-0.3907</td>
<td>0.0165</td>
<td>-5.43e-05</td>
<td>-0.0019</td>
<td>-0.0011</td>
<td>-0.0004</td>
<td>-0.0013</td>
<td>-0.0007</td>
<td>-0.0004</td>
</tr>
<tr>
<td>Median</td>
<td>0.0096</td>
<td>0.0993</td>
<td>0.0026</td>
<td>0.0109</td>
<td>-0.0002</td>
<td>0.0003</td>
<td>-0.0003</td>
<td>0.0002</td>
<td>-7.99e-05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Z_2$</th>
<th>$z_{2,1}$</th>
<th>$z_{2,2}$</th>
<th>$z_{2,3}$</th>
<th>$z_{2,4}$</th>
<th>$z_{2,5}$</th>
<th>$z_{2,6}$</th>
<th>$z_{2,7}$</th>
<th>$z_{2,8}$</th>
<th>$z_{2,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.1025</td>
<td>0.0019</td>
<td>-0.0027</td>
<td>-0.0136</td>
<td>0.0050</td>
<td>-0.0066</td>
<td>0.0033</td>
<td>-0.0021</td>
<td>-0.0003</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.2305</td>
<td>0.0359</td>
<td>0.0053</td>
<td>0.0159</td>
<td>0.0065</td>
<td>0.0013</td>
<td>0.0061</td>
<td>0.0010</td>
<td>0.0034</td>
</tr>
<tr>
<td>max</td>
<td>0.78437</td>
<td>0.0329</td>
<td>0.0041</td>
<td>0.0175</td>
<td>0.0208</td>
<td>-0.0049</td>
<td>0.0184</td>
<td>-0.0006</td>
<td>0.0039</td>
</tr>
<tr>
<td>min</td>
<td>-0.6461</td>
<td>-0.1002</td>
<td>-0.0116</td>
<td>-0.0563</td>
<td>-0.0062</td>
<td>-0.0098</td>
<td>-0.0081</td>
<td>-0.0048</td>
<td>-0.0162</td>
</tr>
<tr>
<td>median</td>
<td>-0.0820</td>
<td>0.0225</td>
<td>-0.0011</td>
<td>-0.0111</td>
<td>0.0061</td>
<td>-0.0063</td>
<td>0.0041</td>
<td>-0.0018</td>
<td>0.0008</td>
</tr>
</tbody>
</table>
Table 16: Fitted State Variables from 3-factor General Volatility Model for Yields–Continued

<table>
<thead>
<tr>
<th>$Z_3$</th>
<th>$z_{3,1}$</th>
<th>$z_{3,2}$</th>
<th>$z_{3,3}$</th>
<th>$z_{3,4}$</th>
<th>$z_{3,5}$</th>
<th>$z_{3,6}$</th>
<th>$z_{3,7}$</th>
<th>$z_{3,8}$</th>
<th>$z_{3,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.0362</td>
<td>-0.0718</td>
<td>0.0009</td>
<td>-0.0157</td>
<td>-0.0037</td>
<td>0.0041</td>
<td>-0.0042</td>
<td>0.0016</td>
<td>-0.0015</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.3009</td>
<td>0.0171</td>
<td>0.0052</td>
<td>0.0139</td>
<td>0.0038</td>
<td>0.0019</td>
<td>0.0035</td>
<td>0.0011</td>
<td>0.0020</td>
</tr>
<tr>
<td>max</td>
<td>0.7466</td>
<td>-0.0073</td>
<td>0.0099</td>
<td>0.0077</td>
<td>0.0016</td>
<td>0.0089</td>
<td>0.0013</td>
<td>0.0052</td>
<td>0.0007</td>
</tr>
<tr>
<td>min</td>
<td>-0.9727</td>
<td>-0.1005</td>
<td>-0.0057</td>
<td>-0.0411</td>
<td>-0.0166</td>
<td>0.0010</td>
<td>-0.0167</td>
<td>-0.0002</td>
<td>-0.0129</td>
</tr>
<tr>
<td>median</td>
<td>0.0168</td>
<td>-0.0743</td>
<td>-0.0010</td>
<td>-0.0145</td>
<td>-0.0031</td>
<td>0.0037</td>
<td>-0.0038</td>
<td>0.0013</td>
<td>-0.0010</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$V$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.2443</td>
<td>0.1586</td>
<td>0.3121</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.0515</td>
<td>0.0665</td>
<td>0.1032</td>
</tr>
<tr>
<td>max</td>
<td>0.4329</td>
<td>0.4119</td>
<td>0.6951</td>
</tr>
<tr>
<td>min</td>
<td>-0.0147</td>
<td>-0.0547</td>
<td>-0.0137</td>
</tr>
<tr>
<td>median</td>
<td>0.2454</td>
<td>0.1569</td>
<td>0.3007</td>
</tr>
</tbody>
</table>

$v_i$ should be positive. But due to the numerical process, there are several negative values generated. One way to deal with this issue is to truncate the value at zero as in Duffee (1999). Without truncating the value, this paper instead allows the optimizer to choose a nearby point to pass through the negative one.
Table 17: Fitted State Variables from 3-factor General Volatility Model for LIBOR/Swap Rates

<table>
<thead>
<tr>
<th>$Z_1$</th>
<th>$z_{1,1}$</th>
<th>$z_{1,2}$</th>
<th>$z_{1,3}$</th>
<th>$z_{1,4}$</th>
<th>$z_{1,5}$</th>
<th>$z_{1,6}$</th>
<th>$z_{1,7}$</th>
<th>$z_{1,8}$</th>
<th>$z_{1,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.5294</td>
<td>0.4143</td>
<td>0.0105</td>
<td>0.0281</td>
<td>-0.0011</td>
<td>0.0022</td>
<td>-0.0008</td>
<td>0.0007</td>
<td>-0.0002</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.7993</td>
<td>0.2138</td>
<td>0.0058</td>
<td>0.0215</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0003</td>
<td>0.0002</td>
<td>5.36e-05</td>
</tr>
<tr>
<td>max</td>
<td>1.8335</td>
<td>0.8929</td>
<td>0.0244</td>
<td>0.0614</td>
<td>0.0009</td>
<td>0.0037</td>
<td>0.0002</td>
<td>0.0013</td>
<td>-5.15e-05</td>
</tr>
<tr>
<td>min</td>
<td>-1.0076</td>
<td>0.0229</td>
<td>0.0024</td>
<td>-0.0082</td>
<td>-0.0020</td>
<td>0.0015</td>
<td>-0.0014</td>
<td>0.0004</td>
<td>-0.0003</td>
</tr>
<tr>
<td>median</td>
<td>0.6828</td>
<td>0.3448</td>
<td>0.0093</td>
<td>0.0343</td>
<td>-0.0010</td>
<td>0.0020</td>
<td>-0.0008</td>
<td>0.0007</td>
<td>-0.0002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$Z_2$</th>
<th>$z_{2,1}$</th>
<th>$z_{2,2}$</th>
<th>$z_{2,3}$</th>
<th>$z_{2,4}$</th>
<th>$z_{2,5}$</th>
<th>$z_{2,6}$</th>
<th>$z_{2,7}$</th>
<th>$z_{2,8}$</th>
<th>$z_{2,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.2440</td>
<td>-0.0083</td>
<td>-0.0059</td>
<td>-0.0106</td>
<td>0.0006</td>
<td>-0.0043</td>
<td>1.41e-05</td>
<td>-0.0013</td>
<td>-0.0008</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.3214</td>
<td>0.0838</td>
<td>0.0090</td>
<td>0.0133</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0020</td>
<td>0.0008</td>
<td>0.0026</td>
</tr>
<tr>
<td>max</td>
<td>1.0407</td>
<td>0.0647</td>
<td>0.0029</td>
<td>0.0170</td>
<td>0.0040</td>
<td>-0.0024</td>
<td>0.0057</td>
<td>-0.0004</td>
<td>0.0007</td>
</tr>
<tr>
<td>min</td>
<td>-0.9675</td>
<td>-0.2639</td>
<td>-0.0314</td>
<td>-0.0427</td>
<td>-0.0023</td>
<td>-0.0072</td>
<td>-0.0052</td>
<td>-0.0047</td>
<td>-0.0149</td>
</tr>
<tr>
<td>median</td>
<td>-0.2155</td>
<td>0.0342</td>
<td>-0.0039</td>
<td>-0.0080</td>
<td>0.0009</td>
<td>-0.0037</td>
<td>0.0007</td>
<td>-0.0009</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Table 18: Fitted State Variables from 3-factor General Volatility Model for LIBOR/Swap Rates–Continued

<table>
<thead>
<tr>
<th>$Z_3$</th>
<th>$z_{3,1}$</th>
<th>$z_{3,2}$</th>
<th>$z_{3,3}$</th>
<th>$z_{3,4}$</th>
<th>$z_{3,5}$</th>
<th>$z_{3,6}$</th>
<th>$z_{3,7}$</th>
<th>$z_{3,8}$</th>
<th>$z_{3,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.2084</td>
<td>-0.3547</td>
<td>-0.0032</td>
<td>-0.0354</td>
<td>-0.0004</td>
<td>0.0014</td>
<td>-0.0010</td>
<td>0.0005</td>
<td>-0.0007</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.5724</td>
<td>0.1333</td>
<td>0.0035</td>
<td>0.0298</td>
<td>0.0008</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0007</td>
<td>0.0018</td>
</tr>
<tr>
<td>max</td>
<td>0.7370</td>
<td>0.0229</td>
<td>0.0087</td>
<td>0.0111</td>
<td>0.0024</td>
<td>0.0059</td>
<td>0.0018</td>
<td>0.0038</td>
<td>0.0004</td>
</tr>
<tr>
<td>min</td>
<td>-1.2932</td>
<td>-0.5917</td>
<td>-0.0068</td>
<td>-0.0866</td>
<td>-0.0025</td>
<td>-0.0009</td>
<td>-0.0061</td>
<td>-0.0008</td>
<td>-0.0120</td>
</tr>
<tr>
<td>median</td>
<td>-0.2170</td>
<td>-0.3349</td>
<td>-0.0046</td>
<td>-0.0368</td>
<td>-0.0005</td>
<td>0.0007</td>
<td>-0.0006</td>
<td>0.0003</td>
<td>-0.0002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$V_1$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.3529</td>
<td>0.1539</td>
<td>0.2772</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.1228</td>
<td>0.0607</td>
<td>0.0682</td>
</tr>
<tr>
<td>max</td>
<td>0.8013</td>
<td>0.3973</td>
<td>0.4908</td>
</tr>
<tr>
<td>min</td>
<td>-0.2027</td>
<td>-0.0412</td>
<td>-0.0026</td>
</tr>
<tr>
<td>median</td>
<td>0.3566</td>
<td>0.1551</td>
<td>0.2767</td>
</tr>
</tbody>
</table>

$v_i$ should be positive. But due to the numerical process, there are several negative values generated. One way to deal with this issue is to truncate the value at zero as in Duffee (1999). Without truncating the value, this paper instead allows the optimizer to choose a nearby point to pass through the negative one.
Table 19: Fitted State Variables from Augmented Nelson-Siegel Model for Yields

<table>
<thead>
<tr>
<th>Z</th>
<th>$z_{1,1}$</th>
<th>$z_{1,2}$</th>
<th>$z_{2,1}$</th>
<th>$z_{2,2}$</th>
<th>$z_{3,1}$</th>
<th>$z_{3,2}$</th>
<th>$z_{3,3}$</th>
<th>$z_{3,4}$</th>
<th>$z_{3,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0492</td>
<td>0.0003</td>
<td>-0.0041</td>
<td>-0.0058</td>
<td>-0.0084</td>
<td>0.0002</td>
<td>-9.42e-05</td>
<td>-0.0227</td>
<td>-0.0015</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.0123</td>
<td>0.0008</td>
<td>0.0230</td>
<td>0.0002</td>
<td>0.0131</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.0159</td>
<td>0.0032</td>
</tr>
<tr>
<td>max</td>
<td>0.0732</td>
<td>0.0035</td>
<td>0.0316</td>
<td>-0.0055</td>
<td>0.0314</td>
<td>0.0019</td>
<td>0.0004</td>
<td>0.0124</td>
<td>0.0005</td>
</tr>
<tr>
<td>min</td>
<td>0.0159</td>
<td>-0.0012</td>
<td>-0.0469</td>
<td>-0.0067</td>
<td>-0.0344</td>
<td>-0.0004</td>
<td>-0.0005</td>
<td>-0.0559</td>
<td>-0.0149</td>
</tr>
<tr>
<td>median</td>
<td>0.0508</td>
<td>0.0003</td>
<td>-0.0046</td>
<td>-0.0058</td>
<td>-0.0088</td>
<td>0.0002</td>
<td>-5.90e-05</td>
<td>-0.0221</td>
<td>-3.82e-05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\nu_i$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.3879</td>
<td>0.6115</td>
<td>0.0511</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.0740</td>
<td>0.0337</td>
<td>0.0141</td>
</tr>
<tr>
<td>max</td>
<td>0.7249</td>
<td>0.8617</td>
<td>0.1073</td>
</tr>
<tr>
<td>min</td>
<td>0.0173</td>
<td>0.5515</td>
<td>-0.0430</td>
</tr>
<tr>
<td>median</td>
<td>0.3865</td>
<td>0.6066</td>
<td>0.0511</td>
</tr>
</tbody>
</table>

$v_i$ should be positive. But due to the numerical process, there are several negative values generated. One way to deal with this issue is to truncate the value at zero as in Duffee (1999). Without truncating the value, this paper instead allows the optimizer to choose a nearby point to pass through the negative one.
Table 20: Fitted State Variables from Augmented Nelson-Siegel Model for LIBOR/Swap Rates

<table>
<thead>
<tr>
<th>Z</th>
<th>$z_{1,1}$</th>
<th>$z_{1,2}$</th>
<th>$z_{2,1}$</th>
<th>$z_{2,2}$</th>
<th>$z_{3,1}$</th>
<th>$z_{3,2}$</th>
<th>$z_{3,3}$</th>
<th>$z_{3,4}$</th>
<th>$z_{3,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0444</td>
<td>0.0010</td>
<td>-0.0026</td>
<td>-0.0003</td>
<td>-0.0100</td>
<td>-1.03e-05</td>
<td>-0.0002</td>
<td>-0.0200</td>
<td>-0.0009</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.0093</td>
<td>0.0013</td>
<td>0.0183</td>
<td>0.0005</td>
<td>0.0205</td>
<td>0.0010</td>
<td>0.006</td>
<td>0.0208</td>
<td>0.0028</td>
</tr>
<tr>
<td>max</td>
<td>0.0667</td>
<td>0.0054</td>
<td>0.0437</td>
<td>-0.0001</td>
<td>0.0538</td>
<td>0.0019</td>
<td>0.0009</td>
<td>0.0157</td>
<td>0.0014</td>
</tr>
<tr>
<td>min</td>
<td>0.0217</td>
<td>-0.0013</td>
<td>-0.0384</td>
<td>-0.0032</td>
<td>-0.0463</td>
<td>-0.0017</td>
<td>-0.0023</td>
<td>-0.0563</td>
<td>-0.0135</td>
</tr>
<tr>
<td>median</td>
<td>0.0460</td>
<td>0.0010</td>
<td>0.0037</td>
<td>-0.0001</td>
<td>-0.0116</td>
<td>-5.89e-06</td>
<td>-0.0002</td>
<td>-0.0181</td>
<td>0.00020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$V_1$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.2553</td>
<td>0.9976</td>
<td>0.2431</td>
</tr>
<tr>
<td>std.dev.</td>
<td>0.0652</td>
<td>0.1105</td>
<td>0.0608</td>
</tr>
<tr>
<td>max</td>
<td>0.4600</td>
<td>1.3664</td>
<td>0.6629</td>
</tr>
<tr>
<td>min</td>
<td>0.0469</td>
<td>0.5842</td>
<td>0.0106</td>
</tr>
<tr>
<td>median</td>
<td>0.2549</td>
<td>0.9881</td>
<td>0.2484</td>
</tr>
</tbody>
</table>

$v_i$ should be positive. But due to the numerical process, there are several negative values generated. One way to deal with this issue is to truncate the value at zero as in Duffee (1999). Without truncating the value, this paper instead allows the optimizer to choose a nearby point to pass the negative one.
Figure 1: Parameter Loading in Nelson-Siegel Model

\begin{figure}
\centering
\begin{tikzpicture}
\begin{axis}[
    xlabel={Time to Maturity},
    ylabel={Parameter Loading},
    xmin=0, xmax=15,
    ymin=0, ymax=1,
    legend style={at={(0.5,0.65)},anchor=west},
]
\addplot [red, domain=0:15] {1};
\addplot [blue, domain=0:15] {exp(-0.617*x)};\addlegendentry{level=1}
\addplot [black, domain=0:15] {x*exp(-0.617*x)};\addlegendentry{slope=exp(-0.617*x)}
\addplot [black, domain=0:15] {x^2*exp(-0.617*x)};\addlegendentry{curvature=x*exp(-0.617*x)}
\end{axis}
\end{tikzpicture}
\end{figure}
Figure 2: Constant Maturity Treasury (CMT) Yields
Figure 3: Zero-coupon Bond Yields Converted from CMT Yields
Figure 4: Fitted Yields on Zero-coupon Bond with Three-factor General Volatility Model
Figure 5: Fitted Errors of Zero-coupon Bond Yields with Three-factor General Volatility Model
Figure 6: Forecast of Zero-coupon Bond Yields with Three-factor General Volatility Model
Figure 7: Forecast Errors of Zero-coupon Bond Yields with Three-factor General Volatility Model
Figure 8: Root-Mean-Square Fitted Errors of Zero-coupon Bond Yields with Three-factor General Volatility Model

Figure 9: Root-Mean-Square Forecast Errors of Zero-coupon Bond Yields with Three-factor General Volatility Model
Figure 10: Yield and Volatility Factors From Three-factor General Volatility Model

Figure 11: Yield and Volatility Factors From Three-factor General Volatility Model

Figure 12: Yield and Volatility Factors From Three-factor General Volatility Model
Figure 13: Yield Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 14: Yield Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 15: Yield Factor and Its Shadow Mean From Three-factor General Volatility Model
Figure 16: Level Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 17: Slope Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 18: Curvature Factor and Its Shadow Mean From Three-factor General Volatility Model
Figure 19: Information Utilization for 3.6-month, 1-year Yield

Figure 20: Information Utilization for 2,3,5-year Yield
Figure 21: Information Utilization for 7, 10, 20-year Yield
Figure 22: Fitted Yields on Zero-coupon Bond with Augmented Nelson-Siegel Model
Figure 23: Fitted Errors of Zero-coupon Bond Yields with Augmented Nelson-Siegel Model
Figure 24: One-period Ahead Forecast of Zero-coupon Bond Yields with Augmented Nelson-Siegel Model
Figure 25: Forecast Errors of Zero-coupon Bond Yields with Augmented Nelson-Siegel Model
Figure 26: Root-Mean-Square Fitted Errors of Zero-coupon Bond Yields with Augmented Nelson-Siegel Model

Figure 27: Root-Mean-Square Forecast Errors of Zero-coupon Bond Yields with Augmented Nelson-Siegel Model
Figure 28: Yield and Volatility Factors From Augmented Nelson-Siegel Model

Figure 29: Yield and Volatility Factors From Augmented Nelson-Siegel Model

Figure 30: Yield and Volatility Factors From Augmented Nelson-Siegel Model
Figure 31: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model

Figure 32: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model

Figure 33: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model
Figure 34: Information Utilization for 3,6-month, 1-year Yield from Augmented Nelson-Siegel Model

Figure 35: Information Utilization for 2,3,5-year Yield from Augmented Nelson-Siegel Model
Figure 36: Information Utilization for 7,10,20-year Yield from Augmented Nelson-Siegel Model
Figure 37: LIBOR and Swap Rates
Figure 38: Fitted LIBOR and Swap Rates with Three-factor General Volatility Model
Figure 39: Fitted Errors of LIBOR/Swap Rates with Three-factor General Volatility Model
Figure 40: One-period Ahead Forecast of LIBOR/Swap Rates with Three-factor General Volatility Model
Figure 41: Forecast Errors of LIBOR and Swap Rates with Three-factor General Volatility Model
Figure 42: Root-Mean-Square Fitted Errors of LIBOR/ Swap with Three-factor General Volatility Model

Figure 43: Root-Mean-Square Forecast Errors of LIBOR/ Swap with Three-factor General Volatility Model
Figure 44: LIBOR/Swap Rates Factor and Volatility Factor From Three-factor General Volatility Model

Figure 45: LIBOR/Swap Rates Factor and Volatility Factor From Three-factor General Volatility Model

Figure 46: LIBOR/Swap Rates Factor and Volatility Factor From Three-factor General Volatility Model

143
Figure 47: LIBOR/Swap Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 48: LIBOR/Swap Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 49: LIBOR/Swap Factor and Its Shadow Mean From Three-factor General Volatility Model
Figure 50: Level Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 51: Slope Factor and Its Shadow Mean From Three-factor General Volatility Model

Figure 52: Curvature Factor and Its Shadow Mean From Three-factor General Volatility Model
Figure 53: Information Utilization for 3,6,9-month and LIBOR

Figure 54: Information Utilization for 1,2,3-year and LIBOR
Figure 55: Information Utilization for 4, 5, 7-year and LIBOR

Figure 56: Information Utilization for 10, 15-year and LIBOR
Figure 57: Fitted LIBOR and Swap Rates with Augmented Nelson-Siegel Model
Figure 58: Fitted Errors of LIBOR/Swap Rates with Augmented Nelson-Siegel Model
Figure 59: One-period Ahead Forecast of LIBOR/Swap Rates with Augmented Nelson-Siegel Model
Figure 60: Forecast Errors of LIBOR and Swap Rates with Augmented Nelson-Siegel Model
Figure 61: Root-Mean-Square Fitted Errors of LIBOR/Swap with Augmented Nelson-Siegel Model

Figure 62: Root-Mean-Square Forecast Errors of LIBOR/Swap with Augmented Nelson-Siegel Model
Figure 63: LIBOR/Swap Rates Factor and Volatility Factor From Augmented Nelson-Siegel Model

Figure 64: LIBOR/Swap Rates Factor and Volatility Factor From Augmented Nelson-Siegel Model

Figure 65: LIBOR/Swap Rates Factor and Volatility Factor From Augmented Nelson-Siegel Model
Figure 66: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model

Figure 67: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model

Figure 68: Yield Factor and Its Shadow Mean From Augmented Nelson-Siegel Model
Figure 71: Information Utilization for 4, 5, 7-year LIBOR

Figure 72: Information Utilization for 10, 15-year LIBOR