An Assessment of Econometric Methods used in the Estimation of Affine Term Structure Models

by

Januj Juneja

A Dissertation Submitted to the Faculty of the

Committee on Business Administration

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

With a Major in Management

In the Graduate College

The University of Arizona

2 0 1 0

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Januj Juneja
entitled An Assessment of Econometric Methods used in the Estimation of Affine Term Structure Models
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

____________________ Date: 06/04/2010
Christopher Lamoureux
____________________ Date: 06/04/2010
George Jiang
____________________ Date: 06/04/2010
Edward Dyl
____________________ Date: 06/04/2010
Joseph Watkins
____________________ Date: 06/04/2010
Ronald Oaxaca

Final Approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____________________ Date: 06/04/2010
Dissertation Director: Christopher Lamoureux

Statement by Author

Acknowledgments

I am deeply indebted to my dissertation committee members Chris Lamoureux, George Jiang, Edward Dyl, Joseph Watkins, and Ronald Oaxaca for invaluable guidance, support, inspiration, encouragement, expert tutelage, patience, understanding, and extremely valuable comments and suggestions which very strongly contributed to my growth and development both as a researcher and an upstanding citizen of the community. I have also greatly benefited from excellent comments, suggestions, and guidance from Amar Dev Amar, Ivalina Kalcheva, Harpriye Amar, Jordan B. Neyland, Zafer Yuksel, Amilcar Armando Menichini, John L. Campbell, Logan B. Steele, and seminar participants at the University of Arizona and the 2009 FMA Doctoral student seminar. This dissertation evolved out of several great conversations with all of the above mentioned people.

Finally, I would like to thank my parents for their moral support, guidance, patience, uncompromising love, and emotional support throughout my life. Without them, this would not have been possible.

Preface

From June 2007 until June 2009, the notional amount outstanding of interest rate contracts rose from 347.3 billion to 437.2 billion.[1] In terms of gross market values, this is a three-fold growth in the usage of interest rate derivatives. In a recent survey put together by the International Swap Dealers Association (ISDA), 99% of banks report using interest rate derivatives in their everyday practice.[2] Banks' usage of interest rate derivatives has exploded in the recent years; it is clear that they have significant interest rate risk exposures. Furthermore, in the decades prior to the global financial crisis, a plethora of complex interest rate sensitive products was developed. Large exposures and more complex interest rate sensitive products imply increased sensitivity of banks' portfolios to changes in interest rates. The convoluted nature of these positions also makes it tougher to measure risk or uncertainty or the impact on the global financial system. Thus, a better model of interest rates would greatly enhance our understanding of the nature of the uncertainty inherent in these phenomena.

This dissertation is focused on helping both practitioners and researchers model the term structure of interest rates with the overarching goal of arriving at implications for risk management; e.g., assessing and managing exposure to interest rate risk or exchange rate risk, strategic portfolio allocation, and the formation of optimal hedging strategies for global fixed income positions.

It is achieved by three themes, each covered in a paper in this dissertation. Employing

[1] Bank of International Settlements Semiannual OTC derivatives statistics at end-June 2009

[2] 2009 International Swap Dealers Association Derivatives Usage Survey

recent advances in the term structure estimation literature, the first paper empirically evaluates how researchers estimate the class of affine term structure models. This evaluation evolved out of my desire to fill a gap in the literature related to researchers' finding that, while grounded in a strong theoretical foundation, the class of affine term structure models empirically does not perform well. This evaluation consists of two dimensions, both motivated to seek to improve the empirical performance of affine term structure models. The second paper gives a thorough examination to the implications of using principal components analysis in the estimation of affine term structure models. The third examines the extent to which the prevalence of estimation risk in using numerical integration to solve Ricatti ordinary differential equations to extract state variables creates bias, inefficiencies, and inaccurate results in the widely used class of affine term structure models.

Table of Contents

Table of Contents—*Continued*

List of Tables

List of Tables—*Continued*

Abstract

The first essay empirically evaluates recently developed techniques that have been proposed to improve the estimation of affine term structure models. The evaluation presented here is performed on two dimensions. On the first dimension, I find that invariant transformations and rotations can be used to reduce the number of free parameters needed to estimate the model and subsequently, improve the empirical performance of affine term structure models. The second dimension of this evaluation surrounds the comparison between estimating an affine term structure model using the model-free method and the inversion method. Using daily LIBOR rate and swap rate quotes from June 1996 to July 2008 to extract a panel of 3,034 time-series observations and 14 cross sections, this paper shows that, a term structure model that is estimated using the model-free method does not perform significantly better in fitting yields, at any horizon, than the more traditional methods available in the literature.

The second essay attempts explores implications of using principal components analysis in the estimation of affine term structure models. Early work employing principal component analysis focused on portfolio formation and trading strategies. Recent work, however, has moved the usage of principal components analysis into more formal applications such as the direct involvement of principal component based factors within an affine term structure model. It is this usage of principal components analysis in formal model settings that warrants a study of potential econometric implications of its application to term structure modeling. Serial correlation in interest rate data, for example, has been documented by several authors. The majority of the

literature has focused on strong persistence in state variables as giving rise to this phenomena. In this paper, I take yields as given, and hence document the effects of whitening on the model-implied state-dependent factors, subsequently estimated by the principal component based model-free method. These results imply that the process of pre-whitening the data does play a critical role in model estimation. Results are robust to Monte Carlo Simulations. Empirical results are obtained from using daily LIBOR rate and swap rate quotes from June 1996 to July 2008 to extract a panel of zero-coupon yields consisting of 3,034 time-series observations and 14 cross sections.

The third essay examines the extent to which the prevalence of estimation risk in numerical integration creates bias, inefficiencies, and inaccurate results in the widely used class of affine term structure models. In its most general form, this class of models relies on the solution to a system of non-linear Ricatti equations to back out the state-factor coefficients. Only in certain cases does this class of models admit explicit, and thus analytically tractable, solutions for the state factor coefficients. Generally, and for more economically plausible scenarios, explicit closed form solutions do not exist and the application of Runge-Kutta methods must be employed to obtain numerical estimates of the coefficients for the state variables. Using a panel of 3,034 yields and 14 cross-sections, this paper examines what perils, if any, exist in this tradeoff of analytical tractability against economic flexibility. Robustness checks via Monte Carlo Simulations are provided. In specific, while the usage of analytical methods needs less computational time, numerical methods can be used to estimate a broader set of economic scenerios. Regardless of the data generating process, the generalized Gaussian process seems to dominate the Vasicek model in terms of bias and efficiency. However, when the data are generated from a Vasicek model, the

Vasicek model performs better than the generalized Gaussian process for fitting the yield curve. These results impart new and important information about the tradeoff that exists between using analytical methods and numerical methods for estimate affine term structure models.

Chapter 1

An Empirical Evaluation of the Estimation of Affine Term Structure Models: How invariant affine transformations lead to improved empirical performance

1.1 Introduction

Employing methods that can improve the estimation precision of interest rate models can lead researchers to a higher level of accuracy in the pricing of interest rate derivatives, increased efficiency with regards to interest rate risk management, and a deeper understanding of potential sources of economic crises. The purpose of this paper is to empirically evaluate how we estimate affine term structure models. The evaluation is performed on two dimensions, with the overall goal of identifying how researchers can take advantage of recently developed techniques to improve the empirical performance of affine term structure models.

This evaluation of the class of affine term structure models of Duffie and Kan (1996) is necessary because, while grounded in a strong theoretical foundation, researchers find that the class of affine term structure models does not perform well (see for example, Ang and Piazzesi (2003), Christensen, Diebold, and Rudebusch (2008), Collin-Dufresne, Goldstein, and Jones (2008), Kim and Orphanides (2005), Duffee (2002)). Additionally, most scholars note general difficulties in optimizing the log-likelihood function, in specific; large parameter spaces, flat log-likelihood function surfaces, a high number of saddle points, etc. Despite its inability to fit the yield curve and the noted problems with estimation, the class of affine term structure models of Duffie and Kan (1996) is widely studied for several reasons; most notably, it has a good theoretical foundation, is empirically tractable, and has the flexibility to accommodate different specifications for risk premia and the volatility of the state variable. Recently, several researchers have developed various techniques designed to improve the empirical performance of affine term structure models (see for example, Dai and Singleton (2000), Collin-Dufresne, Goldstein, and Jones (2008), Joslin, Singleton, and Zhu (2010), Singleton (2006), Christensen, Diebold, and Rudebusch (2008), Christensen, Diebold, and Rudebusch (2009)). The overall goal of this paper is to empirically evaluate these techniques to see the extent to which they lead to an improvement in the empirical performance of affine term structure models, with the primary objective being to examine the effect of invariant affine transformations on the empirical performance of affine term structure models. I focus on the Gaussian subclass of affine term structure models, because it is widely regarded as a benchmark for the study of dynamic term structure models (see for example, Vasicek (1977), Langetieg(1980), Jegadeesh and Pennacchi(1996), Christensen, Diebold, and Rudebusch (2008,2009), and Joslin, Singleton, and Zhu (2010a)).

Dai and Singleton (2000) (henceforth called DS (2000)) placed certain Gaussian dynamic term structure models within an encompassing taxonomy which characterizes the econometric specifications of various affine term structure models. As part of this classification scheme, the authors provided several facts about the admissibility and identification of these models. As will be demonstrated later in this paper, these facts proved to serve as key insights into the improvement of the empirical performance of affine term structure models.

Traditionally, researchers who work with affine term structure models work with the physical (P-) measure and the risk-neutral (Q-) measure. If only cross-sectional information on prices is used then one only has enough information to work in the Q-measure, as both time-series and cross-section data are necessary to allow for the identification of the market prices of risk (Dejong, 2000). From a theoretical standpoint, DS (2000) show that it is not possible to identify all the parameters in the 3 factor Gaussian subclass of affine term structure models and so they impose identifying restrictions in the P-measure. It is not possible to identify all the parameters in the model due to the fact that state variables may rotate without changing the distribution for interest rates (see Christensen, Diebold, and Rudebusch (2008) for more details). Christensen, Diebold, and Rudebusch (2008) (henceforth called CDR (2008)) use insights from DS (2000) to obtain a representation of Gaussian affine term structure models that is equivalent to that of Dai and Singleton (2000), but imposes identifying restrictions in the Q-measure rather than the P-measure. Within the setting of this equivalent representation for the dynamics of state variables, CDR (2008), which is generalized in Christensen, Diebold, and Rudebusch (2009), are able to impose further restrictions to make the resulting affine model consistent with the Nelson-Siegel (1987) class of models as an attempt to improve the empirical tractability of the model. This representation combines the theoretically desirable properties of affine term structure models with the empirical advantages of modeling within the Nelson-Siegel (1987) framework.

Insights of CDR(2008) applied to the $A_0(3)$ model of DS (2000) with the identifying restrictions imposed in the Q-measure lead to a representation that has 7 parameters in the Q-measure (see Appendix C for more details). This research, by virtue of incorporating the above insights, uses rotations and invariant transformations to completely shift the estimation of the parameters associated with the drift term of the dynamics for the state variable into the diffusion process. Orthogonal transformations reduce the number of free parameters needed to estimate the model to 6, and this leads to an improvement in the empirical performance of the model. This shift in estimation is consistent with the notion that one cannot learn much from the drift, but in small sample one can learn a lot from the covariance matrix of the Brownian motion process (see for example, Merton (1980) or Sam and Jiang (2008)).

The application of rotations and invariant transformations constitutes some of the key insights that led to the algebraic transformation of the model of CDR (2008) into a model that retains the empirical advantages of modeling within the Nelson-Siegel framework, while yielding state variables that are consistent with the recently developed model-free method of Christensen, Goldstein, and Jones (2008) (henceforth called CDGJ (2008)).

The resulting model falls within the $A_0(3)$ sub-class of models and preserves its "canonical" nature, in the language of Singleton (2006). While the resulting model is affine in nature, it is less general than the model of CDGJ (2008), which nests the model developed by this research (see Appendix A and/or Appendix C for more details). This transformation is permissible because the canonical affine term structure representations of DS (2000) consist of latent factors (see for example, Singleton(2006), DS (2000), or CDGJ (2008)). DS (2000) remark that their representations are equivalence classes and hence, not unique to invariant transformations which implies that it is possible to convert one $A_0(3)$ model to another and yet remain within that class. However, this also implies that two models that may seem superficially different could in fact be the same model (See for example, Joslin, Singleton, and Zhu (2010a)or Babbs and Nowman (1999)). Kim and Orphanides (2005) remark further that this phenomenon leads to the possibility of multiple optimizers that can all lead to very different implications for economic behavior. Moreover, CDGJ(2008), CDR(2008), Singleton (2006) DS (2000), and Joslin, Singleton, and Zhu (2010a) suggest that there are benefits to carrying out these kinds of transformations. Furthermore, this transformation yields a model that is appropriate for comparing alternative econometric methods used to estimate affine term structure models.

The above transformed model enables me to explore another avenue for empirically evaluating the way that we estimate affine term structure models. Along this dimension, I compare the model-free method of CDGJ(2008) with the widely used inversion method of Duffie and Singleton (1997) to see which one demonstrates better empirical performance, as measured through the root mean squared error for actual yields, for the same benchmark term structure model. The model-free method of CDGJ(2008) represents an alternative way for estimating state variables, and subsequently for estimating yields. Because the method imposes "minimal" parametric structure without being dependent upon any particular model, they call this a model-free method. The authors claim that this parametric structure leads to advantages in estimation and interpretation of the yield curve, especially when examining short maturities (CDGJ, 2008).

CDGJ(2008) state that their framework is superior to more traditional frameworks used to measure state variables due to its direct tie to the short-end of the yield curve, which makes it observable.[1] CDGJ(2008) point out, for example, that in their setting,

[1]Joslin, Singleton, and Zhu (2010a) describe certain relevant practical drawbacks that are associated with the CDGJ(2008) normalization, including assuming that the short term interest rate is observable.

the three state variables r,μ_1, and μ_2, by definition, are taken to be the level state variable (which is akin to the level factor), its drift (which is akin to a slope factor), and the drift of the drift (which is akin to a curvature factor), respectively, under the Q-measure. As CDGJ(2008) argue, since the level of the short rate is tied directly to the yield curve, its drift and the subsequent drift of the drift are also tied to the yield curve directly and should provide advantages for estimation, at least at short maturities, to a method that does not directly use known information about the yield curve at short maturities. So, the state variables within the CDGJ(2008) set up consist of r and its first $N-1$ moments. The direct tie of the state variables to the yield curve is a result of rotations of the state vector consisting of the first few terms of a Taylor series expansion of the yield curve around a maturity of zero and their quadratic covariations (CDGJ, 2008).

My empirical results show that the usage of invariant transformations and Brownian motion rotations does improve the empirical performance of affine term structure models. On average, the rotation led to an improvement of about 1 basis point at the short end of the yield curve. While this may seem small, keep in mind that the rotation only involved 1 parameter. This result is consistent with CDGJ (2008) by noting that the level state variable corresponds to the shortest maturity bond yield. They argue that this construction of the state variable dynamics translates into superior empirical performance at the short-end of the yield curve, with respect to the inversion method of Duffie and Singleton (1997). This argument corresponds to the second constituent of my evaluation.

My empirical results also show that, within a less general yet still affine setting, the inversion method, which I refer to in the traditional sense of Pearson and Sun (1994) or Duffie and Singleton (1997), as an econometric method of choice used to estimate state variables does not perform significantly worse at any horizon than the method suggested by CDGJ (2008). Robustness of the estimation results is verified through estimation by a Kalman Filter, which, as the literature notes, is the most efficient method of parameter estimation. (See, for example, Dejong (2000), Hamilton (1994), or Chen and Scott(1993)). The stark differences in empirical performance obtained in the recent literature therefore reflect the of the choice of yields and are not related to the choice of econometric methods used to estimate the term structure model. Furthermore, Joslin, Singleton, and Zhu (2010a) find that the models of both CDGJ(2008) and CDR(2008,2009) are special cases of their model and DS (2000). In fact, they show that the CDGJ(2008) set up is observationally equivalent to the class of Gaussian Dynamic Term Structure Models of DS (2000), as well. Along these lines, this paper serves as an empirical counterpart to the discussion in Joslin, Singleton, and Zhu (2010a) by demonstrating empirically that the models of CDR (2008) and CDGJ(2008) are just normalizations of DS(2000).

The rest of the paper is organized as follows: Section II revisits the standard class of affine term structure models. Section III describes the data and the extrapolation procedure used to obtain zero-coupon treasury yields from swap rate quotes. Section IV presents and interprets the results. Section V presents and interprets the yield errors. Finally, Section VI gives the conclusion.

1.2 The Affine Class of Term Structure Models

The standard affine class of term structure models of Duffie and Kan (1996) and DS (2000) postulate state variables, X that evolve according to the following Markov N-dimensional transition dynamics under the equivalent risk neutral measure

$$dX_t = \kappa^Q(\theta^Q - X_t)dt + \Sigma\sqrt{S_t}dZ_t^Q$$

where κ^Q is an NxN matrix, θ^Q is an N-dimensional vector, Σ is an NxM matrix, and dZ_t^Q is a vector of M independent brownian motions, where $(M \geq N)$. Furthermore, S is a diagonal MxM matrix that takes the following form

$$S_{ii,t} = \alpha_i + \beta_i^T X_t$$

α is an $Nx1$ vector of coefficients and β_i is the i^{th} row of β, an NxN matrix of coefficients.

The spot rate is also taken to be an affine function of X_t

$$r_t = \delta_o + \delta_1^T X_t$$

where δ_1 is an N-dimensional vector. According to this model, for a given maturity τ, zero coupon bond prices take the following exponential affine form

$$P(t,\tau) = \exp\left(a(\tau) + b(\tau)^T X_t\right)$$

where the parameters $a(\tau)$ and $b(\tau))$ solve the following system of ordinary differential equations

$$\frac{da(\tau)}{d\tau} = -\theta^{Q^T}\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\alpha_i - \delta_0$$
$$\frac{db(\tau)}{d\tau} = -\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\beta_i - \delta_1$$

These ordinary differential equations or Ricatti equations can be solved using standard Runge Kutta numerical methods with initial conditions

$$a(0) = 0$$

$$b(0) = 0$$

Yields are defined, in terms of price, to be $P(t, \tau) = \exp(-\tau Y(t, \tau))$ which means that they can be written more succinctly as

$$Y(t, \tau) = A(\tau) + B(\tau)^T X_t$$

where $A(\tau) = -\frac{a(\tau)}{\tau}$ and $B(\tau) = -\frac{(b(\tau))}{\tau}$

The class of affine term structure models contains several models as special cases. This list of models includes, but is not limited to the model of Vasicek (1977), Langetieg(1980), Cox, Ingersoll, and Ross (1985), Jegadeesh and Pennacchi(1996), CDR(2008), and CDGJ(2008) (see Dejong (2000) for a more comprehensive list). As scholars have noted, these proposed models have later been determined to be a normalized or transformed version of a model that falls within the affine class of term structure models. All models that are estimated here fall within that general class. More specifically, they fall within the $A_0(3)$ class of models that are Gaussian in nature.

1.2.1 The $A_0(3)$ model

This paper focuses on the $A_0(3)$ since many papers use it as a base to study term structure models (see for example, Langetieg(1980), CDGJ(2008), Joslin, Singleton, and Zhu (2010a)). In this spirit, I present the $A_0(3)$ model in terms of the state variables

$$d\begin{bmatrix} r_t \\ \mu^Q_{1,t} \\ \mu^Q_{2,t} \end{bmatrix} = \begin{bmatrix} \mu^Q_{1,t} \\ \mu^Q_{2,t} \\ -(\lambda^2\mu^Q_{1,t} + 2\lambda\mu^Q_{2,t}) \end{bmatrix} dt + \Sigma^R dB^Q_t$$

where Σ^R is an orthogonal matrix that is estimated numerically due to certain constraints associated with the transformation process (see Appendix A and/or Appendix B for more details). Yields are then taken to be a function of these state variables. According to CDGJ(2008), the presentation of the above system of stochastic differential equations for the state variable transition dynamics expedites the estimation and interpretation of multi-factor models. The three state variables r,μ_1, and μ_2, by definition, are taken to be the level state variable, its drift, and the drift of the drift, respectively, under the Q- measure.

1.3 Data

The data set comprises daily observations of LIBOR, with maturities of 1-month, 3-month, 6-month, 9-month, and 12-months and swap rate quotes for 2-year, 3-year, 4-year, 5-year,7-year, and 10-years, from June 21, 1996 through July 02,2008, from Bloomberg. This resulted in $3,034$ observations, after eliminating special days like Christmas or New Year's when no trading took place. Given LIBOR rates and swap rate quotes, it is possible to construct a panel of zero coupon yields via an extrapolation procedure. The details of that procedure are explained in Appendix C. Some descriptive statistics for the sample yields are shown in tables 1.1 to 1.3.

The minimum of 0.013836525 for the 10 year rate is due to the fact that the 7-year swap rate quote on that day was 0%. Excluding that minimum, the minimum of all other days in the sample was 0.029772228. For the 7-year rates, the minimum of

0.025563812 was due to the fact that several LIBOR quotes were around 1%.

1.4 An Overview of the model-free method for estimation of the state variables

Using the CDGJ (2008) method, the set of state variables is estimated using the insight that a tractable way to transform from unobservable state variables to observable state variables is to use a Taylor Series expansion around $\tau = 0$. CDGJ(2008) contend that the main advantage to their approach is that it allows for error reduction by imposing some parametric structure, while retaining flexible structure.

This approach is grounded in the result of Litterman and Schienkman (1991) that only three factors drive the variation in yield curve behavior. So, the first step in extracting the state variables is to obtain the principal component factor loadings from PCA on the 1-month, 3-month, 6-month, 12-month, 2-year, 3-year, 5-year, and 7-year yields. The remaining maturities will be used elsewhere in the analysis, and therefore were left out in order to avoid using the same data. After running PCA on the 8 yields above, and over the entire sample period, I find that the first three factors cumulatively explain about 99.8% of the variation in the 8 bond yields.

Using these three principal components, I can form or reconstruct yields with the following approximation.

$Y(t,\tau) \approx \sum_{k=1}^{3} f_k(t)P_k(t)$

$P_k(t)$ is the realization of the k^{th} principal component, while $f_k(t)$ is a loading function that will be extrapolated to a low order polynomial which will bring the curve down to zero. This gets at the idea of using the Taylor series at short maturities, since, with the formula above, I am in a sense using the derivatives corresponding to the

Taylor series to construct the approximate yield function. Although it is approximate, the above relation holds with great accuracy and I can use it to obtain the derivatives of the yield function to get at the level, slope, and curvature. To see this symbolically, note that if I take the derivative of the expression above

$$\frac{\partial^n Y(t,\tau)}{\partial \tau^n} = \sum_{k=1}^{3} \frac{\partial^n f_k(\tau)}{\partial \tau^n} P_k(t)$$

So, it can explain about 99% of the variation in the yield curve, so using it as the starting point for obtaining estimates may prove to be fruitful.

Upon obtaining the principal components, the next step in the process is to extrapolate the yield curve down to 0, which can be accomplished using a low-order polynomial. Following CDGJ (2008), I use a linear polynomial to approximate $f_1(\tau)$, a quadratic polynomial to approximate $f_2(\tau)$, and a cubic polynomial to approximate $f_3(\tau)$.

Analytic derivatives can be computed for each polynomial and then I can use those to approximate the yield function derivatives. Upon obtaining the yield function derivatives, I can back out the state variables as described in section 2. Descriptive statistics for each of the state variables are shown in table 1.4.

In Appendix G, I consider a small Monte Carlo Simulation study that checks the accuracy and robustness of this state variable estimation procedure. From this monte-carlo simulation, it is clear that the first and second state variables are estimated with reasonable accuracy, but there is some bias in the third state variable. These results are consistent with CDGJ (2008).

1.5 Model and Estimation

I estimate the model using Quasi-Maximum Likelihood (QML) Estimation. More specifically, for the inversion method, I assume three bonds are measured with error and that three bonds are measured without error. In this analysis, I take the 3 month, 2 year, and 4 year to be measured without error, while the 6 month, 3 year, and the 5 year are measured with error. Bonds with similar characteristics are used in order to minimize the informational disparity across state variables. This is the same method used in Duffie and Singleton (1997) and others. For the bonds that are measured without error, I follow Pearson and Sun (1994) and exactly transform the yields into state variables. This requires the use of a Jacobian to carry out the transformation which will be multiplied by the transition density, a multivariate normal. Hence, I am only required to know the first 2 conditional moments of the state variables, which simplifies the analysis. (Please see Appendix E for the details of those calculations.) Symbolically,

$$Y(t,\tau) = \frac{A(\tau)}{\tau} + \frac{(B(\tau))^T}{\tau} X_t$$

More compactly,

$$Y(t) = A + BX(t)$$

where

$$Y(t) = \begin{bmatrix} y_{t,t+\tau_1} \\ \vdots \\ y_{t,t+\tau_N} \end{bmatrix}$$

$$A = \begin{bmatrix} \sum_{i=1}^{F} \frac{A_i \tau_1}{\tau_1} \\ \vdots \\ \sum_{i=1}^{F} \frac{A_i \tau_N}{\tau_N} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{B_{\tau_1}}{\tau_1} \\ \vdots \\ \frac{B_{\tau_N}}{\tau_N} \end{bmatrix} = \begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} \dots \frac{B_N(\tau_1)}{\tau_1} \\ \ddots \\ \frac{B_1(\tau_N)}{\tau_N} \dots \frac{B_N(\tau_N)}{\tau_N} \end{bmatrix}$$

N=the number of bonds

F=the number of factors

and hence,

$X(t) = (B)^{-1}(Y(t) - A)$

The final form of this part of the likelihood is

$L_1 = \frac{1}{det(B)} f(X_{t+1}|X_t)$

For the three bonds which are measured with error, I cannot simply invert in order to get the state variables. Relying on distributional assumptions made about the error process, three yields are characterized by the following relation,

$\widetilde{Y}(t, \tau) = \widetilde{A} + \widetilde{B}X(t) + \varepsilon(t)$

where the tildes are used to differentiate the bonds that are assumed to be measured with error from those that are measured without error.

The conditional distribution of the errors is

$g(\varepsilon(t_i)|\varepsilon(t_{i-1})) \sim \aleph(0, \Sigma)$

where Σ is the covariance matrix for the measurement errors. Now, the log-likelihood function is the sum of the two Gaussian log-likelihood functions (without error and with error). For the CDGJ (2008) method of estimating state variables, rather than inverting yields to obtain state variables, the state variables are taken to be exactly those obtained based upon the procedure outlined in Section 2 and Section 4, while maintaining the assumption that the remaining yields are observed with Gaussian error. For both methods, the parameter space is given by:

$$\Theta = \{\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \sigma_{32}, \sigma_{33}, ME_1, ME_2, ME_3, ME_4, ME_5, ME_6\}$$

Given that canonical representation, there is no loss in generality in fixing the mean of the model under the Q-measure to be 0, while leaving it to be estimated under the actual dynamics (see for example CDR(2008) or DS(2000)). Furthermore, recent work by Joslin, Singleton, and Zhu (2009) finds that the forecasting ability associated with the P-dynamics of the CDR(2008) model is no better than an unconstrained VAR(1) model.

Moreover, while I work in the equivalent risk-neutral measure, I am working within the model framework of CDR (2008), which includes their set of assumptions associated with the Q-dynamics. In this case, for the inversion method, the eigenvalues for the κ matrix are; λ_1=0, λ_2=3.9386 and λ_3=-1. For the model-free method, the eigenvalues are λ_1=0, λ_2=2.112963823 and λ_3=2.113020511. Further, stationarity of the state variables is ensured if the eigenvalues of the κ^Q matrix are real (or if the complex part is positive, see Ahn, Dittmar, and Gallant (2002)). Stationarity is not a prerequisite for the process to be well-defined (Christensen, Diebold, & Rudebusch, 2008).

Table 1.6 and 1.7 present the results from the maximum likelihood estimation, for each method respectively. More specifically, each table contains parameter estimates and asymptotic standard errors (in parentheses) from the full sample. Please note that the elements of the measurement error matrix and their standard errors were multiplied by 10000.

Standard errors were computed using the Berndt, Hall, Hall, and Hausman (1974) approach. Please also note that previous studies (see, for example, Duffee (2002)), used a Cholesky factorization of the terms constituting the covariance matrix for the measurement error, which is symmetric. On the other hand, I worked with the covariance matrix directly, which has an impact on the scaling of the standard error estimates. The optimized log-likelihood value on $n = 3,034$ observations was 204,476.709 for the CDGJ (2008) method, while on that same number of observations the log-likelihood value for the inversion method was 197,660.505.[2] Finally, note that the parameters δ_1 and δ_2 are related to the orthogonal rotation and therefore do not affect the optimization process and hence do not have an associated standard error. Like λ, they are completely determined through the orthogonality constraints that are imposed on the Σ^R matrix.

As CDGJ(2008) note, their method did provide good starting conditions for the otherwise troublesome search for a global optimum. The CDGJ (2008) method also was characterized by a likelihood surface over which a unique global optimum could easily be found by a gradient-based search algorithm for each dimension of the parameter space. When used as initial conditions for the inversion method, the solution from the CDGJ(2008) method reduced the potentially large space over which to search for the optimum.

From table 1.8, it is clear that this transformation had a stronger affect on the short-end of the yield curve, than at the long-end of the curve. On average, the rotation led to an improvement of about one basis point. While this may seem small, keep in mind that the rotation only involved one parameter. This result confirms intuition that has been suggested in CDGJ(2008).

[2]The entire optimization procedure for both methods, for my parametrization and with a Intel Pentium 4 CPU processor, takes about 2 days using sequential quadratic programming as implemented in MATLAB's FMINCON.

1.5.1 Kalman Filter Estimation

This section presents estimation results via the Kalman Filter method. The Kalman Filter method is the most efficient way to check the robustness of the conclusions regarding the comparison between the CDGJ(2008) method and the Duffie and Singleton (1997) method.

The Kalman Filter is a system of recursive equations that specifies both the state space transition dynamics of the state variable as well as a measurement equation in which the actual filtration of the noise occurs. In the term structure literature, this noise could be associated with data-entry error or bid-ask spread of bonds, among other things.

Following CDR(2008) and others who use the Kalman Filter to estimate continuous-time affine terms structure models, I start the algorithm at the unconditional mean vector and the conditional covariance matrix. Once, I start the algorithm, I use the conditional mean vector and conditional covariance matrix to run the algorithm. These are given in appendix E. For more details of the implementation, please see Dejong (2000).

The results from the Kalman Filter estimation, using the same 6 bonds as in the maximum likelihood estimation for consistency, are shown in table 1.9. In this framework, I assume that all 6 of the yields- the 3 month, 6 month, 2 year, 3 year, 4 year, and 5 year-are observed with error.

The log-likelihood function on $n = 3034$ observations is 202,132.699.

1.6 Yield Errors

In this section, I present yield errors for each method based upon the maximum likelihood parameter estimates from the tables above. A priori, I would expect the CDGJ (2008) method to have a smaller error, relatively, at the very short-end of the yield curve due to its construction, while on the relatively longer-end I would expect the inversion method to have smaller error due to its incorporation of more information. The inversion method will be associated with smaller error at the long-end of the yield curve only if I use long term yields to back out the state variables. Consequently, I would expect my results to be different from those of CDGJ (2008) because while they assume the 3-month yield, the 2-year yield, and the 10-year yield to be priced with no error, I use 3-month, 2-year, and the 4-year yields to invert to obtain the state variables. Therefore, across both methods, there is less disparity in the content of information that I am incorporating into the model relative to CDGJ (2008). This means that I would expect that at the very short-end the CDGJ (2008) method would do better, for middle maturities I would expect the inversion method to do better, while at the long-end of the curve I would expect both to not perform that well. However, although the CDGJ (2008) method involves the direct extraction of state variables from the yield curve and employs regression based methods to obtain these variables, I believe, a priori that the CDGJ (2008) method may perform better due to the fact that my principal components analysis involved 8 yields of varying maturities out 10 years. Moreover, I also believe that had I used more information from the yield curve with the inversion method, it too would have performed better at the long-end of the yield curve.

Some notes are in order. Table 1.10 contains root mean squared errors (RMSE) computed from yields over the 1996-2008 sample for the $A_0(3)$ model. The model-implied yields are computed as $Y(t,\tau) = \frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau}X(t)$ and errors are defined as actual minus model-implied yields measured in basis points. In Panel A, the inversion method inverts the state vector $X(t)$ by inverting the 3 month, 2 year, and 4 year yields, while in Panel B the inversion method inverts the state vector $X(t)$ with the 1 month, 3 year, and 10 year yields. The goal of Panel B is to gain a more coherent understanding of the impact of changing the yields used to invert to obtain the state variable on the ability of the model to fit the yield curve. The CDGJ(2008) method uses estimates of r, μ_1, and μ_2 as actual state variables. The Kalman Filter uses a dynamic recursive state space formulation to estimate the transition dynamics, $X(t)$.

On the whole, the results generally support my hypotheses. It is clear that, for the inversion method, the errors are larger for maturities greater than or equal to 7 years. For the method proposed by CDGJ (2008), at the long end of the yield curve the error is larger. One potential explanation is that the information coming from the 2-year, 3-year, and 4-year maturities is not useful for fitting maturities greater than six years. For the CDGJ (2008) method, generally, the short-end of the yield curve is characterized by smaller error than the longer end. The graph below plots the RMSE against various horizons, and provides an alternative view of the fact that there are no discernible pattern in the model's explanatory power.

It is clear that there is no distinguishable pattern in the explanatory power of the model for both methods. Estimation via the Kalman Filter reinforces these conclusions as it is clear that the economic implications of the root mean square errors are not very different across any horizon. This implies that independent of one's computational method of choice for state variable transition dynamics estimation, there should not be any pattern in the model explanatory power that is dependent upon the method I used to estimate the state variables.

1.6.1 Impact of changing the yields used in the inversion method on the ability of the model to fit the yield curve

This subsection presents estimation results with an alternative choice of yields that are used to invert to obtain state variables. This selection seems to be an arbitrary procedure with at least one important implication. The choice of yields impacts where information comes from with regards to the yield curve. The yields that were used to invert to obtain the state variables were the 1-month, 9-month, and the 10-year. In this manner, I attempted to examine the impact of changing the yields that are assumed to be priced without error on the estimation results. Presumably, this should have a significant influence on the root mean squared error of all the yields. This claim is verified in panel B of table 9. Notice that the largest errors are at the short end and that as we move further out the errors are relatively small. The large errors at the short end could be a manifestation of tainted LIBOR quotes. This additional error may have resulted from banks understating LIBOR in an effort to reduce their costs of borrowing during the financial crisis of 2008.

1.7 Conclusion

Employing recent advances in the term structure estimation literature, this paper empirically evaluates how researchers estimate the class of affine term structure models. This evaluation evolved out of a desire to fill a gap in the literature related to researchers' finding that while grounded in a strong theoretical foundation, the class of affine term structure models empirically does not perform well. This evaluation consists of two dimensions, both motivated to seek to improve the empirical performance of affine term structure models.

This paper reveals that the application of invariant affine transformations to the class of affine term structure models does improve the empirical performance of affine term structure models. More specifically, this research employs rotations to completely shift the estimation of the parameters associated with the drift term of the dynamics for the state variable into the diffusion process. Orthogonal transformations reduce the number of free parameters needed to estimate the model from 7 to 6, and this leads to an improvement in the empirical performance of the model.

This finding that orthogonal rotations improved the empirical performance of the affine model by shifting the estimation of the drift parameters into the diffusion parameters is consistent with the findings in the literature. Merton (1980) and Sam and Jiang (2008) both speak to the lack of estimation precision in the drift parameter estimates relative to the estimation precision diffusion parameter estimates.

This paper also argued that the inversion method, which I refer to in the traditional sense of Pearson and Sun (1994) or Duffie and Singleton (1997), as an econometric method of choice used to estimate state variables does not empirically perform significantly worse, at any horizon, than the CDGJ(2008) method.

This conclusion is a simple reflection of the fact that a normalization of an affine term structure model should not perform very differently from another affine term structure model. This result is consistent with Joslin, Singleton, and Zhu (2010a) in that the model of CDGJ (2008) is observationally equivalent to DS (2000). Furthermore, the stark differences in empirical performance obtained in CDGJ (2008) seem to have been a manifestation of the choice of yields and not related to the choice of econometric methods used to estimate the term structure model.

I worked in the Q-measure. Hence, while I could not forecast out of sample, my research finds preliminary evidence that is suggestive of empirical improvements in existing methods used to estimate affine term structure models.

Both dimensions taken together, my results impart new and important information about the empirical estimation of affine term structure models.

Table 1.1. Objective Function Summary Statistics I

SumStat	*Obj*1	*Obj*2	*Obj*3	*Obj*4	*Obj*5	*Obj*6
Max	9.93382E-07	9.91E-07	9.99E-07	9.995E-07	3.834E-06	6.16E-06
Mean	5.7768E-08	4.19E-08	1.002E-07	2.59E-08	5.11E-08	9.52E-08
Min	0	0	0	0	0	0

TABLE 1.2. Spot Rate Summary Statistics I

SumStat	1*mo*	3*mo*	6*mo*	9*mo*	12*mo*	2*yr*	3*yr*
Mean	.041452	0.04165	0.04238	0.04306	0.04388	0.04615	0.04689
Stdev	.018007	0.01817	0.01811	0.01800	0.01775	0.01558	0.01502
Max	.068213	0.068688	0.071088	0.073325	0.075013	0.076023	0.0761435
Min	.0102	0.01	0.0098	0.0098	0.0099	0.01274	0.0140

Table 1.3. Spot Rate Summary Statistics II

$4yr$	$5yr$	$6yr$	$7yr$	$8yr$	$9yr$	$10yr$
0.04892	0.05063	0.05096	0.05247	0.05433	0.05479	0.05474
0.01361	0.01247	0.01272	0.01131	0.00945	0.0092	0.01025
0.0767	0.0771	0.0771	0.0775	0.0787	0.0777	0.0780
0.0174	0.0210	0.0177	0.02556	0.0222	0.0258	0.0138

Table 1.4. Summary Statistics of State Variables

Statistic	r	μ_1	μ_2
mean	0.0311	0.2427	20.1513
$st-dev$	0.0119	0.0366	0.1248
max	0.0541	0.3327	20.6052
min	0.0060	0.1261	19.8112

Table 1.5. Check for Accuracy and Unbiasedness of State Variable Estimates.

SV		r			μ_1			μ_2	
$M.E.$	$10^6 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$10^6 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$10000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2
0	.0000	1.000	1.000	.0000	1.000	1.000	.0000	1.000	1.000
0.5	-0.0001	1.000	1.000	-0.014	1.000	1.000	-0.028	1.001	0.999
2.0	.00575	0.999	0.999	-0.120	0.999	0.999	0.17	0.997	0.991

TABLE 1.6. Maximum Likelihood Estimation of $A_0(3)$ system using the Inversion Method to estimate state variables.

Parameter	Max. Likelihood Estimate (Std. Error)
δ_1	-0.1261
δ_2	0.1472
λ	1.969
σ_{11}	$2.2226e^{-5}$ $(1.751e^{-9})$
σ_{21}	$4.00e^{-3}$ $(2.409e^{-7})$
σ_{22}	$8.06e^{-4}$ $(2.216e^{-7})$
σ_{31}	$2.95e^{-3}$ $(2.156e^{-6})$
σ_{32}	$1.1e^{-4}$ $(2.284e^{-6})$
σ_{33}	$9.05e^{-4}$ $(2.186e^{-6})$
Σ_{11}	0.005091 (0.0107986)
Σ_{21}	-0.0035 (0.417081)
Σ_{22}	0.01010 (0.29754)
Σ_{31}	0.02280 (0.544253)
Σ_{32}	-0.0294 (0.18495)
Σ_{33}	0.1266 (2.47298)

Table 1.7. Maximum Likelihood Estimation of $A_0(3)$ system using the CDGJ(2008) Method to estimate state variables.

Parameter	Max. Likelihood Estimate (Std. Error)
δ_1	-0.0908
δ_2	0.1282
λ	2.113
σ_{11}	$1.712e^{-5}$ $(4.670e^{-9})$
σ_{21}	$4.044e^{-3}$ $(1.641e^{-9})$
σ_{22}	$9.84e^{-4}$ $(2.294e^{-8})$
σ_{31}	$2.95e^{-3}$ $(4.4215e^{-9})$
σ_{32}	$1.1e^{-4}$ $(2.783e^{-6})$
σ_{33}	$9.05e^{-4}$ $(1.596e^{-6})$
$\Sigma_{me_{11}}$	0.0051 (0.044415)
$\Sigma_{me_{21}}$	-0.003 (0.027970)
$\Sigma_{me_{22}}$	0.0101 (0.4994276)
$\Sigma_{me_{31}}$	0.0228 (0.1447076)
$\Sigma_{me_{32}}$	-0.029 (0.631626)
$\Sigma_{me_{33}}$	0.1266 (1.3336224)

TABLE 1.8. Assessment of Invariant Transformation on the Empirical Performance of the $A_0(3)$ model

Yield	$RMSE_{before}$	$RMSE_{After}$
$1-month$	10.009	3.286
$3-month$	10.332	5.749
$6-month$	10.740	20.665
$9-month$	11.086	12.389
$12-month$	11.412	7.966
$2year$	11.804	11.168
$3year$	11.905	11.941
$4year$	12.220	14.841
$5year$	12.507	13.669
$6year$	12.611	13.815
$7year$	12.739	14.080
$8year$	12.653	13.344
$9year$	12.778	13.955
$10year$	13.122	14.953

TABLE 1.9. Maximum Likelihood Estimation of $A_0(3)$ system using a Kalman Filter State Space Formulation.

Parameter	Max. Likelihood Estimate (Std. Error)
δ_1	0.2220
δ_2	0.0424
λ	3.718
σ_{11}	$1.232e^{-5}$ ($8.084e^{-7}$)
σ_{21}	$2.029e^{-3}$ ($1.078e^{-7}$)
σ_{22}	$7.989e^{-4}$ ($2.685e^{-7}$)
σ_{31}	$3.599e^{-3}$ ($6.053e^{-8}$)
σ_{32}	$1.105e^{-4}$ ($6.183e^{-7}$)
σ_{33}	$8.039e^{-4}$ ($5.732e^{-7}$)
$\Sigma_{me_{11}}$	0.00509 (0.000001)
$\Sigma_{me_{21}}$	−0.0035 (0.002235)
$\Sigma_{me_{22}}$	0.01010 (0.00092)
$\Sigma_{me_{31}}$	0.02280 (0.00338)
$\Sigma_{me_{32}}$	−0.0294 (0.00001)
$\Sigma_{me_{33}}$	0.1266 (0.00200)
$\Sigma_{me_{41}}$	0.01593 (0.000003)
$\Sigma_{me_{42}}$	$4.8318e^{-7}$ (0.000001)
$\Sigma_{me_{43}}$	0.03513 (0.000001)
$\Sigma_{me_{44}}$	0.01695 (0.000001)
$\Sigma_{me_{45}}$	−0.0079 (0.0000027)
$\Sigma_{me_{51}}$	0.09995 (0.000038)
$\Sigma_{me_{52}}$	0.04486 (0.000001)
$\Sigma_{me_{53}}$	−0.0450 (0.0000030)
$\Sigma_{me_{55}}$	0.03022 (0.0000008)
$\Sigma_{me_{56}}$	−0.4043 (0.0000007)
$\Sigma_{me_{61}}$	0.01693 (0.000003)
$\Sigma_{me_{62}}$	−0.1295 (0.000036)
$\Sigma_{me_{63}}$	0.09999 (0.000023)
$\Sigma_{me_{64}}$	−0.0905 (0.0000011)
$\Sigma_{me_{66}}$	0.49618 (0.000002)

Table 1.10. $A_0(3)$ Yield Errors

	Panel A		
Yield	Inversion Method	CDGJ(2008) Method	Kalman Filter
$1-month$	3.287	3.286	3.286
$3-month$	-	1.479	5.749
$6-month$	5.181	15.079	20.665
$9-month$	1.762	5.356	12.389
$12-month$	5.901	4.202	7.966
$2year$	-	19.002	11.168
$3year$	3.760	2.083	11.941
$4year$	-	1.771	14.841
$5year$	0.730	3.163	13.669
$6year$	3.700	4.782	13.815
$7year$	8.284	6.938	14.080
$8year$	9.583	8.841	13.344
$9year$	12.343	10.115	13.955
$10year$	15.186	11.131	14.953
	Panel B		
Yield	Inversion Method	CDGJ(2008) Method	Kalman Filter
$1-month$	-	3.286	3.286
$3-month$	19.249	1.479	5.749
$6-month$	7.979	15.079	20.665
$9-month$	-	5.356	12.389
$12-month$	0.319	4.202	7.966
$2year$	3.760	19.002	11.168
$3year$	0.139	2.083	11.941
$4year$	3.574	1.771	14.841
$5year$	1.174	3.163	13.669
$6year$	1.341	4.782	13.815
$7year$	0.960	6.938	14.080
$8year$	0.134	8.841	13.344
$9year$	0.091	10.115	13.955
$10year$	-	11.131	14.953

Chapter 2

Simulation-based inference for Principal Components Analysis with implications for term structure estimation

2.1 Introduction

The popularity of principal components analysis for analyzing the yield curve has grown quite dramatically since Litterman and Shienkman (1991). This popularity is in part due to the fact that principal components analysis allows researchers to summarize the systematic risk of many tradeable financial assets by a smaller number of sources or constructed factors (Diebold, Piazzesi, and Rudebusch, 2005). Given this, it is not surprising that principal components analysis is also popular as a tool used for hedging risk exposure in various settings (see for example, Litterman and Shienkman (1991) or Jamshidian and Zhu (1996)).

Dai and Singleton (2000) note that there are a few observable factors that drive term structure behavior. Perignon, Smith, and Villa (2007) and Piazzesi (2003) further remark that much of the intuition for yield curve dynamics can be provided by PCA. While there have been numerous attempts to incorporate principal components analysis into term structure applications, to the best of my knowledge, no one has yet examined the types of errors or additional complexities that get introduced through the application of principal components analysis. This is the goal of this paper.

Certain applications of principal components analysis do not prompt extra scrutiny. In particular, many applications involve an assessment of risk exposure (see for example, Litterman and Schienkman (1991), Jamshidian and Zhu (1996), Perignon and Villa (2006), Perignon, Smith and Villa (2007), Scherer and Avelleneda (2002), Driessen, Melenberg, and Nijman (2003), or Collin-Dufresne, Goldstein, and Martin (2001)). More recently, however, authors have attempted to bring principal components analysis into a more formal model setting (see for example, Ang and Piazzesi (2003), Collin-Dufresne, Goldstein, and Jones (2008) (CDGJ(2008)), and Joslin, Singleton, and Zhu (2010)). It is this acceleration of bringing to bear principal components analysis in formal model settings that warrants a study of the potential econometric implications of its application to term structure modeling.

The first objective of this paper is to document and assess the extent to which yield curve persistence affects state variables obtained through the usage of principal components analysis. This approach takes the serial correlation in yields as given, and attempts to examine the extent to which certain econometric methods such as pre-whitening can improve model estimation.

Economic data, such as yields, are highly autocorrelated (see for example, Piazzesi(2003) or Gorodnichenko and Ng (2010)). While the literature has focused on short-term horizons, changes in the serial correlation of yields, also known as persistence, can have both cross-sectional and time-varying effects upon the yield curve (Watson, 1999). Piazzesi (2003) documents strange changes in persistence in various short term interest rates. Watson (1999) goes on to note that changes in persistence of the overnight Fed Funds rate have implications for Fed Policy. Persistence in yields can also create practical problems in the estimation of certain economic models (Piazzesi, 2003).

The literature has traditionally documented persistence in yields as arising from persistence in factors. Affine term structure models specify yields as an affine function of factors. This means that within the affine term structure model setting, persistence in yields comes from persistence in factors. Using monthly data from January 1964 to December 2001, Piazzesi (2003) documents that the level factor demonstrates the most persistence, followed by the slope factor, and then finally the curvature factor. Several authors including Chen and Scott (1993), Pearson and Sun (1994) and Campbell and Viceira (1997) document strong persistence in the state variables (see Duarte (2004) for a more complete list). Accounting for persistence in factors draws upon knowledge of various aspects of the term structure model. Duarte (2004) finds that the strong persistence in factors may have been caused, at least in part, by restrictions to the parametrization of the market price of risk within the affine term structure model setting.

On this dimension, I find that whitening does play an important role in the estimation of state variables that obtain from data-driven methods, such as principal components analysis. Moreover, recent attempts to incorporate principal components analysis into a more formal model setting, such as that of CDGJ (2008), do in fact stand to benefit from whitening the data before computation of the state variables. Evidence from monte-carlo simulations leads me to believe that the estimation of the first state variable stands to gain the least. As CDGJ(2008) noted, it is the least negatively affected by serial correlation. However, the second and third state variables are more severely impacted by serial correlation. The process of pre-whitening the data does matter for model estimation.

The second contribution of this paper is to isolate the effect of persistence on principal component based econometric methods used to extract state variables en route to estimating affine term structure models. The econometric estimation procedures used to estimate affine term structure models specify observed yields as an input, and this could introduce additional sources of error. Furthermore, as I describe below, the application of principal components analysis could introduce additional complexities that may adversely affect the estimation procedure. These issues are also explored from a theoretical perspective through a demonstration that the log-likelihood functions under the two methods are observationally equivalent, in the presence of serial-correlation and if the data are Gaussian. This demonstration will serve to extend work done by Joslin, Singleton, and Zhu (2010) to the log-likelihood function in their setting with that of CDGJ (2008).

Principal components analysis is a cross-sectional tool used to obtain loadings at any point in time. Repeated applications of principal components analysis to various subsets of data can provide implications for the stability of the factors across time. Nevertheless, principal components analysis is a data-driven tool. So any problems with the data can lead to bias in the method's application, as principal components analysis inherits problems in the data. Still, the application of principal components analysis imposes certain restrictions on the factor loadings matrix and the true factors, so that if the data do comply with these restrictions then the principal components estimator is estimating the true factors and factor loadings matrix in the absence of rotation (Bai and Ng, 2010). Estimates from principal components analysis are rotations of the original factors and loadings. Bai and Ng (2010) study the situation in which the true factors can be obtained asymptotically and without the need for rotation. Using insight from Bai (2003) and Bai and Ng (2006), they derive the limiting distribution for these factors. Essentially, factors equivalent to those obtained from principal components analysis can be arrived at from a rotation and scaling exercise, since that is a mechanistic description of principal components analysis.

As mentioned above, the application of principal components analysis imposes restrictions, perhaps unwillingly, onto the data. Yet, researchers seem to apply principal components analysis to a variety of economic and financial problems without considering whether or not its introduction creates additional concerns for the econometric estimation procedure (see for example, CDGJ(2008), Joslin, Singleton, and Zhu (2010), Litterman and Schienkman (1991), Collin-Dufresne, Goldstein, and Martin (2001)).

On this dimension, I find that usage of principal component based factors does indeed impose restrictions that create deviations from an empirical density that relies on linearity and a normal distribution. While principal components does rely on linearity, I show, using a Kolmogorov Smirnov test, that formal employment of principal component based state variables does lead to a different empirical characterization of the log-likelihood density function for Gaussian affine term structure models.

The rest of the paper is organized as follows. Section II revisits the standard class of affine term structure models. Section III describes the data and extrapolation procedure. Section IV documents the persistence in the data on yields. Section V provides a comparison of the econometric bias introduced into the model through serial correlation in the yields in the eyes of two competing methods. Section VI shows the conditions under which the log-likelihood function under the two methods are asymptotically equivalent. Section VII concludes.

2.2 The Affine Class of Term Structure Models

The standard affine class of term structure models of Duffie and Kan (1996) and DS (2000) postulate state variables, X that evolve according to the following Markov N-dimensional transition dynamics under the equivalent risk neutral measure

$$dX_t = \kappa^Q(\theta^Q - X_t)dt + \Sigma\sqrt{S_t}dZ_t^Q$$

where κ^Q is an NxN matrix, θ^Q is an N-dimensional vector, Σ is an NxM matrix, and dZ_t^Q is a vector of M independent Brownian motions, where $(M \geq N)$. Furthermore, S is a diagonal MxM matrix that takes the following form

$$S_{ii,t} = \alpha_i + \beta_i^T X_t$$

α is an $Nx1$ vector of coefficients and β_i is the i^{th} row of β, an NxN matrix of coefficients.

The spot rate is also taken to be an affine function of X_t

$$r_t = \delta_o + \delta_1^T X_t$$

where δ_1 is an N-dimensional vector. According to this model, for a given maturity τ, zero coupon bond prices take the following exponential affine form

$$P(t,\tau) = \exp\left(a(\tau) + b(\tau)^T X_t\right)$$

where the state variables $a(\tau)$ and $b(\tau))$ solve the following system of ordinary differential equations

$$\frac{da(\tau)}{d\tau} = -\theta^{Q^T}\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\alpha_i - \delta_0$$

$$\frac{db(\tau)}{d\tau} = -\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\beta_i - \delta_1$$

These ordinary differential equations or Ricatti equations can be solved using standard Runge-Kutta numerical methods with initial conditions

$$a(0) = 0$$

$$b(0) = 0$$

Yields are defined, in terms of price, to be $P(t,\tau) = \exp(-\tau Y(t,\tau))$ which means that they can be written more succinctly as

$$Y(t,\tau) = A(\tau) + B(\tau)^T X_t$$

where $A(\tau) = -\frac{a(\tau)}{\tau}$ and $B(\tau) = -\frac{(b(\tau))}{\tau}$

The class of affine term structure models contains several models as special cases. This list of models includes, but is not limited to the models of Vasicek (1977), Langetieg(1980), Cox, Ingersoll, and Ross (1985), Jegadeesh and Pennacchi(1996), Christensen, Diebold, and Rudebusch (2008), and CDGJ(2008) (see Dejong (2000) for a more comprehensive list). As scholars have noted, these proposed models have later been determined to be a normalized or transformed version of a model that falls within the affine class of term structure models. All models that are estimated here fall within that general class. More specifically, they fall within the $A_0(3)$ class of models that are Gaussian in nature.

2.2.1 The $A_0(3)$ model

This paper focuses on the $A_0(3)$ since many papers use it as a base to study term structure models (see for example, Langetieg(1980), CDGJ(2008), Joslin, Singleton, and Zhu (2010a)). The nice closed form solutions afforded by the Gaussian setting yields a clean model environment from which new learning about the data can happen. In this spirit, I present the $A_0(3)$ model in terms of the state variables

$$d\begin{bmatrix} r_t \\ \mu^Q_{1,t} \\ \mu^Q_{2,t} \end{bmatrix} = \begin{bmatrix} \mu^Q_{1,t} \\ \mu^Q_{2,t} \\ -(\lambda^2\mu^Q_{1,t} + 2\lambda\mu^Q_{2,t}) \end{bmatrix} dt + \Sigma^R dB^Q_t$$

where Σ^R is an orthogonal matrix that is estimated numerically due to certain constraints associated with the transformation process (see Appendix A and/or Appendix B for more details). Yields are then taken to be a function of these state variables. According to CDGJ(2008), the presentation of the above system of stochastic differential equations for the state variable transition dynamics expedites the estimation and interpretation of multi-factor models.

The three state variables r,μ_1, and μ_2, by definition, are taken to be the level state variable, its drift, and the drift of the drift, respectively, under the equivalent risk-neutral measure.

2.3 Data

The dataset comprises daily observations of Libor rates with maturities of 1-month, 3-month, 6-month, 9-month, and 12-month and swap rate quotes for 2-years, 3-years, 4-years, 5-years, 7-years, and 10-years from June 21, 1996 until July 02, 2008, that

I collected from Bloomberg. This results in $3,034$ observations, after eliminating special days like Christmas or New Year's when no trading took place. Since Libor rates and swap rate quotes are given on a constant maturity basis, it is possible to construct a panel of zero coupon yields with maturities that matched the underlying instrument on one side of the swap. According to CDGJ(2008), this aspect of the Libor market and the swap market makes it more advantageous to use these quotes over Treasury quoted zero-coupon bond yields. This advantage comes at a cost which is that the swap rate quotes must be used to extrapolate the zero coupon yields. The details of that procedure are explained in Appendix C. Some descriptive statistics for the yields which resulted from those optimizations are shown in tables 1.1 and 1.2.

The minimum of 0.013836525 for the 10 year rate is due to the fact that the 7-year swap rate quote on that day was 0%. Excluding that minimum, the minimum of all other days in the sample was 0.029772228. For the 7-year rates, the minimum of 0.025563812 was due to the fact that several Libor quotes were around 1%.

2.4 Documenting Persistence in the data

To document persistence in the data, I follow Ludvigson and Ng (2009) and others who assume a first order vector auto regressive model for yields. Further, I compute the $AR(1)$ coefficient for each yield. I also report the R^2 from each regression. Those results are shown in table 2.1.

These results are also consistent with Ang and Piazzesi (2003) who find significant autocorrelation in the 1-month, 3-month, 12-month, 36-month, and 60-month yields. The autocorrelation is definitely the strongest at the short-end of the yield curve. Although, it tapers off as we move further out on the yield curve, it is quite strong for all horizons. Correlations in the actual data are similar to the correlations in the simulated data. At the short-end of the curve, the simulated data demonstrated smaller levels of persistence, while at the longer-end the actual yields displayed a lower level of persistence.

2.5 Analysis of Model-free methods

Potential sources of error in any method that is data driven arise from various aspects of the data. For instance, if data are noisy or are characterized by autocorrelation, then this could bias estimates and affect inference. These issues stem from the fact that the model-free method uses yield curve data to back out the state variables. I begin this section with a brief overview of this method.

CDGJ(2008) devise a method for estimating state variables, and subsequently for estimating yields. Because the model imposes "minimal" parametric structure without being dependent upon any particular model, they call this a model-free method. The authors claim that this parametric structure leads to advantages in estimation and interpretation of the yield curve, especially when examining short maturities (CDGJ, 2008).

CDGJ(2008) state that their framework is superior to more traditional frameworks used to measure state variables due to its direct tie to the short-end of the yield curve. CDGJ(2008) point out, for example, that in their setting, the three state variables r,μ_1, and μ_2, by definition, are taken to be the level state variable (which is akin to the level factor), its drift (which is akin to a slope factor), and the drift of the drift (which is akin to a curvature factor), respectively, under the equivalent risk-neutral measure. As CDGJ(2008) argue, since the level of the short rate is tied directly to the yield curve, its drift and the subsequent drift of the drift are also tied to the yield curve. According to the authors, these direct ties to the yield curve imply observability.[1] So, the state variables within the CDGJ(2008) set up consist of r and its first $N-1$ moments

The direct tie of the state variables to the yield curve is a result of rotations of the state vector consisting of the first few terms of a Taylor series expansion of the yield curve around a maturity of zero and their quadratic covariations (CDGJ, 2008). From an empirical standpoint, this method depends upon the usage of principal components analysis to extract the first three factors that explain the variation in the term structure of interest rates. It then uses least-square regressions to extract information from those factor loadings at short term maturities. Using a polynomial expansion, one can then tie the derivative expressions for these loadings back to derivatives of the yield curve. Finally, using a Taylor Series expansion and a simple but involved application of Ito's lemma, one can tie these derivatives to the state variables.

To assess what, if any, impact persistence or autocorrelation in the data have upon the estimation of state variables using the model-free method, I take the raw yield curve data from two models and run principal components analysis on two generated data sets. One data set consists of the raw data, the result of the application of the model-free method to this data set leads to the set of actual state variables, while the other involves pre-whitening the data in order to move toward a structure that has minimal dependence due to serial correlation. Pre-whitening the data involves removing the time-series autocorrelation from the data and has scarcely been applied

[1] Joslin, Singleton, and Zhu (2010) describe certain relevant practical drawbacks that are associated with the CDGJ(2008) normalization, including assuming that the short term interest rate is observable.

in finance applications, but seems to be widely used in other scientific fields (see for example, Andrews and Monahan (1992), Kao, Tamhane, and Mah (1992), Hamed (2009), Vargas-Guzman, Warrick, and Myers (1999), Edwards and Richardson (2004), Pyper and Peterman (1998), Milicich, Meekan, and Doherty (1992), or Quinn and Niebauer (1995)). In order to remove the time-series autocorrelation, I run a vector auto-regression for yields at each time series point on up to 5 lags, this intends to eliminate any persistence occuring during the previous week. The residual from the regression represents that information which is not explained by the yield. Applying principal components analysis on the resultant set of residuals provides a new set of yields from which to estimate a new set of state variables. Finally, I form the re-colored factors by taking the product of the whitened factor loadings and the actual time-series observations of the factors, as presumably the factor loadings are estimated with the noise inherited from the data. I call the state variables resulting from the application of the Taylor-series expansion to the colored factors the pre-whitened set of state variables, and compare them to the state variables arrived at from the simulated data without pre-whitening.

The results from this exercise are shown in the table 2.2. These results imply that the first state variable obtained from the model-free method is the most robust in the presence of serial correlation. It is less robust in the estimation of the second and third state variables. The correlation between the state variables before and after pre-whitening were 0.9897 for the first state variable, 0.2940 for the second state variable, and 0.4479 for the third state variable. These results taken together imply that the first state variable are estimated with the most precision, while the second and third state variables are estimated with less precision and are consistent with CDGJ(2008). These ideas can be reinforced from the table above into an easily interpreted conclusion that essentially points to the impact of whitening on model estimation. On the whole, it is clear that whitening does affect model estimation, as it gives two different sets of results. Monte Carlo simulations can provide the necessary details to determine the direction of the effect of whitening on model estimation.

2.6 Simulation-based Analysis of Model-free Method

In this section, I construct a Monte Carlo algorithm that simulates yield curve data from a three factor Vasicek model and then runs principal components analysis on the generated data set. The parameters used were κ_1=2, κ_2=0.2, κ_3=1, $\sigma_1 = 0.04$, σ_2=0.002, σ_3=0.005, θ_1=0.01231, θ_2=0.0062, θ_3=0.0031, λ_1=0, λ_2=0, λ_3=0. These parameters were chosen for various reasons. The parameters provide a long run mean of about 4.25%, which is close to 4.16%, which is the mean of the 1-month Libor rate in the dataset. The long run variance is about 1% for the first factor, 0.5% for the second factor, and 0.25% for the third factor. For each factor, λ was set to 0 in order to minimize the discrepancy from the model of CDGJ(2008), who work in the equivalent risk-neutral measure, and so this simplification allows the historical measure and the equivalent martingale measure to be identical. Finally, it should be noted that these parameters were chosen so as to prevent the yield curve from going into negative space. The procedure then repeats the above for a large number of trials, with one little twist.

The Monte Carlo Simulation procedure provides the actual state variable, following the model-free method, that is estimated from the simulated yield curve. Hence, in each simulation, two sets of regressions are carried out; the estimated pre-whitened state variable on the actual state variable and the estimated state variable without pre-whitening on the actual state variable.[2] These regressions provide coefficients for α, β, R^2,from the regressions. Following CDGJ (2008), I consider three cases, a zero error case, a low-error case in which I add an error standard deviation of 0.5 basis points, and a high error case in which I add an error standard deviation of 2.0 basis points.

The procedure essentially replicates the construction of the model-free state variables. The table below reports means of the coefficients and the R^2 from the following regression $Estimated\ State\ Variable(t) = \alpha + \beta\ Actual\ State\ Variable(t) + \epsilon(t)$ based upon 100 simulations.

The descriptive statistics from those simulations are presented in table 2.3. The results above provide further evidence for the effects of whitening on the estimation of model free state variables. Again, it is clear that even in a high-error scenario in which I add Gaussian errors with a 2 basis point standard deviation, the first state variable is estimated fairly accurately. Correlation coefficients between the estimated model-free state variables and the actual state variables were consistent with the above explanation. Specifically, before prewhitening, for the zero-error case, $\rho_{r_{act},r_{est}}$=1.000, $\rho_{\mu_{1,act},\mu_{1,est}}$=1.000, and $\rho_{\mu_{2,act},\mu_{2,est}}$=1.000. In the low-error case, $\rho_{r_{act},r_{est}}$=1.000, $\rho_{\mu_{1,act},\mu_{1,est}}$=1.000, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.999. In the high-error case, $\rho_{r_{act},r_{est}}$=1.000, $\rho_{\mu_{1,act},\mu_{1,est}}$=1.000, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.998. After prewhitening, the correlations were, for the zero-error case, $\rho_{r_{act},r_{est}}$=0.999, $\rho_{\mu_{1,act},\mu_{1,est}}$=0.966, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.962. For the low error case, $\rho_{r_{act},r_{est}}$=0.999, $\rho_{\mu_{1,act},\mu_{1,est}}$=0.966, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.962. Finally, in the high-error case, $\rho_{r_{act},r_{est}}$=0.999, $\rho_{\mu_{1,act},\mu_{1,est}}$=0.966, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.958. The first state variable is the least affected, while the second

[2] More detailed examinations of prewhitening within a simulation based study have been carried out in other fields (see, for example, Bayazit and Onoz (2009) , Yue and Wang (2002), Christou and Pittis (2002), or Cappuccio and Lubian (1996)).

and third state variables, on the other hand, are more affected. Notice that following the application of the prewhitening procedure, each state variable, in all three cases, is characterized by similar coefficients. Should the zero-error case serve as a benchmark, this speaks to the application of the whitening procedure. Recall from CDGJ (2008) that if the model-independent state variable estimates are accurate and unbiased, then we should expect to see intercepts close to 0 and slope coefficients close to 1.

Results were robust to the application of a Kolmogorov-Smirnov test, as for the zero error case, whitening did not seem to influence the underlying distribution of any of the state variables. The test failed to reject the null hypothesis that the state variables were from the same underlying distribution. However, after prewhitening and for the high-error case, the test was, on average, more likely to reject the aforementioned null hypothesis than for the low-error and zero-error case. This was especially true for the third state variable.

Whitening the data seems to play a bigger role in the low and high error cases. Consequently, in the table below, I repeat the same exercise except that I add a twist. For the table below, I eliminate the zero-error case and instead I add serially correlated errors to the simulated yields. I assume that the errors follow a first order auto regressive process with an autocorrelation coefficient of 0.8.[3] I consider a low-error case in which I assume that incorporated into the yield curve is serially correlated errors with a 0.5 basis point standard deviation and similarly I add a high error case which has a 2.0 basis point standard deviation. These standard deviations correspond to 0.3 basis points for the low error case and 1.2 basis points for the high

[3]This application of prewhitening to remove serial correlation due to the use of an AR(1) is also employed by Bryson and Henrikson (1968). Its application to remove serial correlation arising from more general processes has been explored by Kao et. al (1992) and others

error case, after I adjust the variance of the draw to be identical to the variance of the auto regressive process. Adding the errors as described above to each simulated set of yields gives me a new set of yields. I repeat the construction of the model-free state variables several times with different data sets generated based upon the above criteria.

In table 2.4, I report the results of the simulations above with $N = 100$. These results have some very interesting implications. The first implication is that, for the low-error case, each state variable seems to be estimated relatively precisely. I draw this conclusion from the fact that the estimated state variables have relatively high R^2 values, the α values are close to 0, and β is close to 1. Recall from CDGJ (2008) that if the model-independent state variable estimates are accurate and unbiased, then we should expect to see intercepts close to 0 and slope coefficients close to 1. On the whole, whitening seems to improve the model estimation in the low error case. These results are essentially the same as the previous exercise.

With regards to the high error case, the state variables are definitely more biased. After prewhitening, there is almost no difference between the low and high error cases, which is consistent with the above exercise. It is again clear that pre-whitening has had an effect on the two state variables to the effect of making them more aligned on average. After pre-whitening and for the high error case, the Kolmogorov-Smirnov test was more likely to reject the null hypothesis that the state variables arise from the same underlying distribution than in the low-error case. It was especially pronounced for the third state variable.

Correlation coefficients between the estimated model-free state variables and the actual state variables were consistent with the above explanation. Specifically, before prewhitening, in the low-error case, $\rho_{r_{act},r_{est}}$=1.000, $\rho_{\mu_{1,act},\mu_{1,est}}$=1.000, and $\rho_{\mu_{2,act},\mu_{2,est}}$=1.000.

In the high-error case, $\rho_{r_{act},r_{est}}$=1.000, $\rho_{\mu_{1,act},\mu_{1,est}}$=1.000, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.994. After prewhitening, the correlations were, for the low error case, $\rho_{r_{act},r_{est}}$=0.999, $\rho_{\mu_{1,act},\mu_{1,est}}$=0.967, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.960. Finally, in the high-error case, $\rho_{r_{act},r_{est}}$=0.999, $\rho_{\mu_{1,act},\mu_{1,est}}$=0.967, and $\rho_{\mu_{2,act},\mu_{2,est}}$=0.958.

On the whole, prewhitening the data seems to improve the model estimation in the presence of persistence. The results from the Kolmogorov-Smirnov test indicate that in the presence of prewhitening the state variables are generally characterized by the same underlying distribution. There are a handful of cases where prewhitening the state variables seemed to affect the characterization of the underlying distribution of the state variables.

Overall, these results are consistent with the actual data. Prewhitening does impact the estimation of state variables as estimated via principal components analysis. These results indicate that pre-whitening does matter, and should be considered in the application of principal components analysis.

2.7 Observational Equivalence of Log-Likelihood Function for model-free method and inversion method

In this section, I show the observational equivalence between principal component factors, as derived in CDGJ(2008) and the factors obtained from Joslin, Singleton, and Zhu (2010) or Dai and Singleton (2000). The purpose of this exercise is to obtain a representation for the model-free method that can be used for direct comparison to Dai and Singleton (2000), from the standpoint of empirical density construction. I start with a review the method that CDGJ (2008) use to compute state variables.

The method associated with CDGJ(2008) is contingent upon the fact that three factors drive most of the explained variation the yield curve according to Litterman and Schienkman (1991). This approach imposes some parametric structure while not being dependent upon any model. It uses principal components analysis to form derivatives of the yield curve and then a Taylor Series Expansion about a maturity of zero to tie these to the back to the level of the yields at the short end of the curve. A simple but messy application of Ito's lemma ties these derivatives to state variables.

Specifically, it relies upon the use of principal components analysis to extract factor loadings from yields with the following maturities; 1 month, 3 month, 6 month, 12 month, 2 year, 3 year, 5 year, and 7 year. Then, I use orthogonality of the resulting factor loading matrix to obtain the principal component factors. The next step is to carry out the following regressions; regress the first factor on a vector of ones and a vector τ which is a vector that contains the times to maturity at very short horizons, regress the second factor on the set above plus a term quadratic in the time to maturity vec, and finally regress the third factor on the set used in the second regression plus a term that is cubic in the time to maturity vector. These regressions give me intercept terms plus coefficients that I use to construct n^{th} order derivatives. Please note the following approximation which is a result of the work of Litterman and Schienkman (1991) and others:

$y(t,\tau) \cong \sum_{k=1}^{3} f_k(\tau) P_k(t)$

where $P_k(t)$ is the k^{th} principal component factor and $f_k(t)$ is the factor loading function (i.e. polynomial expansion for the k^{th} factor).

Taking the n^{th} order derivative with respect to τ of both sides,

$\frac{\partial Y^n(t,\tau)}{\partial \tau}|_{\tau=0} \cong \sum_{k=1}^{3} \frac{\partial^n f_k(\tau)}{\partial \tau^n} P_k(t)$

A Taylor Series expansion of the yield curve about short maturities connects the above derivatives to state variables. Specifically, the application of Ito's lemma on the Markov state vector $\{X(t)\}$, with length N (essentially N is the number of state variables), shown below, in the equivalent risk-neutral measure.

$$dX_i = m_i^Q(X)dt + \sum_{k=1}^{N} \sigma_{ik}(X)dz_i^Q$$

Further, assume that the spot rate is an arbitrary function of the state variables, i.e. $r(X)$. If I apply Ito's formula to the Markov State vector above, I end up the following expression.

$$dr = \sum_{i=1}^{N} \frac{\partial r}{\partial X_i}[m_i^Q dt + \sum_{i=1}^{N} \sigma_{ik} dz_i^Q] + \frac{1}{2}\sum_{i,j,k=1}^{N} \frac{\partial^2 r}{\partial X_i \partial X_j}\sigma_{ik}\sigma_{jk}dt$$

From here we can derive the equivalent risk-neutral drift, its variance and the drift of the drift (although this needs Ito's formula again.) Then, I define the price of a zero coupon bond in terms of state variables with maturity T as follows:

$$P^T(t, X_t) \equiv e^{-\tau Y(X_t,\tau)}$$

Taking first order derivative with respect to time to maturity, state variables, and second order cross-partial derivatives with respect to pairs of state variables allows me to connect those derivatives to the bond pricing equation. Proceeding step by step, the computations of the partial derivatives are shown below.

$$P_\tau = \frac{\partial P}{\partial \tau} = \frac{\partial(e^{-\tau Y(\{X_t\},\tau)})}{\partial \tau} = [-\tau\frac{\partial Y(\{X_t\},\tau)}{\partial \tau} - Y(\{X_t\},\tau)]e^{-\tau Y(\{X_t\},\tau)} = [-\tau Y_\tau - Y]P$$

$$P_i = \frac{\partial P}{\partial X_i} = \frac{\partial(e^{-\tau Y(\{X_t\},\tau)})}{\partial X_i} = -\tau\frac{\partial(Y(\{X_t\},\tau))}{\partial X_i}e^{-\tau Y(\{X_t\},\tau)} = -\tau Y_i P$$

$$P_{ij} = \frac{\partial P_i}{\partial X_j} = \frac{\partial(-\tau Y_i P)}{\partial X_j} = -\tau[\frac{\partial P}{\partial X_j}Y_i + \frac{\partial Y_i}{\partial X_j}P] = \tau[Y_i\tau Y_j P] - \tau Y_{ij}P$$

Plugging the above partial derivatives into the bond pricing equation,

$$rP = -P_\tau + \sum_{i=1}^{N} P_i m_i^Q + \frac{1}{2}\sum_{i,j,k}^{N} P_{ij}\sigma_{ik}\sigma jk$$

$$\rightarrow rP = [-\tau Y_\tau - Y]P + \sum_{i=1}^{N} -\tau Y_i P m_i^Q + \frac{1}{2}\sum_{i,j,k}^{N}[\tau[Y_i\tau Y_j P] - \tau Y_{ij}P]\sigma_{ik}\sigma jk$$

Note that this equation is consistent with the absence of arbitrage. Plugging the derivatives as described above into the bond pricing equation and then using a Taylor Series expansion as such

$$Y(X_t, \tau) \equiv Y^0(X_t) + \tau Y^1(X_t) + \frac{1}{2}\tau^2 Y^2(X_t) + \dots$$

to write yields and plugging that into the bond pricing equation with the derivatives leads to the desired result. Specifically, we need to compute another set of partial derivatives of the yields with respect to the state variables

$$Y_\tau = Y^1(\{X_t\}) + \tau Y^2(\{X_t\}) + \dots$$

$$Y_i = Y_i^0(\{X_t\}) + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \dots$$

$$Y_j = Y_j^0(\{X_t\}) + \tau Y_j^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_j^2(\{X_t\}) + \dots$$

$$Y_{ij} = Y_{ij}^0(\{X_t\}) + \tau Y_{ij}^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_{ij}^2(\{X_t\}) + \dots$$

Then, plugging these derivative expressions into the above bond pricing equation, collecting terms and simplifying leads to

$$\rightarrow r = ([Y^0(\{X_t\}) + \tau Y^1(\{X\}) + \frac{1}{2}\tau^2 Y^2(\{X_t\})] + \tau[Y^1(\{X_t\}) + \tau Y^2(\{X_t\}) + \dots]) - \tau \sum_{i=1}^N (Y_i^0(\{X_t\}) + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \dots) m_i^Q + \frac{1}{2}\sum_{i,j,k}^N \tau^2([(Y_i^0 + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \dots)(Y_j^0(\{X_t\}) + \tau Y_j^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_j^2(\{X_t\}) + \dots)] - \tau[Y_{ij}^0(\{X_t\}) + \tau Y_{ij}^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_{ij}^2(\{X_t\})])\sigma_{ij}\sigma_{jk}$$

Finally, collecting terms in τ,

$$r = Y^0(\{X_t\}) + \tau[Y^1(\{X_t\}) + Y^1(\{X_t\}) - \sum_{i=1}^N Y_i^0(\{X_t\}) m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^N Y_{ij}^0(\{X_t\})\sigma_{ij}\sigma_{jk}] + \tau^2[\frac{1}{2}Y^2(\{X_t\}) + Y^2(\{X_t\}) - \sum_{i=1}^N Y_i^1(\{X_t\}) m_i^Q + \frac{1}{2}Y_i^0(\{X_t\})Y_j^0(\{X_t\})\sigma_{ij}\sigma_{jk} - \frac{1}{2}\sum_{i,j,k=1}^N Y_{ij}^1(\{X_t\})\sigma_{ij}\sigma_{jk}]$$

Because r is the level of the short rate (i.e. the first term in a Taylor series is the function evaluated at a, where here a is approximately zero, so r is the level of interest rate at short maturities, or the level of the short rate)

Hence, the first three state variables are

$Y^0(t) = r(X_t)$

$Y^1(X_t) = \frac{1}{2}\mu_1(X_t)$

$Y^2(X_t) = \frac{1}{3}[\mu_2(X_t) - V(t)]$

where $\mu_1(t) = \sum_{i=1}^{N} Y_i^0(\{X_t\})m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^{N} Y_{ij}^0(\{X_t\})\sigma_{ij}\sigma jk$

$Y_i^0(\{X_t\}) = \frac{\partial r}{X_i}$

$Y_{ij}^0(\{X_t\}) = \frac{\partial^2 r}{X_i X_j}$

$V(t) = \sum_{i,j,k=1}^{N} Y_i^0(\{X_t\})Y_j^0(\{X_t\})\sigma_{ij}\sigma_{jk}$

$\mu_2 = \sum_{i=1}^{N} Y_i^1(\{X_t\})m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^{N} Y_{ij}^1(\{X_t\})\sigma_{ij}\sigma_{jk}$

These are my set of state variables.

Joslin, Singleton, and Zhu (2010) show that making the following definition, in the three factor setting, leads to

$$X_t = \{r_t, \mu_1, \mu_2\}$$

with the state variables defined as above, defining the following two matrices,

$$K_{0,CDGJ}^{X,Q} = \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}$$

$$K_{1,CDGJ}^{X,Q} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mu_0 & \mu_1 & \mu_2 \end{bmatrix}$$

Then, under Q, X_t follows,

$$dX_t = (K_{0,CDGJ}^{X,Q} + K_{1,CDGJ}^{X,Q} X_t)dt + \Sigma_X dW_t$$

Following the description above regarding the relationship between the state variables and the Principal Component factors leads to the following system of equations,

$$r = a_1 PC_1 + a_2 PC_2 + a_3 PC_3$$

$$\mu_1 = 2[b_1 PC_1 + b_2 PC_2 + b_3 PC_3]$$

$$\mu_2 = 3[2c_1 PC_2 + 2c_2 PC_3] + var(\Delta y)]$$

In matrix form,

$$\begin{bmatrix} r \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ 2b_1 & 2b_2 & 2b_3 \\ 0 & 6c_1 & 6c_2 \end{bmatrix} \begin{bmatrix} PC_1 \\ PC_2 \\ PC_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ var(\Delta y) \end{bmatrix}$$

Inverting the coefficient matrix, assuming its non-singular, and solving for the principal component factors leads to the following matrix equation,

$$Y = AX + b$$

where the following definitions are in order,

$$Y = \begin{bmatrix} PC_1 \\ PC_2 \\ PC_3 \end{bmatrix}$$

$$A = \begin{bmatrix} \frac{12(b_2c_2 - b_3c_1)}{M} & \frac{6(a_3c_1 - a_2c_2)}{M} & \frac{2(a_2b_3 - b_2a_3)}{M} \\ \frac{6(a_3c_1 - a_2c_2)}{M} & \frac{6a_1c_2}{M} & \frac{2(b_1a_3 - a_1b_3)}{M} \\ \frac{2(a_2b_3 - b_2a_3)}{M} & \frac{-6a_1c_1}{M} & \frac{2(b_2a_1 - a_2b_1)}{M} \end{bmatrix}$$

$$M = 12[a_1b_2c_2 - a_1c_1b_3 - a_2b_1c_2 + b_1c_1a_3]$$

$$b = \begin{bmatrix} 0 \\ 0 \\ -var(\Delta(y)) \end{bmatrix}$$

$$dX_t = (K^{X,Q}_{0,CDGJ} + K^{X,Q}_{1,CDGJ} X_t)dt + \Sigma_X dW_t$$

$$K^{X,Q}_{0,CDGJ} = \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}$$

$$K^{X,Q}_{1,CDGJ} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mu_0 & \mu_1 & \mu_2 \end{bmatrix}$$

A simple application of Ito's lemma allows me to obtain the stochastic differential equation for the principal component factors. This allows one to draw inference for the density function for the principal component factors, which can be used to compare to other sets of state variables implied by the "inversion" method. It is shown below,

$$Y_t = AX_t + b$$

In the expression above, as with an invariant affine transformation, I require that A is square and non-singular, then after taking differentials and applying Ito's lemma,

$$dY_t = AdX_t$$

$$\rightarrow dY_t = A[(K^{X,Q}_{0,CDGJ} + K^{X,Q}_{1,CDGJ} X_t)dt + \Sigma_X dW_t]$$

$$\rightarrow dY_t = (AK^{X,Q}_{0,CDGJ} + AK^{X,Q}_{1,CDGJ} X_t)dt + A\Sigma_X dW_t$$

$$\rightarrow dY_t = (AK^{X,Q}_{1,CDGJ})[(AK^{X,Q}_{1,CDGJ})^{-1} AK^{X,Q}_{0,CDGJ} + ((AK^{X,Q}_{1,CDGJ})^{-1})((AK^{X,Q}_{1,CDGJ}))X_t]dt + A\Sigma_X dW_t$$

$$\rightarrow dY_t = (AK^{X,Q}_{1,CDGJ} A^{-1})[A(AK^{X,Q}_{1,CDGJ})^{-1} AK^{X,Q}_{0,CDGJ} + AX_t - b + b]dt + A\Sigma_X dW_t$$

$$\rightarrow dY_t = (AK^{X,Q}_{1,CDGJ} A^{-1}(K^{X,Q}_{1,CDGJ})^{-1})[(K^{X,Q}_{1,CDGJ})A(AK^{X,Q}_{1,CDGJ})^{-1} AK^{X,Q}_{0,CDGJ} - (K^{X,Q}_{1,CDGJ})b + K^{X,Q}_{1,CDGJ} Y_t]dt + A\Sigma_X dW_t$$

$$\rightarrow dY_t = [AK^{X,Q}_{0,CDGJ} - AK^{X,Q}_{1,CDGJ}A^{-1}b + AK^{X,Q}_{1,CDGJ}A^{-1}Y_t]dt + A\Sigma_X dW_t$$

$$\rightarrow dY_t = [K^{Y,Q}_{0,CDGJ} + K^{Y,Q}_{1,CDGJ}Y_t]dt + \Sigma_Y dW_t$$

where

$$K^{Y,Q}_{0,CDGJ} = AK^{X,Q}_{0,CDGJ} - AK^{X,Q}_{1,CDGJ}A^{-1}b$$

$$K^{Y,Q}_{1,CDGJ} = AK^{X,Q}_{1,CDGJ}A^{-1}$$

$$\Sigma_Y = A\Sigma_X$$

As mentioned above, this expression will allow a link between the empirical density function for an affine term structure model that uses principal component based factors to describe the transition dynamics for the state variable to empirical density functions that have been traditionally employed to estimate affine term structure models. If the data are Gaussian, then there should not be any meaningful statistical differences across the two methods. These conclusions are verified through a comparison between the empirical density function using the above method and the empirical density function as obtained using the inversion method when the data are Gaussian so that the maximum likelihood estimation procedure is exact. The results of a Kolmogorov-Smirnov test are shown in table 2.5.

It should be noted that in the implementation of the test above is in a setting that is less general than that of CDGJ (2008), yet still remains in the affine class. Specifically, the results are applicable if, while retaining the setting of CDGJ (2008), we restrict $\gamma = 0$, $\mu_0 = 0$, $\mu_1 = \lambda^2$, and $\mu_2 = 2\lambda$.

The results are very interesting. Notice that the first comparison is a type of benchmark test in the following sense. The application of Ito's lemma to the state variables obtained in CDGJ (2008) leads to a term structure model that directly employs principal component based factors as state variables. Hence, one would expect that an algebraic transformation of the same model would not lead to different probabilistic description for the empirical log-likelihood function. This is indeed the case. Not surprisingly, with an asymptotic $p - value$ of 0.0971, at the 1% level, we fail to reject the null hypothesis that the empirical density function for the model-free method is different from the empirical density function described by principal component based state variables. On the other hand, the probabilistic description of the empirical density function associated with a Kalman Filter is in fact different from that which employs principal component based state variables to estimate the likelihood density function. These results taken together imply that despite the fact that, among other things, PCA assumes linearity and that the mean and variance are sufficient statistics for the probability density function, its usage in a term structure model seems to introduce additional complexities that create deviations from a term structure model that also, by virtue of the implementation using a Kalman filter, relies on linearity and normality (Shlens, 2005). Normality assumes that the mean and variance are sufficient statistics for the data, but if the mean and variance are the only parameters needed to characterize the data that does not imply that a Gaussian description is adequate for the state variables transition dynamics.

From a theoretical perspective, these results provide further empirical support for the findings of Joslin, Singleton, and Zhu (2010) that the two methods are in fact observationally equivalent. However, empirically the results of this paper provide impetus for further study.

2.8 Conclusion

This paper examined the estimation of affine term structure models when data are persistent. Observed yields are highly autocorrelated. Traditionally, scholars have attributed strong persistence in yields to strong persistence in factors. This paper took a different approach. Taking yields, including all their properties like serial correlation, as given, I compared alternative computational methods for state-dependent model-implied factors.

Serial correlation does seem to affect the estimation of state variables and hence model-implied factors. This suggests several avenues of future research. Gorodnichenko, Mikusheva, and Ng (2009) obtain a persistence robust estimator using vector auto-regressive frameworks and some extensions. Using the ideas in that paper, it may be interesting to extend the analysis to affine term structure models.

Overall, this paper provides new insight into how to estimate state variables corresponding to affine term structure models in a noisy environment.

TABLE 2.1. Persistence in the Panel of Yields

i	$AR1(Y_{it})$	R_i^2
1	0.9997	0.9994
2	0.9999	0.9997
3	0.9998	0.9996
4	0.9997	0.9995
5	0.9995	0.9993
6	0.9990	0.9984
7	0.9990	0.9983
8	0.9986	0.9977
9	0.9983	0.9972
10	0.9969	0.9943
11	0.9969	0.9944
12	0.9567	0.9157
13	0.9731	0.9476
14	0.9885	0.9778

For $i = 1, \ldots, 14$ where i is the time to maturity of the bond. The data is taken at daily frequencies from June 1996 through July 2008, a total of 3,034 observations. The model used to obtain these coefficients is

$$Y_{i,t+1} = \alpha + \beta Y_{i,t} + \epsilon$$

Table 2.2. Check for the effect of pre-whitening the data on the fit of the factors.

SV		r			μ_1			μ_2	
Par	$\widehat{\alpha}$	$\widehat{\beta}$	R^2	$\widehat{\alpha}$	$\widehat{\beta}$	R^2	$\widehat{\alpha}$	$\widehat{\beta}$	R^2
	-0.0092	1.5290	0.9795	0.0068	0.0288	0.0865	-0.0045	0.0795	0.2006

The table reports descriptive statistics of the coefficients and the R^2 from the following regression

$$Non{-}Prewhitened\ State\ Variable(t) = \alpha + \beta * Prewhitened\ State\ Variable(t) + \epsilon(t)$$

TABLE 2.3. This table reports regression results based upon simulated interest rates.

Panel A									
SV		r			μ_1			μ_2	
Par	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2
0	-0.000	1.000	1.000	-0.000	1.000	1.000	-0.000	1.000	1.000
0.5	0.000	1.000	1.000	-0.009	1.000	1.000	0.002	0.997	0.999
2.0	-0.000	1.000	1.000	0.002	1.000	1.000	0.005	0.999	0.996
Panel B									
SV		r			μ_1			μ_2	
Par	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2
0	-0.129	1.100	0.998	0.495	1.310	0.937	-0.501	0.861	0.926
0.5	-0.128	1.100	0.998	0.485	1.310	0.937	-0.375	0.870	0.927
2.0	-0.131	1.100	0.998	0.520	1.310	0.936	-0.765	0.850	0.920

TABLE 2.4. This table reports regression results based upon simulated interest rates with serially correlated errors.

Panel A									
SV		r			μ_1			μ_2	
Par	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2
0.5	-0.000	1.000	1.000	-0.000	1.000	1.000	-0.000	0.998	0.999
2.0	-0.001	1.000	1.000	0.000	1.000	0.999	-0.001	0.992	0.988
Panel B									
SV		r			μ_1			μ_2	
Par	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2	$1000 * \widehat{\alpha}$	$\widehat{\beta}$	R^2
0.5	0.003	1.102	0.999	0.006	1.344	0.937	0.031	0.863	0.923
2.0	0.003	1.102	0.999	0.006	1.344	0.936	0.031	0.883	0.920

TABLE 2.5. Application of Kolmogorov-Smirnov Test to compare empirical density functions under alternative computational methods used to estimate an $A_0(3)$ model

$KSTest$	$Model - Free$ v.s. PCA	$Kalman\ Filter$ v.s. PCA
$TestStat$	1	0.8916
$p - Value$	0.0971	0

Chapter 3

On the Estimation Risk in the Numerical Integration of Affine Term Structure Models.

3.1 Introduction

The class of affine term structure models of Duffie and Kan (1996) is widely studied for several reasons; it has a good theoretical foundation, is empirically tractable, and has the flexibility to accommodate different specifications for risk premia and the volatility of the state variable. Duffie, Pan, and Singleton (2000) use transform analysis to obtain affine term structure models, in the presence of jump diffusions, which can characterize analytically tractable pricing relations for many financial valuation problems. Chen and Joslin (2009) significantly extend the analysis of Duffie, Pan, and Singleton (2000) to account for defaultable securities. Joslin, Singleton, and Zhu (2010) develop a new class of canonical Gaussian term structure models that is both analytically tractable and has appealing properties from an econometric standpoint. (See Collin-Dufresne, Goldstein, and Jones (2008) for a more complete list of examples demonstrating the strong tractability of Affine Term Structure Models).

In its canonical form, the class of affine term structure models is characterized by N state variables, $X = \{X_1, X_2, \ldots, X_n\}$, that are postulated to follow Markov transition dynamics. Prices are then taken to be exponential affine functions of these latent or unobservable state variables. The components of this relationship also include state factor coefficients that describe how the state variables relate back to model-implied yields. These state factor coefficients solve standard matrix Ricatti equations (see, for example, Dai and Singleton (2000), Collin-Dufresne, Goldstein, and Jones (2009), Boulder (2001), Piazzesi (2003), or Fisher and Gilles (1996) for a more expository explanation). As is well known in the literature, save for the case where the κ matrix and the Σ matrix are diagonal (i.e. all the factors are multi-independent). There is not a general solution to these matrix-valued differential equations (Grasselli and Tebaldi, 2007). Instead, researchers must rely on numerical algorithms, such as Runge-Kutta methods, to solve such non-linear matrix differential equations (see for example, Piazzesi (2003), Duffee (2002), Dai and Singleton (2000)). This implies that we give up analytic tractability in order to explore more complex, and perhaps more realistic descriptions of the term structure of interest rates. Boulder (2001) notes that for a multi-independent CIR (1985) term structure model the desire for tractability forces us to neglect any correlation between state variables. It is reasonable to believe that state variables may be correlated (see, for example, Christensen, Diebold, and Rudebusch (2008,2009) or Collin-Dufresne, Goldstein, and Jones (2008), Collin-Dufrense, Goldstein, and Jones (2002), or Singleton (2006)) so that, it is worth the cost of additional complexity and estimation risk. Furthermore, these latent state variables attempt to proxy for macroeconomic variables. Certain macroeconomic variables tend to be related, for example, the short term interest rate and inflation would be strongly correlated (see for example, Ang, Dong, and Piazzesi (2007) or Ang and Piazzesi (2003)).

While this class of models has been the source of much study by leading scholars in the field of finance, no one has yet studied the econometric tradeoffs between analytical tractability and computational flexibility regarding the estimation of affine term structure models. This paper fills in this gap.

The primary objective of this paper is to examine the estimation risk associated with numerical integration. I examine this issue by simulating data from two versions of a Gaussian dynamic term structure model. The first version of the model is restricted in such a way that there exists a closed form solution for the state factor coefficients and so the data generating process obtains from a model that is analytically tractable. The second version of the model is completely unrestricted so that Runge-Kutta methods need to be used to solve the Ricatti ordinary differential equations in order to arrive at the state factor coefficients. In this case, the data generating process obtains from a model that while it is completely unrestricted so that it can provide consistency to more economic scenarios, it has an additional layer of complexity that stems from a lack of analytic tractability.

For each version of the model, I generate a panel of zero coupon yields. I then estimate the Gaussian dynamic term structure model in both the completely unrestricted and hence generalized framework, and in the restricted framework. I repeat this process for a large number of trials. Finally, across the two methods, I compare the statistical bias, statistical efficiency, and economic accuracy for actual yields.

For the sake of computational feasibility, I estimate and perform Monte Carlo simulations within the 2 factor setting. This paper concludes that regardless of the data generating process, the generalized Gaussian process seems to lead to less biased parameter estimates and more efficiently generated yields than a Vasicek model. However, when the data are generated from the Vasicek model, on the whole, the Vasicek model is more accurate. Moreover, as suggested by Bolder (2001), usage of the more tractable Vasicek model implies giving up correlation between state variables. The generalized Gaussian model is more accurate than the Vasicek model in the more general setting. This suggests that is necessary to use numerical Runge-Kutta methods to solve the Ricatti ordinary differential equations to arrive at the state variable factor coefficients.

However, not surprisingly, the one advantage to estimating the model via the restricted Vasicek model is the reduced computational time with regards to the optimization process. Herein, lies the tradeoff. Specifically, when the data was generated from a Vasicek model, the average time to complete the maximum likelihood estimation for a Vasicek model was 277.44 seconds which is about 5 minutes, while the average time to completion for the generalized Gaussian model was 7,313.89 seconds which is about 122 minutes or just over 2 hours. Similarly, when data was generated from the generalized Gaussian model, the average time to completion for the Vasicek model was 382.78 seconds or about 7 minutes, while the generalized Gaussian model optimized in about 145 minutes, about 2.5 hours.

The rest of the paper is organized as follows. Section II revisits the standard class of affine term structure models. Section III describes the data and extrapolation procedure. Section IV examines the estimation risk of numerical integration within a 3 factor Gaussian dynamic term structure model. Section V uses monte carlo simulation based methods to examine the estimation risk in numerical integration within a 2 factor setting. Section VI examines the stability of the state variables estimated under analytic and numerical methods. Section VII examines the role that the error term plays within the class of affine term structure models. Section VIII concludes.

3.2 The Affine Class of Term Structure Models

The standard affine class of term structure models of Duffie and Kan (1996) and DS (2000) postulate state variables, X that evolve according to the following Markov N-dimensional transition dynamics under the equivalent risk neutral measure

$$dX_t = \kappa^Q(\theta^Q - X_t)dt + \Sigma\sqrt{S_t}dZ_t^Q$$

where κ^Q is an NxN matrix, θ^Q is an N-dimensional vector, Σ is an NxM matrix, and dZ_t^Q is a vector of M independent brownian motions, where $(M \geq N)$. Furthermore, S is a diagonal MxM matrix that takes the following form

$$S_{ii,t} = \alpha_i + \beta_i^T X_t$$

α is an $Nx1$ vector of coefficients and β_i is the i^{th} row of β, an NxN matrix of coefficients.

The spot rate is also taken to be an affine function of X_t

$$r_t = \delta_o + \delta_1^T X_t$$

where δ_1 is an N-dimensional vector. According to this model, for a given maturity τ, zero coupon bond prices take the following exponential affine form

$$P(t,\tau) = \exp\left(a(\tau) + b(\tau)^T X_t\right)$$

where the state variables $a(\tau)$ and $b(\tau))$ solve the following system of ordinary differential equations

$$\frac{da(\tau)}{d\tau} = -\theta^{Q^T}\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\alpha_i - \delta_0$$

$$\frac{db(\tau)}{d\tau} = -\kappa^{Q^T}b(\tau) + \frac{1}{2}\sum_{i=1}^{M}[\Sigma^T b(\tau)]_i^2\beta_i - \delta_1$$

These ordinary differential equations or Ricatti equations can be solved using standard Runge-Kutta numerical methods with initial conditions

$$a(0) = 0$$

$$b(0) = 0$$

This paper focuses on differences as regards parameter estimates and model fit when using the aforementioned standard Runge-Kutta methods for which Matlab provides solutions through the **odesolver** toolkit versus when these state variable coefficients can be solved for explicitly through the following expressions.

$b_i(\tau) = \frac{1}{\kappa_i}(1 - exp(-\kappa_i\ \tau))$

$a_i(\tau) = \frac{\gamma_i(b_i(\tau)-\tau)}{\kappa_i*\kappa_i} - \frac{\sigma_i*\sigma_i*b_i*b_i}{4*\kappa_i}$

where i delineates the i^{th} factor (see Bolder (2001) for more details).

Yields are defined, in terms of price, to be $P(t,\tau) = \exp(-\tau Y(t,\tau))$ which means that they can be written more succinctly as

$$Y(t,\tau) = A(\tau) + B(\tau)^T X_t$$

where $A(\tau) = -\frac{a(\tau)}{\tau}$ and $B(\tau) = -\frac{(b(\tau))}{\tau}$

The class of affine term structure models contains several models as special cases. This list of models includes, but is not limited to the model of Vasicek (1977), Langetieg(1980), Cox, Ingersoll, and Ross (1985), Jegadeesh and Pennacchi(1996), Christensen, Diebold, and Rudebusch (2008), Christensen, Diebold, and Rudebusch (2009), and CDGJ(2008) (see Dejong (2000) for a more comprehensive list).

3.3 Data

The dataset used here derives from daily observations of Libor rates of maturities of 1-month,3-month,6-month,9-month, and 12-month and swap rate quotes for 2-year,3-year,4-year,5-year,7-year, and 10-years from June 21, 1996 until July 02, 2008, and were collected from Bloomberg. This resulted in $3,034$ observations, after eliminating special days like Christmas or New Year's when no trading took place. Since Libor rates and swap rate quotes are given on a constant maturity basis, it is possible to construct a panel of zero coupon yields with maturities that matched the underlying instrument on one side of the swap. According to CDGJ(2008), this aspect of the Libor market and the swap market makes it more advantageous to use these quotes over Treasury quoted zero-coupon bond yields. This advantage comes at a cost which is that the swap rate quotes must be used to extrapolate the zero coupon yields. Some descriptive statistics for the extrapolated yields are shown in table 1.1 and 1.2.

The minimum of 0.013836525 for the 10 year rate is due to the fact that the 7-year swap rate quote on that day was 0%. Excluding that minimum, the minimum of all other days in the sample was 0.029772228. For the 7-year rates, the minimum of 0.025563812 was due to the fact that several Libor quotes were around 1%.

3.4 Estimation Risk of Numerical Integration

In this section, I provide details regarding the estimation risk of numerical integration, subsequently estimated within two alternative frameworks. I, first, simulate data from two versions of a Gaussian dynamic term structure model. The first version of the model is restricted in such a way that there exists a closed form solution for the state factor coefficients and so the data generating process obtains from a model that is analytically tractable. The second version of the model is completely unrestricted so that Runge-Kutta methods need to be used to solve the Ricatti ordinary differential equations in order to arrive at the state factor coefficients. In this case, the data generating process obtains from a model that while it is completely unrestricted so that it can provide consistency to more economic scenarios, it has an additional layer of complexity that stems from a lack of analytic tractability.

For each version of the model, I generate a panel of zero coupon yields. I then estimate the Gaussian dynamic term structure model in both the completely unrestricted and hence generalized framework, and in the restricted framework. The estimations are carried out through the use of the inversion method of Duffie and Singleton (1997). I assume that I have two factors and two bonds. In this way, I can avoid estimating additional parameters, like those associated with the measurement error. Issues related to the computation or introduction of measurement error are not directly related to the focus of this study. Below, I provide a brief overview of the Inversion method.

The method involves following Pearson and Sun (1994) and exactly transforming the yields into state variables. This of course requires the use of a Jacobian to carry out the transformation which will be multiplied by the transition density, a multivariate normal. Hence, I am only required to know the first two conditional moments of the state variables, which simplifies the analysis. Symbolically,

$$Y(t,\tau) = \frac{A(\tau)}{\tau} + \frac{(B(\tau))^T}{\tau} X_t$$

More compactly,

$$Y(t) = A + BX(t)$$

where

$$Y(t) = \begin{bmatrix} y_{t,t+\tau_1} \\ \vdots \\ y_{t,t+\tau_N} \end{bmatrix}$$

$$A = \begin{bmatrix} \sum_{i=1}^{F} \frac{A_i \tau_1}{\tau_1} \\ \vdots \\ \sum_{i=1}^{F} \frac{A_i \tau_N}{\tau_N} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{B_{\tau_1}}{\tau_1} \\ \vdots \\ \frac{B_{\tau_N}}{\tau_N} \end{bmatrix} = \begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} \dots \frac{B_N(\tau_1)}{\tau_1} \\ \ddots \\ \frac{B_1(\tau_N)}{\tau_N} \dots \frac{B_N(\tau_N)}{\tau_N} \end{bmatrix}$$

N=the number of bonds

F=the number of factors

and hence,

$X(t) = (B)^{-1}(Y(t) - A)$

For each observation in the sample, the likelihood function is

$L_{1,t} = (\frac{1}{det(\frac{(B(\tau))^T}{\tau})})f(X_{t+1}|X_t)$

So, the estimated parameter vector θ^* solves the following optimization problem

$\max_{\theta^*} L(\theta) = \sum_{i=1}^{T} L_{1,t}(\theta)$

where $\theta = \{\kappa_1, \kappa_2, \kappa_3, \kappa_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \lambda_1, \lambda_2, \theta_1, \theta_2, \beta_1, \beta_2, \beta_3, \beta_4, \alpha_1, \alpha_2, \delta_0, \delta_1, \delta_2\}$

for the generalized framework, and

$\theta = \{\kappa_1, \kappa_2, \sigma_1, \sigma_2, \lambda_1, \lambda_2, \theta_1, \theta_2\}$

for the restricted framework

This results in a parameter space that consists of twenty one parameters, when estimating the generalized Gaussian term structure model, and eight parameters when estimating the restricted Vasicek model. Several authors have noted the difficulties of optimizing over a large parameter space (see, for example, Ang and Piazzesi (2003), Christensen, Diebold, and Rudebusch (2008,2009), Collin-Dufresne, and Jones (2008), Kim and Orphanides (2005), Duffee (2002), Chib and Ergashev (2009), Duffee and Stanton (2007), Duffee (2009), Joslin, Singleton, and Zhu (2010)).

In order to overcome a difficult task of optimizing a highly non-linear parameter space, I follow Ang and Piazzesi (2003), Dai and Singleton (2002), and others who estimate the model in several iterative rounds. I estimate the two factor model in several iterative rounds as well. I repeat this process for a large number of trials. Finally, across the two methods, I compare the statistical bias, statistical efficiency, and economic accuracy for actual yields.

The results from the Monte Carlo simulations are provided in the tables 3.1 and 3.2.

The results from the tables above are intuitive, yet striking at the same time. Table 3.1 presents statistical bias associated with parameters estimated from a generalized Gaussian term structure model, with panel A focusing on a generalized Gaussian term structure model and panel B the restricted Gaussian term structure model. Not surprisingly, the generalized Gaussian model leads to less biased results. Table 3.2 presents efficiency, and accuracy for estimations based upon data generated from a generalized two factor Gaussian term structure model. Results from F-tests for the equality of variances seem to indicate that for both sets of data a normal distribution but with unequal variances. These tests indicated that we reject the null hypothesis that the data were generated from a normal distribution with the same variances.

When the data are generated from an unrestricted Gaussian model, the generalized Gaussian model performs better, on the whole, in its ability to fit the yield curve, and the estimated yields are more efficient than when estimated from a Vasicek model.

The results, as shown in tables 3.3 and 3.4, are similar when the data are generated from the Vasicek model. For estimation via the Vasicek model, we see less efficient parameter estimates with more statistical bias. The generalized Gaussian data generating process also led to more efficiently generated yields than the Vasicek model. Although, when the data are generated from the Vasicek model, the Vasicek model leads to more efficiently generated yields than when the data generating process is a generalized Gaussian process. Just as with the case for data generated from a generalized Gaussian process, results from F-tests for the equality of variances seem to indicate that for both sets of data a normal distribution but with unequal variances. These tests indicated that we reject the null hypothesis that the data were generated from a normal distribution with the same variances.

On the whole, estimation of the generalized Gaussian term structure model seems to lead to less biased parameter estimates and more efficiently generated yields than estimation by the Vasicek model. However, with regards to accuracy for fitting actual yields, the Vasicek model outperformed the generalized Gaussian term structure model when the Vasicek model represented the actual data generating process. In the world of Vasicek, the Vasicek model demonstrated a superior ability to fit the yield curve versus the generalized Gaussian model.

One final observation is in order as regards computational efficiency. The estimation of the Vasicek model took less time that the estimation of the generalized Gaussian model needed to complete each optimization. Usage of the analytical methods always took less time. The average time to completion for the restricted Gaussian process, when the data were generated from a restricted Gaussian process was 277.44 seconds, while it was 382.78 seconds when the data were generated from a generalized Gaussian process. When the data were generated from a generalized Gaussian process, the generalized Gaussian process took about 8,647 seconds, while taking about 7,314 seconds when the data were generated from a restricted Gaussian process.

On the whole, these results speak to the tradeoff involved with the employment of analytical methods, which can only be used in certain cases, to solve the Ricatti equations versus numerical methods which have to be used in the general case. The focus of this tradeoff concerns an observed divide that exists between statistical comparisons (parameter bias and statistical efficiency) which motivate the use of numerical Runge-Kutta methods and economic comparisons (accuracy for fitting actual yields) which seem to motivate the use of analytical methods to solve for the state variable versus numerical methods. The crux of the economic aspect of the tradeoff surrounds the fact that when the data are generated from a restricted Gaussian process, a restricted Gaussian model performs better. For this reason, it seems sensible that researchers continue to use both methods, while being aware of each one's drawbacks.

Table 3.5 presents the results in a slightly different way. Specifically, instead of drawing inference from the parameter space, the results in Table 3.5 presents results in terms of state variables. In specific, the table reports means of the coefficients and the R^2 from the following regression $Estimated\, State\, Variable(t) = \alpha + \beta\, Actual\, State\, Variable(t) + \epsilon(t)$ based upon 4 simulations. The actual state variable obtains from the monte carlo simulation procedure, while the estimated state variable obtains from the parameters arrived at from the maximum likelihood estimation of each simulated dataset for each model. The state variables comprising the independent variables in Panel A, $\{X_{1,Gauss}, X_{2,Gauss}\}$, are generated from a Generalized Gaussian model, while the independent variables in Panel B, $\{X_{1,Vasic}, X_{2,Vasic}\}$, are generated from a Restricted Gaussian model. There were four dependent variables; two were obtained from estimation of the generalized Gaussian process and two from a Vasicek process. Notice that independent of the data generating process, state variables generated from the generalized Gaussian model provided for a more accurate fit to the true state variable as obtained from the Monte Carlo procedure. Similar regressions were run for the Vasicek model in Panel B, although the results were not as strong as those for the generalized Gaussian model. This speaks to the statistical prowess of the generalized Gaussian model.

Results were robust to using a Kolmogorov-Smirnov test to compare underlying empirical densities. Specifically, the results indicate that the state variables generated under the two methods are obtained from the same underlying densities, which seems sensible since both models are governed by Gaussian transition dynamics. For all cases, the p-values were very close 0.

These results, taken together, imply that while the generalized Gaussian term structure model does have the ability to accomodate more realistic scenerios, which model performs better really depends upon the extent to which parametric restrictions impact inference from the model.

Finally, in the Gaussian case, the system of second order ordinary differential equations which produce the state factor coefficients simplifies to linear system of ordinary differential equations for which a closed form solution does indeed exist (Grasselli and Tebaldi, 2007). It may be more interesting to perform a comparison such as the one above in the case of a term structure model for which a closed form solution does not exist for the state factor coefficients, such as a Cox, Ingersoll, and Ross model. This is left for future research.

3.5 Conclusion

Researchers need to use integration to obtain the state factor coefficients which are the result of a system of non-linear ricatti differential equations. However, the analytic solution to the system of differential equations only exists in certain restricted cases. In all remaining cases, the econometrician needs to apply runge-kutta methods in order to numerically integrate the system of differential equations to arrive at the state factor coefficients. Furthermore, the majority of cases which require numerical integration constitute plausible economic scenarios. The desire for analytic tractability forces the researcher to forgo estimation of the model in the cases which lead to economically more plausible scenarios.

The statistical implications of this tradeoff between analytic tractability and economic flexibility was the focus of this study. This paper examined the extent to which the prevalence of estimation risk in numerical integration creates bias, inefficiencies, and inaccurate results in the widely used class of affine term structure models.

This paper highlights an important divide that exists between estimating a term structure model using analytical methods and estimating one using numerical methods. The origins of this divide obtain from statistical implications, for example parameter bias and statistical efficiency, on the one side, and economic implications, such as examining a specific model's fit for the yield curve.

The results of this study indicate that although the generalized model leads to more efficiently generated yields and less biased parameter estimates, if the data is obtained from a restricted model then the restricted Gaussian model generates more accurate yields.

This tradeoff seems to involve the econometrician's ability to weight computational time against statistical bias and efficiency and a model fit.

Overall, this paper provides important and new insight into the estimation of affine term structure models.

Table 3.1. Generalized Gaussian Data Generating Process: Parameter Bias.

Panel A Estimation of Two Factor Generalized Gaussian Model		
Parm	Actual Value	Bias
κ_1	1.4981	-5.02e-11
κ_2	9.8e-007	9.21e-10
κ_3	1.47e-007	-6.09e-10
κ_4	2.9873	-3.35e-05
σ_1	0.07996	-1.02e-06
σ_2	7.45e-007	4.92e-10
σ_3	1.3e-008	-2.28e-10
σ_4	0.09997	4.15e-07
θ_1	0.08999	8.15e-09
θ_2	0.01994	-3.47e-07
λ_1	0.9860	-3.74e-09
λ_2	2.6865	-3.46e-06
β_1	1.59e-008	1.53e-19
β_2	7.16e-009	-5.86e-13
β_3	5.7e-010	-1.95e-13
β_4	2.1e-010	-1.95e-13
α_1	1.0001	5.97e-14
α_2	1.0001	7.42e-14
δ_0	0.00082	-9.80e-10
δ_1	1.00024	-9.28e-09
δ_2	1.00250	-4.88e-08

Panel B Estimation of Two Factor Restricted Gaussian Model		
Parm	Actual Value	Bias
κ_1	0.00000634	3.34e-006
κ_2	9.63427	14.87
σ_1	0.04166	-0.015
σ_2	0.11218	-0.11
θ_1	0.001337	0.238
θ_2	0.296123	-0.086
λ_1	0.682343	-0.041
λ_2	19.47832	-42.39

Table 3.2. Generalized Gaussian Data Generating Process: Efficiency and Accuracy.

Panel A Generalized Gaussian Data Generating Process		
yield	1000*Efficiency	Accuracy
1-month	0.1955	16.815
3-month	0.1076	7.56e-15
6-month	0.0974	49.136
9-month	0.0959	53.769
12-year	0.09545	38.586
2-year	0.09479	27.547
3-year	0.09482	13.448
4-year	0.09462	7.4105
5-year	0.09471	0.9524
6-year	0.09469	1.5440
7-year	0.09473	0.2137
8-year	0.09474	6.5426
9-year	0.09466	0
10-year	0.09468	8.0919

Panel B Restricted Gaussian Data Generating Process		
yield	1000*Efficiency	Accuracy
1-month	0.4133	25.905
3-month	0.4109	5.04e-15
6-month	0.4104	57.565
9-month	0.4103	13.319
12-year	0.4103	14.586
2-year	0.4106	54.318
3-year	0.4109	13.434
4-year	0.4113	11.455
5-year	0.4117	9.2913
6-year	0.4121	7.0305
7-year	0.4124	4.7177
8-year	0.4128	2.3701
9-year	0.4132	5.04e-15
10-year	0.4136	2.3872

Table 3.3. Restricted Gaussian Data Generating Process: Parameter Bias

Panel A Estimation of a Generalized Gaussian Model		
Parm	Actual Value	Bias
κ_1	1.4981	-5.02e-11
κ_2	9.8e-007	9.21e-10
κ_3	1.47e-007	-1.02e-09
κ_4	2.9873	-3.71e-05
σ_1	0.07996	8.75e-06
σ_2	7.45e-007	8.75e-06
σ_3	1.3e-008	-3.64e-09
σ_4	0.09997	6.63e-06
θ_1	0.08999	-4.27e-012
θ_2	0.01994	-2.12e-011
λ_1	0.9860	-5.98e-09
λ_2	2.6865	-5.53e-06
β_1	1.59e-008	7.28e-23
β_2	7.16e-009	-2.86e-16
β_3	5.7e-010	-9.53e-17
β_4	2.1e-010	-9.53e-17
α_1	1.0001	2.22e-16
α_2	1.0001	0
δ_0	0.00082	-2.17e-19
δ_1	1.00024	0
δ_2	1.00250	0

Panel B Estimation of a Restricted Gaussian Model		
Parm	Actual Value	Bias
κ_1	0.00000634	0.0972
κ_2	9.63427	5.6656
σ_1	0.04166	-0.0291
σ_2	0.11218	0.8878
θ_1	0.001337	0.1578
θ_2	0.296123	-0.0612
λ_1	0.682343	3.432
λ_2	19.47832	-17.244

Table 3.4. Restricted Gaussian Data Generating Process: Efficiency and Accuracy

Panel A Generalized Gaussian data generating process		
yield	1000*Efficiency	Accuracy
1-month	0.2323	16.815
3-month	0.0907	7.56e-015
6-month	0.0904	49.1365
9-month	0.0909	53.768
12-year	0.09112	38.59
2-year	0.09286	27.55
3-year	0.09327	13.45
4-year	0.09368	7.4105
5-year	0.09392	0.9524
6-year	0.09389	1.54404
7-year	0.09408	0.21367
8-year	0.09417	6.54262
9-year	0.09423	0
10-year	0.09428	8.0919

Panel A Restricted Gaussian data generating process		
yield	1000*Efficiency	Accuracy
1-month	0.1087	17.455
3-month	0.1089	3.78e-015
6-month	0.1095	41.2997
9-month	0.1098	7.87693
12-year	0.1100	8.835796
2-year	0.1103	21.51279
3-year	0.1104	8.35972
4-year	0.1105	7.382397
5-year	0.1106	6.235624
6-year	0.1106	4.81088
7-year	0.1106	3.33915
8-year	0.1106	1.59187
9-year	0.1107	3.78e-015
10-year	0.1107	1.4249071

TABLE 3.5. Regression results based upon simulated interest rates generated from two Gaussian data generating processes.

	Panel A: Generalized Gaussian Model			
$Dep.Var.$	$\widehat{\alpha}$	$\widehat{\beta}$	R^2	ρ
$X_{1,Guass}$	-0.00001	1.00011	0.9999	0.9999
$X_{2,Guass}$	-0.000002	1.00007	0.9999	0.9999
$X_{1,Vasic}$	-1.65845	1.91469	0.5230	0.7182
$X_{2,Vasic}$	0.99936	0.85111	0.4158	0.5862
	Panel B: Restricted Gaussian Model			
$Dep.Var.$	$\widehat{\alpha}$	$\widehat{\beta}$	R^2	ρ
$X_{1,Guass}$	0.00001	0.99999	0.9999	0.9999
$X_{2,Guass}$	-0.0000004	1.00001	0.9999	0.9999
$X_{1,Vasic}$	-0.000076	0.4611	0.4506	0.6712
$X_{2,Vasic}$	-0.6882	21.15	0.8772	0.9366

Appendix A

Relevant Mathematical Proofs and Computations

A.1 The Equivalent $A_0(3)$ Model

CDR(2008) start from the insight of Diebold and Li (2006) that the coefficients for the original Nelson-Seigel (1987) model fit the yield curve with a specification that has a functional form which can be written in terms of a time-varying version of the familiar Level, Slope, and curvature factors ((L_t, S_t, C_t), respectively). This represents the full dynamic Nelson-Seigel specification

$y_\tau = L_t + S_t(\frac{(1-e^{-\lambda\tau})}{\lambda\tau}) + C_t(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau})$

where y_τ is the zero-coupon yield with τ years to maturity and λ is a model parameter. Specifically, λ can be interpreted as the mean-reversion rate for the curvature and slope factors. If we, further, let $(L_t, S_t, C_t) = X_t$ and follow Duffie and Kan (1996), prices are exponential affine functions of the state variables, with

$P(t,\tau) = e^{(B(t,\tau)^T X_t + Con(t,\tau))}$

where $Con(t,\tau)$ is the convexity adjustment according to CDR(2008).

Hence,

$y(t,\tau) = -\frac{logP(t,\tau)}{\tau} = \frac{-B(t,\tau)}{\tau}X_t - \frac{Con(t,\tau)}{\tau}$

To follow the notation in CDR (2008),

$$B(t,\tau) = \begin{bmatrix} B^1(\tau) \\ B^2(\tau) \\ B^3(\tau) \end{bmatrix} = \begin{bmatrix} \tau \\ \frac{-1-e^{-\lambda\tau}}{\tau} \\ \tau e^{-\lambda\tau} - \frac{(1-e^{-\lambda\tau})}{\lambda} \end{bmatrix}$$

The proof for this relation is in CDR (2008) and is obtained from putting the Nelson-Seigel (1987) framework within the arbitrage-free class of models of Duffie and Kan (1996). It is easily obtained by noting that CDR (2008) operate under the assumption that volatility is a constant.[1] Under this assumption, the Ricatti equations become a simple first order ordinary differential equation which can be solved analytically by imposing the Nelson-Seigel structure and applying the separation of variables technique.

As the final step needed to place the Nelson-Seigel (1987) representation within the affine term structure class models, CDR (2008) impose certain structure on the κ matrix, in the equivalent risk-neutral world, and the ρ_1 vector obtain the system of stochastic differential equations governing level, slope, and curvature as shown below.

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix} [\begin{bmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{bmatrix} - \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}] dt + \Sigma \begin{bmatrix} dW_t^{(1,Q)} \\ dW_t^{(2,Q)} \\ dW_t^{(3,Q)} \end{bmatrix}$$

CDR (2008) estimate the dynamics of the above process in both the physical measure and the equivalent martingale measure. In this paper, I transform the equivalent risk-neutral dynamics of the 3-dimensional Markov process for the state variables associated with the arbitrage-free dynamic Nelson-Seigel (AFDNS) model into a resulting model that is equivalent to the model of CDR (2008) in the sense of Dai and Singleton (2000), while facilitating a comparison of the inversion method of Duffie and

[1] Koopman, Mallee, and Van Der Wel (2008) extend CDR (2008) by generalizing the constant volatility assumption and find improvements in the forecasting ability of this model.

Singleton (1997) to the model-free method of CDGJ (2008). Furthermore, I am able to allow the model to fall within the $A_0(3)$ sub-class of models as well as preserve the "canonical" nature, in the language of Singleton (2006), of the model. The general mode for a transformation can be an affine invariant transformation and/or a Brownian motion rotation. This paper focuses on an empirical design that involves both the rotation of the Brownian motion term and an affine invariant transformation, since by construction an application of both should fully accommodate a comparison between the two methods used to estimate the benchmark model.

Symbolically, an affine invariant transformation is given by

$$\begin{bmatrix} r_t \\ \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} = AX_t + \vartheta$$

In practice, an invariant transformation consists of the following transformations, which include affine invariant transformations and a non-singular matrix, A, such that:

$$\Sigma_X = A\Sigma_Y, \theta_X^Q = A\theta_Y^Q + \eta, \kappa_X^Q = A\kappa_Y^Q A^{-1}$$

An invariant affine transformation does not affect the Brownian motion terms. To affect the Brownian motion terms, define an orthogonal matrix, O, such that $\top_O(W(t)) = OW(t)$

To obtain the model of CDGJ(2008), specify r_t, since $\mu_{1,t}$ will be obtained as its risk-neutral drift term and the same for $\mu_{2,t}$, except that it is the drift term for $\mu_{1,t}$. To place these variables in context with regards to an affine invariant transformation, define $Y = \{r_t, \mu_{1,t}, \mu_{2,t}\}$. Symbolically,

$$\mu_{1,t} = \frac{1}{dt} E_t^Q[dr]$$

$\mu_{2,t} = \frac{1}{dt} E_t^Q [d\mu_{1,t}]$

This is exactly the process specified in CDGJ(2008), except applied to a more general model setting. This process will then be repeated for $\mu_{2,t}$ to get the process for $\mu_{2,t}$. For the standard affine class of term structure models, r_t is given by

$r_t = \delta_o + \delta_{x_1} X_{1,t} + \delta_{x_2} X_{2,t}$

Note that this assumption follows from the fact that the short rate is an affine function of the state variables within this general class of models.

This setup maintains the assumption of CDR(2008) that the short rate be an affine function of two latent pricing factors, while keeping the model general enough so that it is applicable to the widely studied class of affine term structure models. CDR(2008) note that the third factor, the curvature factor, has the sole role of acting as the time-varying mean of the slope factor, which is the second factor, and therefore does not affect the short term interest rate. In the $A_0(3)$ model, which is the benchmark model studied in this paper, the short term interest rate is taken to be a transformation of the level and slope factors. For this reason, these parameters are associated with the transformation of the state variable process.

In order to find the stochastic process for the short rate, this expression needs to be converted into one that includes differentials. After an application Ito's lemma, the end result is shown below:

$dr_t = \delta_{x_1} dX_{1,t} + \delta_{x_2} dX_{2,t}$

Since, these are expressions for $dX_{1,t}$ and $dX_{2,t}$, they can be plugged into the expression above

$dr_t = \delta_{x_1}((\kappa_{11}^Q(\theta_1^Q - X_t^1) + \kappa_{22}^Q(\theta_2^Q - X_t^2) + \kappa_{13}^Q(\theta_3^Q - X_t^3))dt + \Sigma dW_t^{1,Q}) + \delta_{x_2}((\kappa_{21}^Q(\theta_1^Q - X_t^1) + \kappa_{22}^Q(\theta_2^Q - X_t^2) + \kappa_{23}^Q(\theta_3^Q - X_t^3))dt + \Sigma dW_t^{2,Q})$

The Nelson-Seigel (1987) representation affords some restrictions, namely that $\kappa_{11} = \kappa_{12} = \kappa_{13} = \kappa_{21} = \kappa_{31} = \kappa_{32} = 0$, $\kappa_{22} = \lambda$, and $\kappa_{23} = -\lambda$. Thus, the model above reduces to

$$dr_t = \delta_{x_1} \Sigma dW_t^{1,Q} + \delta_{x_2}(\lambda(\theta_2^Q - X_t^2) - \lambda(\theta_3^Q - X_t^3))dt + \Sigma dW_t^{2,Q})$$

The above equation can be rearranged so that it clearly delineates the deterministic term and the drift term:

$$dr_t = \delta_{x_2}(\lambda(\theta_2^Q - X_t^2) - \lambda(\theta_3^Q - X_t^3))dt + \Sigma^* dW_t^{(r,Q)}$$

The star on the Σ term denotes the fact that there is a modified covariance term after the transformation.

Some observations are in order. Stanton (1997) and Jiang (1998) both find that when interest rates are low, which they define to be less than 14%, the short rate process exhibits little reversion and behaves more like a random walk, but when interest rates are larger the short rate exhibits dramatic mean reversion. Conley, Hansen, Luttmer, and Scheinkman (1997) and Jones (2003) report similar results for a different, but similar range of interest rate levels.

Based upon the literature motivated above, the model specification is indeed supported by the data since the time period over which this study is conducted, short term interest rates were below the 14% level.

The process for the drift of the short rate process can be found as follows. We start with the equivalent risk-neutral drift of the process for the short rate.

$$\mu_t^{(1,Q)} = \delta_{x_2}[\lambda(\theta_2^Q - X_t^2) - \lambda(\theta_3^Q - X_t^3)]$$

By virtue of applying Ito's lemma again,

$$d\mu_t^{(1,Q)} = \delta_{x_2}\lambda(dX_t^3 - dX_t^2)$$

Again, following the steps for the short rate, we get

$$\equiv \delta_{x_2}[2\lambda^2(\theta_3^Q - X_t^3) - \lambda^2(\theta_2^Q - X_t^2)dt + \lambda(\Sigma^{**})dW_t^{(\mu_1,Q)}]$$

According to CDGJ (2008) and Singleton (2006), the process for the curvature factor is related to the equivalent risk-neutral drift of the process for the drift of the short rate process:

$$\mu_t^{(2,Q)} = \delta_{x_2}[2\lambda^2(\theta_3^Q - X_t^3) - \lambda^2(\theta_2^Q - X_t^2)]$$

Applying Ito's lemma again and simplifying,

$$d\mu_t^{(2,Q)} = \delta_{x_2}[[\lambda^3(\theta_2^Q - X_t^2) - 3\lambda^3(\theta_3^Q - X_t^3)]dt - \Sigma^* dW_{\mu_t}^{(2,Q)}$$

Finally, I can solve the system as shown below.

$$\mu_t^{(1,Q)} = \delta_{x_2}\lambda\theta_2^Q - \delta_{x_2}\lambda\theta_3^Q - \delta_{x_2}\lambda X_t^2 + \delta_{x_2}\lambda X_t^3$$
$$\mu_t^{(2,Q)} = -\delta_{x_2}\lambda^2\theta_2^Q + 2\delta_{x_2}\lambda^2\theta_3^Q + \delta_{x_2}\lambda^2 X_t^2 - 2\delta_{x_2}\lambda^2 X_t^3$$

Rearranging and placing the system in matrix form,

$$\begin{bmatrix} \mu_t^{(1,Q)} - \delta_{x_2}\lambda\theta_2^Q + \lambda\delta_{x_2}\theta_3^Q \\ \mu_t^{(2,Q)} + \delta_{x_2}\lambda^2\theta_2^Q - 2\delta_{x_2}\lambda^2\theta_3^Q \end{bmatrix} = \begin{bmatrix} -\lambda\delta_{x_2} & \lambda\delta_{x_2} \\ \lambda^2\delta_{x_2} & -2\lambda^2\delta_{x_2} \end{bmatrix} \begin{bmatrix} X_t^2 \\ X_t^3 \end{bmatrix}$$

Since, the coefficient matrix is nonsingular,

$$\begin{bmatrix} X_t^2 \\ X_t^3 \end{bmatrix} = \begin{bmatrix} -\lambda\delta_{x_2} & \lambda\delta_{x_2} \\ \lambda^2\delta_{x_2} & -2\lambda^2\delta_{x_2} \end{bmatrix}^{-1} \begin{bmatrix} \mu_t^{(1,Q)} - \delta_{x_2}\lambda\theta_2^Q + \lambda\delta_{x_2}\theta_3^Q \\ \mu_t^{(2,Q)} + \delta_{x_2}\lambda^2\theta_2^Q - 2\delta_{x_2}\lambda^2\theta_3^Q \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \frac{-2}{\lambda\delta_{x_2}}u_t^{(1,Q)} + \theta_2^Q - \frac{1}{\lambda^2\delta_{x_2}}u_t^{(2,Q)} \\ \frac{-1}{\lambda\delta_{x_2}}u_t^{(1,Q)} + \theta_3^Q - \frac{1}{\lambda^2\delta_{x_2}}u_t^{(2,Q)} \end{bmatrix} = \begin{bmatrix} X_t^2 \\ X_t^3 \end{bmatrix}$$

Plugging this system into the drift term of the process for $du_t^{(2,Q)}$ leads to the desired result.

An alternative presentation, following the appendices of CDR(2008), for obtaining the invariant affine transformation of the κ^Q matrix involves the following transformation:

$$T_A\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} = \begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix}\begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} + \begin{bmatrix} \delta_o \\ \delta_{x_2}(\lambda\theta_2^Q - \lambda\theta_3^Q) \\ \delta_{x_2}(-\lambda^2\theta_2^Q + 2\lambda^2\theta_3^Q) \end{bmatrix}$$

To obtain the rotated κ matrix, recall that $\kappa_X^Q = A * \kappa_Y^Q * A^{-1}$. The steps needed to obtain the new κ matrix are shown below

$$\kappa_X^Q = A\kappa_Y^Q A^{-1}$$

$$\equiv \begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix}\begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix}^{-1}$$

$$\equiv \begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix}\begin{bmatrix} \frac{1}{\delta_{x_1}} & \frac{2}{\delta_{x_1}\lambda} & \frac{1}{\delta_{x_2}\lambda^2} \\ 0 & -\frac{2}{\delta_{x_2}\lambda} & \frac{-1}{\delta_{x_2}\lambda^2} \\ 0 & -\frac{1}{\delta_{x_2}\lambda} & -\frac{1}{\delta_{x_2}\lambda^2} \end{bmatrix}$$

$$\equiv \begin{bmatrix} 0 & \delta_{x_2}\lambda & -\delta_{x_2}\lambda \\ 0 & -\delta_{x_2}\lambda^2 & 2\delta_{x_2}\lambda^2 \\ 0 & \delta_{x_2}\lambda^3 & -3\delta_{x_2}\lambda^3 \end{bmatrix}\begin{bmatrix} \frac{1}{\delta_{x_1}} & \frac{2}{\delta_{x_1}\lambda} & \frac{1}{\delta_{x_2}\lambda^2} \\ 0 & -\frac{2}{\delta_{x_2}\lambda} & \frac{-1}{\delta_{x_2}\lambda^2} \\ 0 & -\frac{1}{\delta_{x_2}\lambda} & -\frac{1}{\delta_{x_2}\lambda^2} \end{bmatrix}$$

So,

$$\kappa_X^Q = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & \lambda^2 & 2\lambda \end{bmatrix}$$

The steps above highlight the application of an affine invariant transformation which is associated with the drift part of the stochastic process system. With regards to the the diffusion process, the application of a rotation of the Sigma matrix associated with the vector of independent Brownian motion terms is necessary (The details of the computation are presented in Appendix B)

Bond yields are taken to be an affine function of state variables that are the result of these transformations. The resulting equivalent maximal model is given by:

$$d\begin{bmatrix} r_t \\ \mu^Q_{1,t} \\ \mu^Q_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & \lambda^2 & 2\lambda \end{bmatrix} \begin{bmatrix} \delta_0 - r_t \\ 0 - \mu^Q_{1,t} \\ 0 - \mu^Q_{2,t} \end{bmatrix} dt + \Sigma^R dB^Q_t$$

Please note that δ_0 is set to 0, the logic of which follows from CDR(2008). This is a result of the restriction that $\kappa^{X,Q}_{1,1} = 0$, so the first factor is a unit-root process under the equivalent risk-neutral measure. This facilitates the identification of this factor as a level factor, and is a direct consequence of the affine invariant transformation of the canonical form of Dai and Singleton (2000). Also note that the pre-rotated form, i.e. $\Sigma \neq \Sigma^R = A\Sigma$, of the covariance matrix for the state variables is consistent with the correlated-factor model of CDR (2008).[2] This consistency results in a lower triangular Σ matrix and holds for two reasons. First, as discussed in CDGJ (2008), the three state variables are not independent since the drift of the first level state variable is the slope variable and this state variable's drift is the curvature state variable. Secondly, the assumption of a lower triangular Σ matrix ensures the maximal level of model flexibility while still maintaining that the model is fully identified (CDR, 2008).

[2]The Σ^R, the rotated version of the Σ matrix, is rotated by the matrix, O, where O is

$$O = \begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

O must be an orthogonal matrix, and so its usage introduces some constraints. This transformation intends to maintain the orthogonality of O, so that the Brownian motion terms remain independent and in that sense the transformation is invariant. According to Dai and Singleton (2000), when performing an affine invariant transformation the matrix used to transform the drift part of the stochastic process just has to be non-singular, and the brownian motion terms are not affected. In contrast, when a Brownian motion rotation is undertaken, there is a necessity to specify an orthogonal matrix that transforms the vector of independent Brownian motions into another vector of independent Brownian motions.

The new system has a leading column of zeros, which is a result of the imposition of restrictions necessary in order to make the affine class of term structure models consistent with the Nelson-Seigel (1987) framework, and so was inherited from CDR(2008). Since the short rate is an affine function of the first two latent state variables, this lack of dependence upon X_t^1 implies that the new system of equations will be independent of r_t.

Alternatively, the above system can be rewritten as follows,

$$d\begin{bmatrix} r_t \\ \mu_{1,t}^Q \\ \mu_{2,t}^Q \end{bmatrix} = \begin{bmatrix} \mu_{1,t}^Q \\ \mu_{2,t}^Q \\ -(\lambda^2\mu_{1,t}^Q + 2\lambda\mu_{2,t}^Q) \end{bmatrix} dt + \Sigma^R dB_t^Q$$

to allow for a direct comparison to CDGJ (2008). The model of CDGJ (2008) is shown below

$$d\begin{bmatrix} r_t \\ \mu_{1,t}^Q \\ \mu_{2,t}^Q \end{bmatrix} = \left[\begin{bmatrix} 0 \\ 0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \mu_{1,t}^Q \\ \mu_{2,t}^Q \\ (\kappa_0 r_t + \kappa_1\mu_{1,t}^Q + \kappa_2\mu_{2,t}^Q) \end{bmatrix}\right] dt + \Sigma dB_t^Q$$

Please note that if $\gamma_0 = 0$, $\kappa_0 = 0$, $\kappa_1 = \lambda^2$, and $\kappa_2 = 2\lambda$, where λ is as defined in CDR (2008), then the model of CDGJ (2008) is exactly the same as the model developed in this research. For this reason, all the conclusions afforded in this paper that hold, are valid within a less general, yet still affine model setting.

A.1.1 Proof That The Arbitrage-free Dynamic Nelson-Seigel Class of Models Falls Within The $A_0(3)$ Class

This appendix follows CDR (2008). Duffie and Kan (1996) prove that zero coupon bond prices are exponential affine functions of the state variables discussed within the text.

$$P(t,T) = E_t^Q[e^{-\int_t^T r_u du}] = e^{B(t,T)^T X_t + C(t,T)}$$

where $B(t,T)$ and $C(t,T)$ are solutions to the following system of ordinary differential equations

$$\frac{dB(t,T)}{dt} = \rho_1 + (\kappa^Q)^T B(t,T) - \frac{1}{2}\sum_{j=1}^{n}(\Sigma^T B(t,T)B(t,T)^T\Sigma)_{j,j}(\delta^j)^T$$

$$B(T,T) = 0$$

$\frac{dC(t,T)}{dt} = \rho_0 - B(t,T)^T(\kappa^Q)^T\theta^Q - \frac{1}{2}\sum_{j=1}^{n}(\Sigma^T B(t,T)B(t,T)^T\Sigma)_{j,j}\gamma_j$

$C(T,T) = 0$

Since CDR (2007) operate under the assumption of a constant volatility matrix, the above system of differential equations simplifies to:

$\frac{dB(t,T)}{dt} = \rho_1 + (\kappa^Q)^T B(t,T)$

$B(T,T) = 0$

$\frac{dC(t,T)}{dt} = \rho_0 - B(t,T)^T(\kappa^Q)^T\theta^Q - \frac{1}{2}\sum_{j=1}^{n}(\Sigma^T B(t,T)B(t,T)^T\Sigma)_{j,j}$

$C(T,T) = 0$

Solving the first differential equation for $B(t,T)$,

$\int_t^T \frac{d}{ds}[e^{((\kappa^Q)^T(T-s))}B(s,T)] = e^{((\kappa^Q)^T)(T-t)}\frac{dB(t,T)}{dt} - ((\kappa^Q)^T)e^{((\kappa^Q)^T)(T-t)}B(t,T)$

$\int_t^T \frac{d}{ds}[e^{((\kappa^Q)^T(T-s))}B(s,T)] = e^{((\kappa^Q)^T)(T-t)}[\rho_1 + (\kappa^Q)^T B(t,T)] - ((\kappa^Q)^T)e^{((\kappa^Q)^T)(T-t)}B(t,T)$

$\int_t^T \frac{d}{ds}[e^{((\kappa^Q)^T(T-s))}B(s,T)] = e^{((\kappa^Q)^T)(T-t)}\rho_1$

This is equivalent to, through the application of the boundary condition (since at the boundary which is the maturity of the zero coupon bond, the value of this state parameter is 0),

$e^{((\kappa^Q)^T(T-t))}B(t,T) = \int_t^T \rho_1 e^{(\kappa)^{Q^T}(T-s)}ds$

$B(t,T) = e^{-((\kappa^Q)^T(T-t))}\int_t^T \rho_1 e^{(\kappa)^{Q^T}(T-s)}ds$

The next step is imposing the Nelson-Seigel (1987) structure on the κ matrix in the equivalent risk-neutral world and the ρ vector. That structure is shown below:

$$(\kappa^Q)^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{bmatrix}$$

$$\rho_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

So,

$e^{(\kappa^Q)^T(T-t)}$

$$\equiv e^{\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{bmatrix}\right)(T-t)}$$

$$\equiv e^{\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix}\right)}$$

Exponentiating the matrix by exploiting the matrix version of the power series expansion for the exponential function,

So,

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix}$$

then

$$A^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix}$$

$$\equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t)\lambda(T-t) & 0 \\ 0 & -2\lambda^2(T-t)^2 & \lambda(T-t)\lambda(T-t) \end{bmatrix}$$

and

$$A^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t)\lambda(T-t) & 0 \\ 0 & -2\lambda^2(T-t)^2 & \lambda(T-t)\lambda(T-t) \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix}$$

$$\equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t)\lambda(T-t)\lambda(T-t) & 0 \\ 0 & -3\lambda^3(T-t)^3 & \lambda(T-t)\lambda(T-t)\lambda(T-t) \end{bmatrix}$$

Since

$$e^A = \sum_{j=1}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$$

We have that the above is

$$\equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda(T-t) & 0 \\ 0 & -\lambda(T-t) & \lambda(T-t) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\lambda(T-t)\lambda(T-t)}{2!} & 0 \\ 0 & \frac{-2\lambda^2(T-t)^2}{2!} & \frac{\lambda(T-t)\lambda(T-t)}{2!} \end{bmatrix} +$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\lambda(T-t)\lambda(T-t)\lambda(T-t)}{3!} & 0 \\ 0 & \frac{-3\lambda^3(T-t)^3}{3!} & \frac{\lambda(T-t)\lambda(T-t)\lambda(T-t)}{3!} \end{bmatrix} + \dots$$

$$\equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1+\frac{\lambda(T-t)}{1!}+\frac{\lambda^2(T-t)^2}{2!}+\dots & 0 \\ 0 & -\lambda(T-t)-\frac{\lambda^2(T-t)^2}{1!}-\frac{\lambda^3(T-t)^3}{2!}+\dots & 1+\frac{\lambda(T-t)}{1!}+\frac{\lambda^2(T-t)^2}{2!}+\dots \end{bmatrix}$$

Noting the series definition of the exponential function,

$$\equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{(\lambda(T-t))} & 0 \\ 0 & -\lambda(T-t)[1+\frac{\lambda(T-t)}{1!}+\frac{\lambda^2}{2!}+\dots] & e^{(\lambda(T-t))} \end{bmatrix}$$

Finally,

$$e^{-((\kappa^Q)^T(T-t))} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{(\lambda(T-t))} & 0 \\ 0 & -\lambda(T-t)e^{(\lambda(T-t))} & e^{(\lambda(T-t))} \end{bmatrix}$$

Inserting this expression and that for ρ_1 into the expression for $B(t,T)$, integrating and simplifying yields the desired result. Specifically,

$B(t,T) = e^{-((\kappa^Q)^T(T-t))} \int_t^T \rho_1 e^{(\kappa)^{Q^T}(T-s)} ds$

Now, plug in the expression above for $e^{-((\kappa^Q)^T(T-t))}$ and simplifying leads to

$$B(t,T) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{(\lambda(T-t))} & 0 \\ 0 & -\lambda(T-t)e^{(\lambda(T-t))} & e^{(\lambda(T-t))} \end{bmatrix} \int_t^T \begin{bmatrix} 1 \\ e^{(\lambda(T-s))} \\ -\lambda(T-s)e^{(\lambda(T-s))} \end{bmatrix} ds$$

Note that,

$\int_t^T 1 ds = (T-t)$

$\int_t^T e^{(\lambda(T-s))} ds = \frac{-1}{\lambda} e^{(\lambda(T-s))}|_t^T = -[\frac{1}{\lambda} - \frac{e^{\lambda(T-t)}}{\lambda}]$

$\int_t^T -\lambda(T-s)e^{(\lambda(T-s))} ds = \frac{1}{\lambda}\int_{-\lambda(T-t)}^0 xe^x dx$

where $x = -\lambda(T-s)$ and hence $ds = \frac{1}{\lambda}dx$

Applying the integration by parts formula with the following definitions; $u = x$, $dv = e^x dx$, $du = 1dx$, $v = e^x$,

$$\int_{-\lambda(T-t)}^{0} udv = \frac{1}{\lambda}[uv|_{-\lambda(T-t)}^{0} - \int_{-\lambda(T-t)}^{0} vdu]$$

$$\equiv \frac{1}{\lambda}[xe^x|_{-\lambda(T-t)}^{0} - \int_{-\lambda(T-t)}^{0} e^x dx]$$

$$\equiv \frac{1}{\lambda}[-\lambda(T-t)e^{-\lambda(T-t)} - 1 - e^{-\lambda(T-t)}]$$

$$\equiv -(T-t)e^{-\lambda(T-t)} - \frac{1-e^{-\lambda(T-t)}}{\lambda}$$

Upon plugging these expressions into the expression for the yield curve, it remains clear that the resulting model falls within the $A_0(3)$ class.

A.1.2 Zero Coupon Bond Yield Extrapolation Procedure

It is the purpose of this section to detail how I obtain zero coupon yields from the swap rate quotes. Using time-series of LIBOR rate quotes, I extracted forward rates at the short-end of the yield curve. To extrapolate the 2-year spot rate, I follow Fama and Bliss (1987). The 3-year spot rate is then extrapolated using the information up to and including that pertaining to the 2-year spot rate, and so on for the rest of the rates. I also assumed that on both sides of the swap, the day count was 30/360. The extrapolation then involved minimizing the difference between the fixed and floating sides of the swap by choosing the forward rate that was to prevail during the year. Symbolically,

$$\min \ \{Value_{fixedside} - Value_{float_side}\}^2 \ x$$

where x is the set of forward rates. It is a singleton set in obtaining the 2-year,3-year,4-year, and 5-year rates, while it contains two forward rate elements for the 5-year to 7-year swap rate quotes and three forward rate elements for the 7-year to

10-year swap rate quotes.

A.1.3 Rotation of Sigma: Orthogonalization Process

Start with the AFDNS model specification of CDR (2008) as shown below in a general setting:

$dY_t = \kappa_Y^Q[\theta_Y^Q - Y_t]dt + \Sigma_Y dW_t^Q$

Consider $T_{Y_t} : AY_t + \eta$ with A-nonsingular

Then the new process is as follows

$dX_t = \kappa_X^Q[\theta_X^Q - X_t]dt + \Sigma_X dW_t^Q$
$(\Sigma_X = A\Sigma_Y, \theta_X^Q = A\theta_Y^Q + \eta, \kappa_X^Q = A\kappa_Y^Q A^{-1})$

In order to carry out the Brownian motion rotation of the above SDE, one must obtain T, where T is a matrix, such that,

$T(A\Sigma_Y) = O$

where O is an orthogonal matrix. Thus, resulting in an orthogonally transformed Brownian motion term.

$OdW_t^Q = \widehat{dW_t^Q}$

In order to implement the Brownian motion rotation, compute $\Sigma_X = A\Sigma_Y$. This is done below:

$$\Sigma_X = \begin{bmatrix} \delta_1 & \delta_2 & 0 \\ 0 & -\delta_2\lambda & \delta_2\lambda \\ 0 & \delta_2\lambda^2 & -2\delta_2\lambda^2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

$$\equiv \begin{bmatrix} \delta_1\sigma_{11}+\delta_2\sigma_{21} & \delta_2\sigma_{22} & 0 \\ -\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31} & -\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32} & \delta_2\lambda\sigma_{33} \\ \delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31} & \delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32} & -2\delta_2\lambda^2\sigma_{33} \end{bmatrix}$$

So, for orthogonality, take the product of the following matrices

$$\begin{bmatrix} \delta_1\sigma_{11}+\delta_2\sigma_{21} & -\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31} & \delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31} \\ \delta_2\sigma_{22} & -\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32} & \delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32} \\ 0 & \delta_2\lambda\sigma_{33} & -2\delta_2\lambda^2\sigma_{33} \end{bmatrix}$$

$$\begin{bmatrix} \delta_1\sigma_{11}+\delta_2\sigma_{21} & \delta_2\sigma_{22} & 0 \\ -\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31} & -\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32} & \delta_2\lambda\sigma_{33} \\ \delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31} & \delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32} & -2\delta_2\lambda^2\sigma_{33} \end{bmatrix}$$

to be the identity matrix. This imposes the following restrictions on the parameter space.

$\Sigma_{X_{11}} : (\delta_1\sigma_{11}+\delta_2\sigma_{21})^2 + (-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})^2 + (\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})^2$

$\Sigma_{X_{12}} : (\delta_1\sigma_{11}+\delta_2\sigma_{21})(\delta_2\sigma_{22}) + (-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})(-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32}) + (\delta_2\lambda^2\sigma_{21} - 2\delta_2\lambda^2\sigma_{31})(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})$

$\Sigma_{X_{13}} : (-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})(\delta_2\lambda\sigma_{33}) + (\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})(-2\delta_2\lambda^2\sigma_{33})$

$\Sigma_{X_{21}} : (\delta_2\sigma_{22})(\delta_1\sigma_{11}+\delta_2\sigma_{21}) + (-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})$
$(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31}) + (\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})(\sigma_{21}\lambda^2\delta_2-2\delta_2\lambda^2\sigma_{31})$

$\Sigma_{X_{22}} : (\delta_2\sigma_{22})^2 + (-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})^2 + (\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})^2$

$\Sigma_{X_{23}} : (-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})(\delta_2\lambda\sigma_{33}) + (\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})(-2\delta_2\lambda^2\sigma_{33})$

$\Sigma_{X_{31}} : (\delta_2\lambda\sigma_{33})(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31}) + (-2\delta_2\lambda^2\sigma_{33})(\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})$

$\Sigma_{X_{32}} : (\delta_2\lambda\sigma_{33})(-\delta_2\lambda\sigma_{22}+\delta_2\lambda^2\sigma_{32}) + (-2\delta_2\lambda^2\sigma_{33})(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})$

$\Sigma_{X_{33}} : (\delta_2\lambda\sigma_{33})^2 + (-2\delta_2\lambda^2\sigma_{33})^2$

Each constraint that corresponds to a diagonal term is equal to 1, while each off-diagonal term is 0. Since there are several constraints which do not seem to have any immediate intuitive interpretation, some further simplifications are necessary to uncover more information about my parameter space.

$\Sigma_{X_{11}}$: $\delta_1^2\sigma_{11}^2+2\delta_1\delta_2\sigma_{11}\sigma_{21}+\delta_2^2\sigma_{21}^2(\lambda^4+\lambda^2+1)+\delta_2^2\lambda^2\sigma_{31}^2+4\delta_2^2\lambda^4\sigma_{31}^2-2\delta_2^2\sigma_{21}\sigma_{31}(\lambda^2+2\lambda^4)$

$\Sigma_{X_{12}}$: $\delta_2^2\sigma_{21}\sigma_{22}(\lambda^4+\lambda^2+1)+\delta_1\delta_2\sigma_{11}\sigma_{22}-(\delta_2^2\sigma_{21}\sigma_{32}+\delta_2^2\sigma_{31}\sigma_{22})(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{13}}$: $-\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{33}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{21}}$: $\delta_2^2\sigma_{21}\sigma_{22}(\lambda^4+\lambda^2+1)+\delta_1\delta_2\sigma_{11}\sigma_{22}-(\delta_2^2\sigma_{21}\sigma_{32}+\delta_2^2\sigma_{31}\sigma_{22})(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{22}}$: $\delta_2^2\sigma_{22}^2(\lambda^4+\lambda^2+1)-2\delta_2^2\sigma_{22}\sigma_{32}(\lambda^4+\lambda^2)+\delta_2^2\sigma_{32}^2(\lambda^2+4\lambda^4)$

$\Sigma_{X_{23}}$: $-\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{33}\sigma_{32}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{31}}$: $-\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{33}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{32}}$: $-\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{32}\sigma_{33}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{33}}$: $\sigma_{33}^2\delta_2^2\lambda^2(1+4\lambda^2)$

The simplifications have yielded a valuable insight; three of the constraints occur twice and hence the total number of constraints is 6, excluding the three that require the equality of the off-diagonal terms. Specifically,

$\Sigma_{X_{11}}$: $\delta_1^2\sigma_{11}^2+2\delta_1\delta_2\sigma_{11}\sigma_{21}+\delta_2^2\sigma_{21}^2(\lambda^4+\lambda^2+1)+\delta_2^2\lambda^2\sigma_{31}^2+4\delta_2^2\lambda^4\sigma_{31}^2-2\delta_2^2\sigma_{21}\sigma_{31}(\lambda^2+2\lambda^4)$

$\Sigma_{X_{12}}$: $\delta_2^2\sigma_{21}\sigma_{22}(\lambda^4+\lambda^2+1)+\delta_1\delta_2\sigma_{11}\sigma_{22}-(\delta_2^2\sigma_{21}\sigma_{32}+\delta_2^2\sigma_{31}\sigma_{22})(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{13}}$: $-\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{33}(\lambda^2+4\lambda^4)$

$\Sigma_{X_{22}}$: $\delta_2^2\sigma_{22}^2(\lambda^4+\lambda^2+1)-2\delta_2^2\sigma_{22}\sigma_{32}(\lambda^4+\lambda^2)+\delta_2^2\sigma_{32}^2(\lambda^2+4\lambda^4)$

$\Sigma_{X_{23}}: \; -\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{33}\sigma_{32}(\lambda^2 + 4\lambda^4)$

$\Sigma_{X_{33}}: \; \sigma_{33}^2\delta_2^2(\lambda^2 + 4\lambda^4)$

It is possible to further simplify the above equations, through a characterization of those parameters that are considered to be free and those that are not. This characterization affords an elimination of three of the constraints and the combination of two constraints into one, so that in the end, the parameter space consists of 10 parameters to go along with two constraints. The details of this results are outlined as follows. Start with the constraint for $\Sigma_{X_{33}}$ above and isolate the λ related variables. Hence,

$(\lambda^2 + 4\lambda^4) = \frac{1}{\sigma_{33}^2\delta_2^2}$

By plugging the constraint equation for $\Sigma_{X_{33}}$, in its rearranged form, into the one for $\Sigma_{X_{13}}$, results in a second equation characterizing the λ related variables,

$(\lambda^2 + 2\lambda^4) = \frac{\sigma_{31}}{\sigma_{33}^2\delta_2^2\sigma_{21}}$

The fact that λ solves this system of equations implies that it is completely determined by the parameter space, i.e. it is a function of the choice parameters, and hence does not affect the optimization process directly. Moreover, the possibility of multiplicity of the roots is eliminated by the imposition that the solution must satisfy $\lambda^2 = \sqrt{\lambda^4}$.

Thus,

$\lambda^2 = \frac{2\sigma_{31} - \sigma_{21}}{\sigma_{33}^2\delta_2^2\sigma_{21}}$

$\lambda^4 = \frac{\sigma_{21} - \sigma_{31}}{2(\sigma_{33}^2\delta_2^2\sigma_{21})}$

Plugging these two expressions into the one for $\Sigma_{X_{22}}$ yields an expression for δ_2^2 as follows:

From the constraint expression for $\Sigma_{X_{22}}$,

$$(\sigma_{22}^2 - 2\sigma_{22}\sigma_{32}) * (\frac{\sigma_{21}-\sigma_{31}}{2\sigma_{33}^2\sigma_{21}}) + (\sigma_{22}^2 - 2\sigma_{22}\sigma_{32}) * (\frac{2\sigma_{31}-\sigma_{21}}{\sigma_{33}^2\sigma_{21}}) + \frac{\sigma_{32}^2}{\sigma_{33}^2} + \delta_2^2\sigma_{22}^2$$

Solving for δ_2^2,

$$\delta_2^2 = (\frac{1}{\sigma_{22}^2}) * (\frac{-\sigma_{21}\sigma_{22}^2 - \sigma_{31}\sigma_{22}^2 - 2\sigma_{31}\sigma_{32}\sigma_{22} + 4\sigma_{22}^2\sigma_{31} - 2\sigma_{21}\sigma_{22}^2 - 2\sigma_{21}\sigma_{32}^2}{2*\sigma_{33}^2\sigma_{21}})$$

Thus, it can be seen that δ_2^2 and λ are dependent variables. Furthermore, the constraints specified by $\Sigma_{X_{13}}$, $\Sigma_{X_{22}}$, and $\Sigma_{X_{33}}$ actually define the δ_2^2, λ, and λ^2 variables. The two constraints that remain are

$$C_1 : \sigma_{31}\sigma_{22} = \sigma_{32}\sigma_{21}$$

from the combination of the constraint for $\Sigma_{X_{13}}$, $\Sigma_{X_{23}}$, and $\Sigma_{X_{33}}$

and

$$C_2 : \frac{\sigma_{31}\sigma_{32} - \delta_1\delta_2\sigma_{11}\sigma_{22}\sigma_{33}^2}{\sigma_{33}^2(\delta_2^2\sigma_{21}\sigma_{31})} = 1 - \delta_1\sigma_{11}^2 - 2\delta_1\delta_2\sigma_{11}\sigma_{21} + \frac{\sigma_{31}^2}{\sigma_{33}^2}$$

from the combination of $\Sigma_{X_{11}}$ and $\Sigma_{X_{12}}$. It is possible to solve the above equation for δ_1, after rearranging the equation. The expression for δ_1 that results is:

$$\delta_1 = \frac{(1+\frac{\sigma_{31}^2}{\sigma_{33}^2}) - \frac{\sigma_{32}}{\sigma_{33}^2(\delta_2^2\sigma_{21})}}{\sigma_{11}^2 - 2\delta_2\sigma_{11}\sigma_{21} - \frac{\sigma_{11}\sigma_{22}}{\sigma_{21}\sigma_{31}}}$$

The solution to all 6 constraints imply expressions for λ, δ_1, and δ_2 and hence these parameters are not associated with the optimization process. δ_1 and δ_2 are parameters associated with the Brownian motion rotation of the diffusion process, and not with the short rate process. Moreover, the interpretation of λ obtains from the AFDNS model.

A.1.4 Parameter Restrictions and The Benchmark Model Setting

In the original specification of the AFDNS class of interest rate models with correlated factors, the P-dynamics are given by

$$d\begin{bmatrix} X_t^{(1,P)} \\ X_t^{(2,P)} \\ X_t^{(P,3)} \end{bmatrix} = \begin{bmatrix} \kappa_{1,1}^{X,P} & \kappa_{1,2}^{X,P} & \kappa_{1,3}^{X,P} \\ \kappa_{2,1}^{X,P} & \kappa_{2,2}^{X,P} & \kappa_{2,3}^{X,P} \\ \kappa_{3,1}^{X,P} & \kappa_{3,2}^{X,P} & \kappa_{3,3}^{X,P} \end{bmatrix} \begin{bmatrix} \theta_1^{(X,P)} - X_t^{(1,P)} \\ \theta_2^{(X,P)} - X_t^{(2,P)} \\ \theta_3^{(X,P)} - X_t^{(3,P)} \end{bmatrix} dt + \begin{bmatrix} \sigma_{1,1}^X & 0 & 0 \\ \sigma_{2,1}^X & \sigma_{2,2}^X & 0 \\ \sigma_{3,1}^X & \sigma_{3,2}^X & \sigma_{3,3}^X \end{bmatrix} dB_t^Q$$

and after certain restrictions which are imposed to obtain the Nelson-Siegel structure (see Appendix B for more details) and to arrive at the AFDNS model from the canonical $A_0(3)$ form dicussed in Dai and Singleton (2000) (see CDR (2008)), the Q-dynamics are given by

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix} [\begin{bmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{bmatrix} - \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}] dt + \begin{bmatrix} \sigma_{1,1}^X & 0 & 0 \\ \sigma_{2,1}^X & \sigma_{2,2}^X & 0 \\ \sigma_{3,1}^X & \sigma_{3,2}^X & \sigma_{3,3}^X \end{bmatrix} \begin{bmatrix} dW_t^{(1,Q)} \\ dW_t^{(2,Q)} \\ dW_t^{(3,Q)} \end{bmatrix}$$

CDR(2008) impose further restrictions on theta in the Q-measure. Specifically, that they are all zero, while free in the P-dynamics. The imposition of these restrictions on the model mean in the Q-measure only has relevance for forecasting out of sample, which is not a central goal of this paper. (see Joslin, Singleton, and Zhu(2009)).

$$\begin{bmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{bmatrix} = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}]dt + \begin{bmatrix} \sigma_{1,1}^X & 0 & 0 \\ \sigma_{2,1}^X & \sigma_{2,2}^X & 0 \\ \sigma_{3,1}^X & \sigma_{3,2}^X & \sigma_{3,3}^X \end{bmatrix} \begin{bmatrix} dW_t^{(1,Q)} \\ dW_t^{(2,Q)} \\ dW_t^{(3,Q)} \end{bmatrix}$$

This leads to 7 parameters to be measured through the application of the maximum likelihood estimation of the parameters of the equivalent risk-neutral dynamics. In the previous section, I noted that the rotation of the model allows λ to be written as a function of the 6 covariance terms for the diffusion process of the state variables. Hence, save for the measurement error matrix, the brownian motion rotations reduced the parameter space from 7 parameters to 6 parameters to be estimated in the risk-neutral dynamics.

A.1.5 Analytic Computation of The First and Second Moments of Observable $A_0(3)$ System

This appendix broadly follows the approach in Fisher and Gilles (1996). For the sake of completeness, the system that I intend to estimate is shown below.

$$d\begin{bmatrix} r_t \\ \mu_t^{(1,Q)} \\ \mu_t^{(2,Q)} \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & \lambda^2 & 2\lambda \end{bmatrix} \begin{bmatrix} 0 - r_t \\ 0 - \mu_t^{(1,Q)} \\ 0 - \mu_t^{(2,Q)} \end{bmatrix} dt + \Sigma^* dB_t^Q$$

In order to preserve some degree of generality, make the following definitions:

$$X_t = \begin{bmatrix} r_t \\ \mu_t^{(1,Q)} \\ \mu_t^{(2,Q)} \end{bmatrix}$$

$$\kappa = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & \lambda^2 & 2\lambda \end{bmatrix}$$

and

$$\theta = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Re-writing the system of observable variables,

$$dX_t = \kappa(\theta - X_t)dt + \Sigma dW_t$$

If we assume $s > t$, then

$$X(s) = X(t) - \int_t^s dX(v)$$

Taking the expectation of both sides at time t and conditioning on $X(t)$, yields,

$$E_t[X(s)|X(t)] = X(t) + \int_t^s E_t[dX(v)]$$

Since the Brownian motion process is a martingale, the integral given above reduces to,

$$E_t[X(s)|X(t)] = X(t) + \int_t^s E_t[d\mu_{x_t}(v)]$$
$$\Rightarrow E_t[X(s)|X(t)] = X(t) + \int_t^s [\kappa\theta - \kappa E_t[X(v)|X(t)]dv$$

Taking the derivative with respect to s, in order to get the above integral equation into the form of a differential equation yields,

$$\frac{\partial}{\partial s}[E_t[X(s)|X(t)] = [\kappa\theta - \kappa[E_t[X(s)|X(t)]]$$

In order to solve this differential equation, choose an integrating factor, $\mu(s)$, multiplying this factor through the above equation,

$$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] = \mu(s)\kappa\theta - \mu(s)\kappa[E_t[X(s)|X(t)]]$$

or

$$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] = \mu(s)\kappa\theta + \mu(s)(-\kappa)[E_t[X(s)|X(t)]]$$

Rearranging, we get

$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] - \mu(s)(-\kappa)[E_t[X(s)|X(t)]] = \mu(s)\kappa\theta$

Note the following relation,

$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] - \mu(s)(-\kappa)[E_t[X(s)|X(t)]] = \frac{d}{ds}(\mu(s))([E_t[X(s)|X(t)]])$

Hence,

$\frac{d}{ds}(\mu(s))([E_t[X(s)|X(t)]]) = \mu(s)\kappa\theta$

Integrating both sides,

$\int_t^T \frac{d}{ds}(\mu(s))([E_t[X(s)|X(t)]]) = \int_t^T \mu(s)\kappa\theta$

Allowing $X(t,s) = [E_t[X(s)|X(t)]]$ and integrating,

$\mu(T)X(t,T) - \mu(t)X(t,t) = \kappa\theta \int_t^T \mu(s)ds$

Isolating $X(t,T)$ yields,

$X(t,T) = \mu^{-1}(T)\mu(t)x(t,t) + \mu^{-1}(T)\int_t^T \mu(s)\kappa\theta ds$

Identifying the exact functional form of the integrating factor,

$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] - \mu(s)(-\kappa)[E_t[X(s)|X(t)]] = \frac{d}{ds}(\mu(s))([E_t[X(s)|X(t)]])$

$\mu(s)\frac{\partial}{\partial s}[E_t[X(s)|X(t)] - \mu(s)(-\kappa)[E_t[X(s)|X(t)]] = \mu(s)\frac{dX(t,s)}{ds} + \frac{d\mu}{ds}X(t,s)$

or

$-\mu(s)(-\kappa)[E_t[X(s)|X(t)]] = \frac{d\mu}{ds}X(t,s)$

Rearranging,

$-(-\kappa)ds = \frac{d\mu}{\mu}$

Integrating both sides from 0 to x,

$\int_0^x -(-\kappa)ds = \int_0^x \frac{d\mu}{\mu}$

Finally,

$\mu(x) = e^{-(-\kappa)x}$

So, let $\kappa' = -\kappa$

$X(t,T) = e^{\kappa'(T-t)}X(t) + e^{\kappa' T}\int_0^{T-t}\kappa\theta e^{-\kappa' s}ds$

$X(t,T) = e^{\kappa'(T-t)}X(t) + \int_t^{T}\kappa\theta e^{-\kappa'(T-u)}du$

$X(t,T) = e^{\kappa'(\Delta t)}X(t) + \int_0^{\Delta t}\kappa\theta e^{-\kappa'(T-u)}du$

Where Δt is the time between observations

For the variance of the state variable, note that

$X(t,T) = e^{\kappa'(\Delta t)}X(t) + \int_0^{\Delta t}\kappa\theta e^{-\kappa'(T-u)}du$

Differentiating both sides of the above expression,

$dX(t,T) = d(e^{\kappa'(T-t)}X(t)) + d(\int_0^{T-t}\kappa\theta e^{-\kappa'(T-u)}du)$

Applying Ito's lemma on the first term and differentiating the second term yields,

$dX(t,T) = e^{\kappa'(T-t)}dX(t) + d(e^{\kappa'(T-t)})X(t) - \kappa\theta e^{-\kappa'(T-t)}dt$

Plug in the expression derived above for $dX(t)$, and further differentiate the second term, to obtain

$dX(t,T) = e^{\kappa'(T-t)}(\kappa(\theta - X_t)dt + \Sigma dW_t) - e^{\kappa'(T-t)}\kappa X(t)dt - \kappa\theta e^{-\kappa'(T-t)}dt$

Hence,

$dX(t,T) = e^{\kappa'(T-t)}\Sigma dW_t$

Further note that,

$X(T) = X(t,T) + \int_t^{T} dX(s,T)$

$X(T) = X(t,T) + \int_t^{T} e^{\kappa'(T-s)}\Sigma dW_s$

$V[X(T)] = \sigma^2[\int_t^T e^{\kappa'(T-s)}\Sigma dW_s]$

By the Ito isometry and the fact that the Brownian motion process is a martingale, the above is equivalent to,

$V[X(T)] = \int_t^T E_t[(e^{\kappa'(T-s)}\Sigma)(e^{\kappa'(T-s)}\Sigma)^T]ds$

$V[X(T)] = \int_t^T e^{\kappa'(T-s)}\Sigma\Sigma^T(e^{\kappa'(T-s)})^T ds$

To implement the above variance computation, apply a Simpson-method adapted quadrature after carrying out some simple algebraic manipulations.

A.1.6 Description of The Model-free Method For Estimating State Variables

This section reviews the method of CDGJ(2008) use to compute state variables. Therefore, nothing new, over and above CDGJ(2008) is being contributed in this section of the appendix.

The method associated with CDGJ(2008) is contingent upon the fact that three factors drive most of the explained variation the yield curve according to Litterman and Schienkman (1991). This approach imposes some parametric structure while not being dependent upon any model. It uses principal components analysis to form derivatives of the yield curve and then a Taylor Series Expansion about a maturity of zero to tie these to the back to the level of the yields at the short end of the curve. A simple but messy application of Ito's lemma ties these derivatives to state variables.

Specifically, it relies upon the use of principal components analysis to extract factor loadings from yields with the following maturities; 1-month, 3-month, 6-month, 12-month, 2-year, 3-year, 5-year, and 7-year. Then, apply orthogonality of the resulting factor loading matrix to obtain the principal component factors. The next step is to carry out the following regressions; regress the first factor on a vector of ones and a vector τ which is a vector that contains the times to maturity at very short horizons, regress the second factor on the set above plus a term quadratic in the time to maturity vec, and finally regress the third factor on the set used in the second regression plus a term that is cubic in the time to maturity vector. The resulting coefficients from the regressions allow me to construct n^{th} order derivatives, through the following approximation which is a result of the work of Litterman and Schienkman (1991) and others:

$y(t,\tau) \cong \sum_{k=1}^{3} f_k(\tau)P_k(t)$

where

$P_k(t) = k^{th} principalcomponentfactor$

$f_k(t) = factorloadingfunction \; (i.e.polynomialexpansionforthek^{th}factor)$

Taking the n^{th} order derivative with respect to τ of both sides,

$\frac{\partial Y^n(t,\tau)}{\partial \tau}|_{\tau=0} \cong \sum_{k=1}^{3} \frac{\partial^n \; f_k(\tau)}{\partial \tau^n} P_k(t)$

A Taylor Series expansion of the yield curve about short maturities connects the above derivatives to state variables. Specifically, the application of Ito's lemma on the Markov state vector $\{X(t)\}$, with length N (essentially N is the number of state variables), shown below, in the equivalent risk-neutral measure.

$dX_i = m_i^Q(X)dt + \sum_{k=1}^{N} \sigma_{ik}(X)dz_i^Q$

Further, assume that the spot rate is an arbitrary function of the state variables and apply Ito's formula to the Markov State vector above,

$dr = \sum_{i=1}^{N} \frac{\partial r}{\partial X_i}[m_i^Q dt + \sum_{i=1}^{N} \sigma_{ik} dz_i^Q] + \frac{1}{2}\sum_{i,j,k=1}^{N} \frac{\partial^2 r}{\partial X_i \partial X_j}\sigma_{ik}\sigma_{jk}dt$

To derive the equivalent risk-neutral drift, its variance and the drift of the drift, start by defining the price of a zero coupon bond in terms of state variables with maturity T as follows:

$P^T(t, X_t) \equiv e^{-\tau Y(X_t,\tau)}$

Next, take first order derivatives with respect to time to maturity, state variables, and second order cross-partial derivatives with respect to pairs of state variables and connect them to the bond pricing equation. Proceeding step by step, the computations of the partial derivatives are shown below.

$P_\tau = \frac{\partial P}{\partial \tau} = \frac{\partial(e^{-\tau Y(\{X_t\},\tau)})}{\partial \tau} = [-\tau \frac{\partial Y(\{X_t\},\tau)}{\partial \tau} - Y(\{X_t\},\tau)]e^{-\tau Y(\{X_t\},\tau)} = [-\tau Y_\tau - Y]P$

$P_i = \frac{\partial P}{\partial X_i} = \frac{\partial(e^{-\tau Y(\{X_t\},\tau)})}{\partial X_i} = -\tau \frac{\partial(Y(\{X_t\},\tau))}{\partial X_i} e^{-\tau Y(\{X_t\},\tau)} = -\tau Y_i P$

$P_{ij} = \frac{\partial P_i}{\partial X_j} = \frac{\partial(-\tau Y_i P)}{\partial X_j} = -\tau[\frac{\partial P}{\partial X_j} Y_i + \frac{\partial Y_i}{\partial X_j} P] = \tau[Y_i \tau Y_j P] - \tau Y_{ij} P$

Plugging the above partial derivatives into the bond pricing equation,

$rP = -P_\tau + \sum_{i=1}^N P_i m_i^Q + \frac{1}{2} \sum_{i,j,k}^N P_{ij} \sigma_{ik} \sigma jk$

$\rightarrow rP = [-\tau Y_\tau - Y]P + \sum_{i=1}^N -\tau Y_i P m_i^Q + \frac{1}{2} \sum_{i,j,k}^N [\tau[Y_i \tau Y_j P] - \tau Y_{ij} P] \sigma_{ik} \sigma jk$

Note that this equation is consistent with the absence of arbitrage. Plugging the derivatives as described above into the bond pricing equation and then using a Taylor Series expansion as such

$Y(X_t, \tau) \equiv Y^0(X_t) + \tau Y^1(X_t) + \frac{1}{2}\tau^2 Y^2(X_t) + \ldots$

to write yields and plugging that into the bond pricing equation with the derivatives leads to the desired result. This involves an additional step to compute another set of partial derivatives of the yields with respect to the state variables

$Y_\tau = Y^1(\{X_t\}) + \tau Y^2(\{X_t\}) + \ldots$

$Y_i = Y_i^0(\{X_t\}) + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \ldots$

$Y_j = Y_j^0(\{X_t\}) + \tau Y_j^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_j^2(\{X_t\}) + \ldots$

$Y_{ij} = Y_{ij}^0(\{X_t\}) + \tau Y_{ij}^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_{ij}^2(\{X_t\}) + \ldots$

Then, plugging these derivative expressions into the above bond pricing equation, collecting terms and simplifying leads to

$\rightarrow r = ([Y^0(\{X_t\}) + \tau Y^1(\{X\}) + \frac{1}{2}\tau^2 Y^2(\{X_t\})] + \tau[Y^1(\{X_t\}) + \tau Y^2(\{X_t\}) + \ldots]) - \tau \sum_{i=1}^N (Y_i^0(\{X_t\}) + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \ldots)m_i^Q + \frac{1}{2}\sum_{i,j,k}^N \tau^2([(Y_i^0 + \tau Y_i^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_i^2(\{X_t\}) + \ldots)(Y_j^0(\{X_t\}) + \tau Y_j^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_j^2(\{X_t\}) + \ldots)] - \tau[Y_{ij}^0(\{X_t\}) + \tau Y_{ij}^1(\{X_t\}) + \frac{1}{2}\tau^2 Y_{ij}^2(\{X_t\})])\sigma_{ij}\sigma_{jk}$

Finally, collecting terms in τ,

$r = Y^0(\{X_t\}) + \tau[Y^1(\{X_t\}) + Y^1(\{X_t\}) - \sum_{i=1}^N Y_i^0(\{X_t\})m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^N Y_{ij}^0(\{X_t\})\sigma_{ij}\sigma_{jk}] + \tau^2[\frac{1}{2}Y^2(\{X_t\}) + Y^2(\{X_t\}) - \sum_{i=1}^N Y_i^1(\{X_t\})m_i^Q + \frac{1}{2}Y_i^0(\{X_t\})Y_j^0(\{X_t\})\sigma_{ij}\sigma_{jk} - \frac{1}{2}\sum_{i,j,k=1}^N Y_{ij}^1(\{X_t\})\sigma_{ij}\sigma_{jk}]$

Because r is the level of the short rate (i.e. the first term in a Taylor series is the function evaluated at a, where here a is approximately zero, so r is the level of interest rate at short maturities, or the level of the short rate)

Hence, the first three state variables are

$Y^0(t) = r(X_t)$

$Y^1(X_t) = \frac{1}{2}\mu_1(X_t)$

$Y^2(X_t) = \frac{1}{3}[\mu_2(X_t) - V(t)]$

where $\mu_1(t) = \sum_{i=1}^N Y_i^0(\{X_t\})m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^N Y_{ij}^0(\{X_t\})\sigma_{ij}\sigma jk$

$Y_i^0(\{X_t\}) = \frac{\partial r}{X_i}$

$Y_{ij}^0(\{X_t\}) = \frac{\partial^2 r}{X_i X_j}$

$V(t) = \sum_{i,j,k=1}^N Y_i^0(\{X_t\})Y_j^0(\{X_t\})\sigma_{ij}\sigma_{jk}$

$\mu_2 = \sum_{i=1}^{N} Y_i^1(\{X_t\})m_i^Q - \frac{1}{2}\sum_{i,j,k=1}^{N} Y_{ij}^1(\{X_t\})\sigma_{ij}\sigma_{jk}$

A.1.7 Monte Carlo Simulation and The Robustness of The Model-Free State Variable Estimation Procedure

This section provides motivation and implementation details regarding the Monte Carlo simulation exercise used to assess the accuracy and level of unbiasedness of parameter estimates and state variable realizations.

The method used to determine the state variable time-series realizations has the advantage that it uses principal components analysis to impose some parametric structure on the data, yet it is still independent of any model (since usually a model, like a log likelihood model can be obtained from the assumed distribution of the data) (CDGJ, 2008). Since the principal component factors have strong ties to the variation in the interest rate structure, using principal component realizations does impose some structure on the estimation problem, this imposition of structure dramatically reduces the estimation error. The approach to carrying out the simulations more or less follows that of CDGJ(2008). The implementation details are discussed below.

The first step in the implementation is to simulate the term structure of interest rates, with exact maturities and length to that of the actual data set, using the preferred essentially affine $A_0(3)$ model of Duffee (2002). For convenience of presentation it is shown below:

$$r_t = \delta_0 + \delta_1 X_{t,1} + \delta_2 X_{t,2} + \delta_3 X_{t,3}$$

$$d\begin{bmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{bmatrix} = \left[\begin{bmatrix} (\kappa\theta)_1 \\ (\kappa\theta)_2 \\ (\kappa\theta)_3 \end{bmatrix} - \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}\begin{bmatrix} X_{t,1} \\ X_{t,2} \\ X_{t,3} \end{bmatrix}\right] dt + S_t dW_t$$

$$S_{t(ii)} = \sqrt{\alpha_i + (\beta_{i1} + \beta_{i2} + \beta_{i3})X_t}$$

$$\Lambda_t = S_t \begin{bmatrix} \lambda_{11} \\ \lambda_{12} \\ \lambda_{13} \end{bmatrix} + S_t^- \begin{bmatrix} \lambda_{2(11)} & \lambda_{2(12)} & \kappa_{2(13)} \\ \lambda_{2(21)} & \lambda_{2(22)} & \lambda_{2(23)} \\ \lambda_{2(31)} & \lambda_{2(32)} & \lambda_{2(33)} \end{bmatrix} X_t$$

S_t^- is defined in Duffee (2002) and for the case of the $A_0(3)$ model, which contains no square root processes, the β vector will be a zero vector and the α vector will be all ones, so that the S_t matrix is a diagonal matrix. This assumption is consistent with the system of SDEs governing the risk-neutral dynamics of the state variables in the original model of CDR (2008). It is also a characteristic of the Vasicek (1977) model, and, hence will lead to some degree of similarity between the model obtained in this paper and that of Vasicek (1977). The rest of the parameters used in the simulation will be obtained from Table III in Duffee (2002). It is the case that since the κ matrix is not diagonal based on the results in Duffee (2002), the Ricatti equations that correspond to $A_0(3)$ do not have an analytic solution and therefore the state variables cannot be solved for explicitly. A closed form solution can only be found if each state variable can be written as independent one-dimensional equations. Should the κ matrix, as in this case, have any off-diagonal terms this would not be the case. Thus, the state variable realizations must be solved for numerically. Runge-Kutta methods are applied for the purpose. Upon obtaining each simulated data set, every aspect of the principal component analysis, polynomial order determination, and state variable determination is repeated $N = 1000$ times.

Finally, after each simulation in order to assess the bias associated with the model-independent estimates, actual state variable realizations are regressed on estimated state variable realizations and record the parameter estimates and the R^2.

More specifically, this table reports the accuracy and bias in state variable estimates produced by the model-free method of CDGJ (2008). The procedure replicates the construction of the model-free state variables. In addition to a 0 error case, I follow CDGJ(2008) and consider two levels of measurement errors (standard deviations of 0.5 basis points and 2 basis points). The table reports the means and standard deviations of the coefficients and the R^2 from the following regression $Actual\ State\ Variable(t) = \alpha + \beta\ Estimated\ State\ Variable(t) + \epsilon(t)$. The basic summary statistics, across all simulations, of these estimates are shown in the table 1.5.

If this procedure is robust, it should result in high R-squared values, $\widehat{\alpha}$ values close to 0, and β values that are close to 1. For the state variables r and μ_1, the procedure is clearly robust, however, there is a some bias in the third state variable μ_2, since the values for $\widehat{\alpha}$ are under 10^{-4} in both the low error case and the high error case. These results are generally consistent with CDGJ(2008), who also find some bias in the third state variable.

References

[1] Andrews, D.W.K. and Monahan, J.C. (1992). An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator. *Econometrica*, 60, 953-966.

[2] Ang, A., Dong S., & Monika Piazzesi (2007). No-Arbitrage Taylor Rules. NBER Working Papers 13448, National Bureau of Economic Research, Inc. Revised and Resubmitted *American Economic Review*.

[3] Ang, A. & Piazzesi, M. (2003). A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50, pp. 745–787.

[4] Babbs, S. and K. Ben Nowman (1999). Kalman Filtering of Generalized Vasicek Term Structure Models. *Journal of Financial and Quantitative Analysis*, 34, 115-130.

[5] Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1), 135-172.

[6] Bai, J. and Ng, S. (2006). Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions. *Econometrica*, 74(4), 1133-1150.

[7] Bai, J. and Ng, S. (2010). Principal Components Estimation and Identification of the Factors. Working Paper.

[8] Bayazit, M. and Onoz, B (2009). To pre-whiten or not to prewhiten in trend analysis. *Hydrological Sciences Journal*, 52(4), 611-624.

[9] Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman (1974). Estimation inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 4, 653- 665.

[10] Boulder, D. (2001). Affine Term-Structure Models: Theory and Implementation. Bank of Canada Working Paper 2001-15.

[11] Bryson, A.E. and Henrikson, L.J. (1968). Estimation Using Sampled Data Containing Sequentially Correlated Noise. *Journal of Spacecraft and Rockets*, 5(6), 662-665.

[12] Campbell, J. and Viceira, L. (1997). Who Should Buy Long-Term Bonds? *American Economic Review*

[13] Cappuccio, N. and Lubian D. (1996). Fully Modified Estimation of Cointegrating Vectors Via VAR Prewhitening: A Simulation Study. *Journal of Italian Statistical Society*, 5(1), 13-37.

[14] Chen, H. and S. Joslin (2009). Generalized Transform Analysis of Affine Processes and Asset Pricing Applications. Working Paper MIT.

[15] Chen, R. and L. Scott (1993). ML Estimation for a Multifactor Equilibrium Model of the Term Structure. *Journal of Fixed Income*, December, 14-31.

[16] Chib, S. and B. Ergashev (2009). Analysis of Multi-Factor Affine Yield Curve Models. Working Paper SSRN.

[17] Christensen,J., F. Diebold, and G. Rudebusch (2008). The Affine Arbitrage-Free Class of Nelson-Seigel Term Structure Models. NBER Working Papers 13611, National Bureau of Economic Research, Inc.

[18] Christensen,J., F.Diebold, and G.Rudebusch (2009). An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model. forthcoming, *The Econometrics Journal*

[19] Christou, C. and Pittis, N. (2002). Kernel and Bandwidth Selection, Prewhitening, and the Performance of the Fully Modified Least Squares Estimation Method. *Econometric Theory*, 18(4), 948-961.

[20] Collin-Dufresne, P., R. Goldstein, and C. Jones (2002). Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance*, 57(4), 1685-1730.

[21] Collin-Dufresne, P., B. Goldstein, and C. Jones (2008). Identification of Maximal Affine Term Structure Models. *Journal of Finance*, 63, 743-795.

[22] Collin-Dufresne, P., B. Goldstein, and C. Jones (2009). Can Interest Rate Volatility be extracted from the cross-section of bond yields? *Journal of Financial Economics*, doi:10.1016/j.jfineco.2008.06.007

[23] Collin-Dufresne, P., R. Goldstein, and S. Martin (2001). The Determinants of Credit Spread Changes *Journal of Finance*, 56, 2177-2208.

[24] Conley, T.C., L.P. Hansen, G.J. Luttmer, and J. Scheinkman (1997). Short-term Interest Rates as Subordinated Diffusions. *Review of Financial Studies*, 10, 525-577.

[25] Cox, J., J. Ingersoll, and S. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53, 385-408.

[26] Dai, Q. and K.J. Singleton (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance*, 55, 1943-1978.

[27] Dai, Q. and KJ Singleton (2002). Expectations Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure. *Journal of Financial Economics*, 63, 415-441.

[28] Dejong, F. (2000). Time Series and Cross-Section Information in Affine Term-Structure Models. *Journal of Business and Economic Statistics*, 18, 300-314.

[29] Diebold, F. and C. Li (2006). Forecasting the Term Structure of government bond yields. *Journal of Econometrics*, 130, 337-364.

[30] Diebold, Francis X. Piazzesi, Monika and Rudebasch, Glenn D. (2005). Modeling Bond Yields in Finance and Macroeconomics. *American Economic Review*, 95, 415–421.

[31] Driessen, J. Melenberg, B. and Nijman, T. (2003). Common factors in international bond returns. *Journal of International Money and Finance*, 22, 629–656.

[32] Duarte, J. (2004). Evaluating an Alternative Risk Preference in Affine Term Structure Models. *Review of Financial Studies*, 17(2), 379-404.

[33] Duffie, D. and R. Kan (1996). A Yield Factor Model of Interest Rates. *Mathematical Finance*, 6, 379-406.

[34] Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68, 1343-1376.

[35] Duffie, D. and K. Singleton (1997). An Econometric model of the term structure of interest-rate swap yields. *Journal of Finance*, 52, 1287-1321.

[36] Duffee, G.R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance*, 57, 405-443.

[37] Duffee, G.R. (2009). Forecasting with the term structure: The role of no-arbitrage. Cal-Berkeley Working Paper.

[38] Edwards, M. and A. Richardson (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Letters of Nature*, 430, 881-884.

[39] Fisher, M. and C. Gilles (1996). Estimating exponential-affine models of the term structure. Federal Reserve Board Working Paper.

[40] Gorodnichenko, Y., Mikusheva. A., and Ng, S. (2009). Estimators for Persistent and Possibly Non-Stationary Data with Classical Properties. Working Paper.

[41] Gorodnichenko, Y. and Ng, S. (2010). Estimation of DSGE Models When the Data are Persistent. *Journal of Monetary Economics* forthcoming.

[42] Grasselli, M. and C. Tebaldi (2007). Stochastic Jacobians and Ricatti ODEs in affine term structure models. *Decisions Econ Finan*, 30, 95-108.

[43] Hamed, K.H. (2009). Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology*, 368(1-4), 143-155.

[44] Hamilton, J. (1994). Time Series Analysis, Princeton, NJ: Princeton University Press.

[45] Jamshidian, F. and Y. Zhu (1996). Scenerio Simulation: Theory and methodology. *Finance and Stochastics*, 1(1), 43-67.

[46] Jegadeesh, N. and Pennacchi, G. (1996). The Behavior of Interest Rates Implied by the Term Structure of EuroDollar Futures. *Journal of Money, Credit, and Banking* 28, 426-446.

[47] Jiang, G.J. (1998). Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Price of Derivative Securities. *Journal of Financial and Quantitative Analysis* 33, 465-497.

[48] Jones, C.S. (2003). Nonlinear Mean Reversion in the Short-term Interest Rate. *Review of Financial Studies* 16, 793-843.

[49] Joslin, S, K. Singleton, and H. Zhu (2010a). A New Perspective on Gaussian DTSM's. forthcoming, *Review of Financial Studies*

[50] Joslin, S, K.Singleton, and H. Zhu (2010b). Supplement to 'a new perspective on Gaussian DTSM's.' Technical Report, Sloan School, MIT.

[51] Kim, D.H. and A. Orphanides (2005). Term Structure Estimation with Survey Data on Interest Rate Forecasts. Finance and Economics Discussion Series, No. 48, Board of Governors of the Federal Reserve System.

[52] Koopman, S. J., M. Mallee, and M. Van Der Wel (2008). Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model with Time-Varying Parameters. forthcoming, *Journal of Business and Economic Statistics*

[53] Lamoureux, C. and K. Roskelley (2008). Empirically Confronting the Stochastic Singularity: An Application of the Cox, Ingersoll, and Ross Model. Available at SSRN: http://ssrn.com/abstract=871443.

[54] Langetieg, T (1980). A Multivariate Model of the Term Structure. *Journal of Finance*, 35, 71-97.

[55] Litterman, B. and J. Shienkman (1991). Common Factors Affecting Bond Returns. Research Paper, Goldman Sachs Financial Strategies Group.

[56] Kao, C, A. TamHane, and R. Mah (1992). A General Prewhitening Procedure for Process and Measurement Noises. *Chemical Engineering Communications*, 118(1), 49-57.

[57] Ludvigson, S. and S. Ng (2009). Macro Factors in bond risk premia. *Review of Financial Studies*. forthcoming.

[58] Nelson, C.R. and A.F. Siegel (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, 60, 473-489.

[59] Milicich, M.J., Meekan, M.G., and P.J. Doherty (1992). Larval Supply: a good predictor of recruitment of three species of reef fish (Pomacentridae). *Mar. Ecol. Prog. Ser.*, 86, p. 153-166.

[60] Merton, R. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8, 323-361.

[61] Pearson, N. and T. Sun (1994). Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model. *Journal of Finance*, 4, 1279-1304.

[62] Perignon, C., Smith, D., and Villa, C. (2007). Why Common Factors in International Bond Returns Are Not So Common. *Journal of International Money and Finance*. Forthcoming.

[63] Perignon, C. and C. Villa (2006). Sources of Time Variation in the Covariance Matrix of Interest Rates. *Journal of Business*, 79(3), 1535-1549.

[64] Piazzesi, M. (2003). Affine Term Structure Models. in L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics.* Amsterdam: North-Holland.

[65] Pyper, B.J. and R.N. Peterman (1998). Comparison of methods to account for autocorrelation in correlation analysesof fish data. *Can. J. Fish Aquat. Sci.*, 55, pp. 2127-2140.

[66] Quinn, T.J. II and Niebauer, H.J. (1995). Relaton of eastern Bering Sea walleye pollock (Theragra chalcogramma) recruitment to environmental and oceanographic variables. In Climatic change and northern fish populations. Edited by R.J. Beamish. *Can. Spec. Publ. Fish. Aquat. Sci.*, 121, pp. 497-507.

[67] Sam, Abdoul G. and George Jiang (2008). Nonparametric Estimation of the Short Rate Diffusion Process from a Panel of Yields. forthcoming, *Journal of Financial and Quantitative Analysis.*

[68] Scherer, K.P. and Avellaneda, M. (2002). All For One... One For All? A Principal Component Analysis of the Latin American Brady Bond Debt from 1994 to 2000. *Journal of International Theoretical and Applied Finance*, 5, 79–107.

[69] Schlens, J. (2005). A Tutorial on Principal Components Analysis. Institute for Nonlinear Sciences, UCSD.

[70] Singleton, Ken. (2006). Empirical Dynamic Asset Pricing. United Kingdom: Princeton University Press.

[71] Stanton, R. (1997). A Nonparametric Regression Under Alternative Data Environments. *Journal of Finance* 52, 1973-2002.

[72] Vargas-Ruzman, J.A., Warrick, A.W., and D.E. Myers (1999). Scale Effect on Principal Component Analysis of Vector Random Functions. *Mathematical Geology*, 31, 701-722.

[73] Vasicek,O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5, 177-188.

[74] Yue, S. and Wang. C. (2002). The influence of serial correlation on the Mann-Whitney test for detecting a shift in Median. *Advances in Water Resources*, 25(3), 325-333.