An Investigation into Alternative methods for estimating Gaussian Affine Term Structure Models

Januj Juneja

September 2009

Employing recent advances in the term structure estimation literature, this paper explores alternative econometric methods for the estimation of the Gaussian 3-factor sub-class of affine term structure models. Bond yields are taken to be an affine function of state variables that are the result of an empirical design consisting of structural rotations and affine invariant transformations. This empirical design would allow for a comparison between estimating an affine term structure model using the method of Collin-Dufresne, Goldstein, and Jones (2008) and the method of Duffie and Singleton (1997), within the context of Christensen, Diebold, and Rudebusch (2008). The contextual reference to Christensen, Diebold, and Rudebusch (2008) introduces restrictions that fully exploit the apriori advantages of modeling within the Nelson-Siegel (1987) framework as regards empirical performance. Using daily LIBOR rate and swap rate quotes from June 1996 to July 2008 to extract a panel of 3,430 time-series observations and 14 cross sections, this paper argues that, in contrast to what the recent literature suggests, a term structure model that is estimated using the method of Collin-Dufresne, Goldstein, and Jones (2008) does not perform significantly better in fitting yields, at any horizon, than the more traditional methods used in this literature. I work in the equivalent risk-neutral measure. Hence, while I cannot forecast out of sample, I also find preliminary evidence that is suggestive of improvements in existing methods. Results are robust to Kalman Filter Estimation.

First Draft: March 2008
This Version: September 2009

Keywords: Affine Term Structure Models, Econometric Methods, Nelson-Siegel, yield curve, Maximum Likelihood Estimation, Kalman Filter Estimation

Please note that this paper was previously titled "A Globally Identifiable Arbitrage-free dynamic Nelson-Seigel Model"

1 Introduction

This paper explores alternative methods for estimating state variables within the context of the Gaussian sub-class of Affine Term Structure Models. It is clear from the literature that 3 factors are sufficient to capture the majority of yield curve dynamics. Hence, this paper focuses on the three factor model.

Through the imposition of restrictions that permit the Gaussian sub-class of Affine Term Structure Models to be consistent with Nelson-Siegel (1987), I take advantage of, apriori, the empirical benefits of modeling within the Nelson-Siegel (1987) structure. For this reason, I compare alternative methods of estimating state variables within the context of the Christensen, Diebold, and Rudebusch (2008) model, which is generalized in Christensen, Diebold, and Rudebusch (2007). Christensen, Diebold, and Rudebusch (2008) enforce the absence of arbitrage, which is inherited from the canonical class of affine term structure models, within the aforementioned Nelson-Seigel (1987) representation, hence the name Arbitrage-Free Dynamic Nelson-Seigel (AFDNS) model. From the standpoint of empirical performance, the Nelson-Siegel (1987) term structure model is able to capture almost 96% of the variation in T-bill yields, however, it lacks a theoretical structure that is consistent with the absence of arbitrage (Nelson and Seigel, 1987), which presumably led Christensen, Diebold, and Rudebusch (2008) to their modification.[1]

[1] Coroneo, Nyholm, and Vidova-Koleva (2008) find that the Nelson-Seigel (1987) model performs just as well as the Christensen, Diebold, and Rudebusch (2008) model, and hence conclude that the enforcement of the absence of arbitrage restriction did not impact empirical performance at all. Almeida and Vicente (2008) find that the imposition of the no-arbitrage restriction improves the ability of the model to produce accurate forecasts. Duffee (2009) develops statistical tests to pin down the role of no-arbitrage on the forecasting ability, and in his application to dynamic term structure models finds that while the no-arbitrage restriction does not impact the ability of the model to produce accurate forecasts, he also argues that it slightly sharpens the results. Joslin, Singleton, and Zhu (2009) note that within the context of the AFDNS model, the no-arbitrage plays no role in the ability of the model to produce accurate forecasts. Consistency with the absence of arbitrage is a precursor for the development of a term structure model since bond markets are very liquid so that any arbitrage opportunities are assumed to be traded away instantaneously leading to the elimination of any risk-less profit opportunities (Ang, Dong, & Piazzesi, 2007)

Nelson and Seigel (1987) inspired models are also advantageous due to their simple and parsimonious structure and their flexibility to accommodate many shapes of the yield curve. Krippner (2006) also develops a class of theoretically consistent Nelson-Seigel models, however, it was formulated in discrete-time. So, although it is derived within the Heath, Jarrow, & Morton (1992) framework, it is not consistent with the majority of the models in our literature.[2]

Through the combination of the insights of Diebold and Li (2006) with those of Duffie and Kan (1996), Christensen, Diebold, and Rudebusch(2008) arrive at a model that is theoretically consistent with the absence of arbitrage and the Nelson-Siegel (1987) framework. While this model possesses some desirable properties, it may be worthwhile to utilize those properties through the model of Collin-Dufresne, Goldstein, and Jones (2008), which they claim has structural appeal that translates into advantages for the estimation and interpretation for model fitting of the yield curve. In order to fully exploit the structural appeal of the model of Collin-Dufresne, Goldstein, and Jones (2008) with the Nelson-Siegel (1987) restrictions of Christensen, Diebold, and Rudebusch (2008), I transform the equivalent risk-neutral dynamics of the arbitrage-free dynamic Nelson-Seigel (AFDNS) model into a model that retains the apriori empirical appeal of Christensen, Diebold, and Rudebusch (2008), while capturing the structural appeal of Collin-Dufresne, Goldstein, and Jones (2008). In this manner, 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 two models.

[2]The Duffie and Kan (1996) formulation fits within the Heath, Jarrow, and Morton (1992) framework which, among other assumptions, specifies a relationship between the forward rate curve and the expected future spot rate. Because of the linearity afforded by the affine representation, the Markov state variables in the Duffie and Kan (1996) can be taken to be the forward rates for given maturities. Thus, Duffie and Kan (1996) can be viewed as a Markov representation of the Heath, Jarrow, & Morton (1992) framework (see for e.g. Duffie and Kan (1996) or Frachot and Lesne (1993)). In Heath Jarrow and Morton (1992), the bond price process is non-Markov because the drift and volatility terms are dependent upon the history of the Brownian motion process

This transformation is permissible because the canonical affine term structure representations of Dai and Singleton (2000) consist of latent factors (see for e.g. Singleton(2006), Dai and Singleton (2000), or Collin-Dufresne, Goldstein, and Jones (2008) (henceforth called CDGJ(2008))). Dai and Singleton (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.[3] Joslin, Singleton, and Zhu (2009) impart new information about the estimation of Gaussian term structure models and suggest that there are benefits to carrying out these kinds of transformations.

The aforementioned transformation will allow me to compare, within the context of Nelson-Siegel (1987), models with state variables that are traditionally used in the literature with models that employ the set of state variables developed within the CDGJ (2008) framework. According to CDGJ(2008), 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, its drift, and the drift of the drift, 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 is also tied to the yield curve directly and should be superior, at least at short maturities, to a method that does not directly use known information about the yield curve at

[3]This also implies that two models that may seem superficially different could in fact be the same model. For example, Joslin, Singleton, and Zhu (2009) find that the AFDNS model of CDR(2008) is an invariant transformation of Dai and Singleton (2000). Another good example of this idea is a model proposed by Babbs and Nowman (1999). This model is "maximal" in the sense of Dai and Singleton (2000), however, there are multiple solutions which satisfy the requirements of a maximum for the conditional likelihood function (see, e.g., Collin-Dufresne, Goldstein, and Jones, 2008). 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.

short maturities. This paper investigates this claim by comparing the method of CDGJ(2008) as a method for estimating state variables against the method of Duffie and Singleton (1997) or Pearson and Sun (1994) for estimating state variables within the context of affine term structure models.

This paper argues 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 perform significantly worse, as regards empirical performance at any horizon, than the method suggested by CDGJ (2008), henceforth called the CDGJ (2008) method. 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 e.g. Dejong (2000), Hamilton (1994), or Chen and Scott(1993) for more information about the Kalman Filter estimation method). The stark differences in empirical performance obtained in the recent literature seem to have been a manifestation of the choice of yields and not related to the superior choice of econometric methods used to estimate the term structure model. This result should not be surprising given the fact that Joslin, Singleton, and Zhu (2009) find the model of CDR (2008) to be an invariant transformation of Dai and Singleton (2000) (Joslin, Singleton, and Zhu, 2009). This research also concludes that econometric methods that have been suggested by the recent term structure literature (see for e.g. Dai and Singleton (2000), Singleton (2006), CDGJ(2008), or Joslin, Singleton, and Zhu (2009)) can be successfully used to improve the empirical performance of affine term structure models. While, I cannot forecast out of sample, preliminary evidence is suggestive of improvements in existing methods.

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 the method of CDGJ (2008)). Section V presents and interprets the results. Section VI presents and interprets the yield errors. Finally, Section VII gives the conclusion.

2 The Affine Class of Term Structure Models

The standard Affine class of term structure models of Duffie and Kan (1996) and Dai and Singleton (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

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

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 model of Vasicek (1977), Langetieg(1980), Cox, Ingersoll, and Ross (1985), Jegadeesh and Pennacchi(1996), Christensen, Diebold, and Rudebusch (2008), and Collin-Dufresne, Goldstein, and Jones (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.

2.1 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 e.g. Langetieg(1980), Collin-Dufresne, Goldstein, and

Jones (2008), Joslin, Singleton, and Zhu (2009)). 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

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

where Σ^R is an orthogonal matrix that is estimated numerically due to certain imposed constraints associated with the transformation process (see Appendix A and/or Appendix D for more details). Yields are then taken to be a function of these state variables. The presentation of the above system of stochastic differential equations for the state variable transition dynamics makes it easy for the reader to understand the hype regarding the structural appeal of the model.

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.* The level of the short rate is tied directly to the yield curve, its drift and the subsequent drift of the drift is also tied to the yield curve, again in the equivalent risk- neutral measure.

3 Data

The dataset used here derives from daily observations of Libor rates of maturities of 1,3,6,9, and 12 months and swap rate quotes for 2,3,4,5,7, 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 quoted 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 Collin-Dufresne, Goldstein, and Jones (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 optimizations are shown below.

Table I about here

Table II about here

Table III about here

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%.

4 An Overview of the CDGJ(2008) method for estimation of the state variables

Using the CDGJ (2008) method, the set of state variables are 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 the approach of CDGJ (2008) is that it allows for error reduction by imposing some parametric structure, while retaining flexible structure. (for more information see Appendix F)

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. Following the insights of CDGJ (2008), 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 down 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 and understanding the relation above, 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 the table below.

Table IV about here

Please see appendix G for details on a small Monte Carlo Simulation study that checks the accuracy and robustness of this state variable estimation procedure.

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. To avoid using the same information twice, three different bonds are assumed to be measured with error (Collin-Dufresne, Goldstein, and Jones, 2008). Bonds with similar characteristics are used in order to minimize the informational disparity across state variables. This method is essentially the same as 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 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 2 conditional moments of the state variables, which simplifies the analysis. (Please see Appendix E for the details of those calculations.) Symbolically,

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

and hence,

$$X(t) = (\frac{(B(\tau))^T}{\tau})^{-1}(Y(t,\tau) - \frac{A(t,\tau)}{\tau})$$

The final form of this part of the likelihood is

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

For the three bonds which are measured with error, I can not 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,

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

The conditional distribution of the errors is

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

where Σ is the variancecovariance matrix for the measurement errors. Using the properties of the Normal distribution, the log-likelihood function can be obtained and so the sum of the two log-likelihood functions (without error and with error) is the log-likelihood function that is to be optimized. 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 that were obtained based upon the procedure outlined in Section 2 and Section 4, while maintaining the assumption for those yields which are assumed to be measured with error. For both methods, the entire parameter space over which I am going to find an optimum is given by:

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

In a moment's reflection, an avid reader may be puzzled by the fact that the parameters governing the value of the long-run mean for each factor do not show up in the parameter space. To stymie any confusions, please note that for any canonical representation equivalent of the AFDNS model with real eigenvalues in the mean reversion matrix under the $Q-measure$, it does not matter which measure I impose identifying restrictions(see Christensen, Diebold, and Rudebusch (2008) for more details). There is no loss in generality if I fix the mean of the model under the Q-measure to be 0, while leaving it to be estimated under the actual dynamics (Christensen, Diebold, & Rudebusch, 2008). Furthermore, recent work by Joslin, Singleton, and Zhu (2009) find that the forecasting ability associated with the P-dynamics of the Christensen, Diebold, and Rudebusch (2008) model to be no better than an unconstrained VAR(1) model. This further highlights the need to better understand the dynamics of the relationship between the P-dynamics and Q-dynamics.

Moreover, while I work in the equivalent risk-neutral measure, I am working within the model framework of Christensen, Diebold, and Rudebusch (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).

The tables below presents the results from the maximum likelihood estimation.

Table VI about here

Table VII about here

Standard errors were computed using the Berndt, Hall, Hall, and Hausman (1974) approach. Please also note that previous studies (see for e.g. 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 $n = 3,430$ 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. 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.

Upon reflection of the maximum likelihood estimation of the two models, several observations are in order. As CDGJ(2008) note, the CDGJ (2008) 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.

For both methods, those parameters that were involved in the rotation of the original model proposed by Christensen, Diebold, and Rudebusch (2008), and of course directly involved in the optimization process, there was a reduction in the nonlinearity of the likelihood surface which has been known to confound the optimization procedure through the introduction of saddle points to the surface. Although the likelihood function surface was still flat, the reduction in the nonlinearity of the surface did facilitate obtaining a global optimum over an already reduced space. Moreover, as concluded by Duffee (2009), the entire optimization procedure for both methods, for my parametrization, took about 2 days.

5.1 Robustness checks via 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 other 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 (please see CDR(2008) for more details). 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 below. Please note that all 6 of these bonds, the 3 month, 6 month, 2 year, 3 year, 4 year, and 5 year, to be measured with error.

Table VIII about here

The log-likelihood function on $n = 3430$ observations is 202,132.699. Please note that, unless otherwise stated, the elements of the measurement error matrix and their standard errors were multiplied by 10000.

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 would it have been the case that had I used more information from the yield curve with the inversion method, that it too would have performed better at the long end of the yield curve.

Table IV about here

On the whole, the results generally support my hypotheses. It is clear that for the inversion method, the errors are larger the further I move out on the yield curve, i.e. 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 two, three, and four year maturities is not useful for fitting maturities greater than 6 years. For the CDGJ (2008) method, generally, the short-end of the yield curve is characterized by smaller error than the longer end. This is due to the fact that the procedure is predicated on the fact that the state variables are a reflection of the information that we have today about zero-coupon yields. 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, independent of one's computational method of choice for state variable transition dynamics estimation.

7 Conclusion

Employing recent advances in the term structure estimation literature, this paper set out to explore alternative methods for estimating state variables within the context of the Gaussian sub-class of Affine Term Structure Models. It is clear from the literature that 3 factors are sufficient to capture the majority of yield curve dynamics. Hence, this paper focused on a three factor model.

Using a unique design specification consisting of both affine invariant transformations and Brownian motion rotations, this paper examined the empirical benefits to term structure modeling using alternative econometric methods for estimating state variables within the Gaussian subclass of 3 factor affine term structure models. The purpose of these translations and rotations constituting the design specification was to allow for a comparison between the CDGJ(2008) method and the Duffie and Singleton(1997) methods, both of which are used in the literature to estimate state variables, within the context of the Christensen, Diebold, and Rudebusch (2008) model. The contextual reference to Christensen, Diebold, and Rudebusch (2008) introduced restrictions that fully exploit the apriori advantages of modeling within the Nelson-Siegel (1987) framework as regards empirical performance.

In this manner, I 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 two models. This paper 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 perform significantly worse, as regards empirical performance 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. 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 choice of econometric methods used to estimate the term structure model. This result should not be surprising given the fact that Joslin, Singleton, and Zhu (2009) find the model of CDR (2008) to be an invariant transformation of Dai and Singleton (2000) (Joslin, Singleton, and Zhu, 2009). I worked in the equivalent risk-neutral measure. Hence, while I could not forecast out of sample, I also found preliminary evidence that is suggestive of empirical improvements in existing methods used to estimate term structure models. Employing the techniques mentioned in the recent literature (see for e.g. Dai and Singleton (2000), Singleton (2006), CDGJ (2008)or Joslin, Singleton, and Zhu (2009)) seem to lead to improvements in existing methods used to estimate Gaussian Term Structure Models.

The model in this paper stands to gain from improvements that can be obtained from the fact that I worked with a Gaussian type structure, which has a prominent limitation in that it allows interest rates to become negative. CIR (1985) type models handle this unlikely, but possible scenario. Koopman, Mallee, & Van der Waal (2008) cast the AFDNS model into one that resembles more closely the CIR (1985) model, at least as regards the time-series properties of CIR (1985). Although Koopman, Mallee, &Van der Waal (2008) use time-series methods like GARCH in conjunction with spline methods to give time-series properties to the volatility component, this could be a starting point for future work in which I attempt to obtain a more flexible structure and one that is consistent with CIR (1985), notably positive interest rates. One of the other drawbacks associated with this model is that the process describing the state variables, while admissible in the language of Dai and Singleton(2000), is characterized by a Brownian motion that can potentially become unstable. Potential stabilizers to the diffusion part of the stochastic process have been introduced by Dixit (1991) and Krugman (1988) and have been termed regulated Brownian motions. The regulation of the Brownian motion term is also consistent with the fact that the Fed has the power to regulate the short term interest rate through its influence on the Federal Funds Rate. This further suggests that models such as those of Farnsworth and Bass (2003) can shed further light on this problem. The inclusion of these insights into these type of models is a very interesting area of future research.

Another potentially interesting area of future research involves the choice of yields that are used to invert to obtain state variables. This selection seems to be an arbitrary procedure with two exceptions. The first is motivation surrounding the utilization of information from different parts of the yield curve. So, I have N arbitrarily chosen yields which I use to back out N state variables. These N yields are then assigned to have a zero pricing error, while additional yields are added to consider an error structure. So, the procedure works because if I have N yields that I take to be free of error and I have M available maturities, then I need $(M - N)$ error terms (Pastorello, Patilea, and Renault, 2003). Hence, the second exception arises from the fact that any other yield (other than any of the N yields chosen for the analysis) can reject the model since it cannot exactly price all observed securities (Lamoureux and Roskelley, 2008).

An alternative way to express the above statements is to say that the variance-covariance matrix on $(N + 1)$ yields is singular (Piazzesi, 2003). This feature of affine term structure models (henceforth, known as ATSM) is known as stochastic singularity.

Some authors argue that this stochastic singularity feature is the reason that the evidence on the empirical performance of the ATSM is not conclusive (Lamoureux and Roskelley, 2008). Much of this consternation is associated with the fact that different researchers use different error structures in the estimation process. Further confusion can be generated by the fact that while the ATSM is understood theoretically, that underlying theory provides no guidance on the construction of the error terms (Lamoureux and Roskelley, 2008). This research also suggests the necessity for the investigation of the role that the error term plays in the estimation and inference of affine term structure models.

While this problem is not simple to solve and has been overlooked by most authors in the literature, a good empirical design can shed light on this problem

(Lamoureux and Roskelley (2008)).

A good empirical design can allow me to learn from the data within the context of the model. This would involve the specification of a prior distribution which could characterize prior beliefs about the structure of error term and the use of a likelihood function or model to update those beliefs to arrive at a more informed view. This suggests that Bayesian methods could play a role in this design. To some extent, the construction of a good empirical design can shed light on how the model and hence the error structure can allow me to learn from the data. This empirical design could involve the usage of a particle filter to simulate the state variables and then Bayesian methods to obtain inferences from the data about the parameters involved with the distribution of the error term (see Pitt and Shepard (1999) for an overview of the particle filter and Polson, Stroud, and Miller (2001) for drawing inference from the error term). Since most of the literature works with a Gaussian distribution assumption, preliminary work could examine whether or not this assumption is consistent with what we see in the data perhaps through attempts to draw inference from 2 parameter families.

8 Appendices

8.1 Appendix A: Demonstration of the equivalent $A_0(3)$ model

Christensen, Diebold, and Rudebusch (2008) (henceforth called 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.[4] Under this assumption, the Ricatti equations become a simple first order ordinary differential equation and can be solved by integration and the imposition of the Nelson-Seigel structure. (Please see Appendix B for more details.)

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 actual world, through the physical measure, and the equivalent risk-neutral world, with 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 permits it to be equivalent to the model of CDR (2008), while containing state variables following the method of CDGJ (2008). In this manner, 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 two models. Since this paper starts from the equivalent risk-neutral dynamics as specified in CDR (2008), I maintain their set of assumptions and sets out to transform the above $A_0(3)$ system governing the equivalent risk-neutral dynamics of unobservable or latent state variables into variables that are observable at the short-end of the yield curve and hence

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

estimable according to CDGJ (2008). CDGJ(2008) point out that the reason for this is that the 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*. This paper investigates this claim. In this context, 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 maximize any attempts to improve the empirical performance associated with the maximum likelihood estimation procedure.

Symbolically, an affine invariant transformation is given by

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

Let's start by specifying 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 $\mu_{1,t}$. So, this process is repeated for $\mu_{2,t}$ to get the process for $\mu_{2,t}$. Let's start with

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

Note that this follows from the fact that the short rate is an affine function within this class of models. In order to find the stochastic process for the short rate, this expression needs to be converted into one that includes differentials. It 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. The expression above is for the short rate process, which, according to CDGJ(2008) is termed an observable variable since it can be obtained from data on fixed income security prices. Hence, the risk-neutral drift of the short rate process can also be observed, given that this process is for an observable variable (Collin-Dufresne, Goldstein, & Jones, 2008). The process for the drift of the short rate process can be found as follows.

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

Hence,

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

Thus,

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

Again, substituting in the expressions for dX_t^3 and dX_t^2 and segregating the drift and diffusion terms.

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

$\Sigma^* dW_t^{(2,Q)})$

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

According to CDGJ (2008), 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:

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

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

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

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

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

Further, note that the process above is incomplete because it still has terms on the right hand side that are in terms of X_t^2 and X_t^3. I eliminate those variables by solving the new system of equations, which is currently in terms of $(\mu_t^{(1,Q)}, \mu_t^{(2,Q)})$ to be in terms of (X_t^2, X_t^3) so that I can plug into the expression above to arrive at the final transformed system. So, taking $(\mu_t^{(1,Q)}, \mu_t^{(2,Q)})$ and all the parameters as given, 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 allows me to put all the given terms on one side and the variables on the other side,

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

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,

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

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

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

$$\Rightarrow \begin{bmatrix} -\frac{2}{\lambda\delta_{x_2}}[\mu_t^{(1,Q)} - \delta_{x_2}\lambda\theta_2^Q + \lambda\delta_{x_2}\theta_3^Q] - \frac{1}{\delta_{x_2}\lambda^2}[\mu_t^{(2,Q)} + \delta_{x_2}\lambda^2\theta_2^Q - 2\delta_{x_2}\lambda^2\theta_3^Q] \\ -\frac{1}{\lambda\delta_{x_2}}[\mu_t^{(1,Q)} - \delta_{x_2}\lambda\theta_2^Q + \lambda\delta_{x_2}\theta_3^Q] - \frac{1}{\delta_{x_2}\lambda}[\mu_t^{(2,Q)} + \delta_{x_2}\lambda^2\theta_2^Q - 2\delta_{x_2}\lambda^2\theta_3^Q] \end{bmatrix}$$

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

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

In the steps above, I perform an affine invariant transformation which is associated with the drift part of the stochastic process system. I am now concerned with the the diffusion part of the stochastic process system. In dealing with the diffusion term, I need to perform what is called a rotation of the Sigma matrix associated with the vector of independent Brownian motion terms. While the details of the computation are presented in Appendix D, I present a short synopsis of those steps here. Essentially, I need to figure out a transformation matrix that is invertible and an orthogonal matrix for the Brownian motion term (See Appendix D for more details). The transformation needs to do the following:

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

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:

$$d \begin{bmatrix} r_t \\ \mu_{1,t}^Q \\ \mu_{2,t}^Q \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & \lambda^2 & 2\lambda \end{bmatrix} \begin{bmatrix} 0 - r_t \\ 0 - \mu_{1,t}^Q \\ 0 - \mu_{2,t}^Q \end{bmatrix} dt + \Sigma^R dB_t^Q$$

Note that the pre-rotated form, i.e. $\Sigma \neq \Sigma^R$, of the covariance matrix for the state variables is consistent with the correlated-factor model of CDR (2008).[5] This consistency results in a lower triangular Σ matrix and holds for 2 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 the drift of the slope state variable 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 (Christensen, Diebold, & Rudebusch, 2008). (For more information on the rotated form of the Σ matrix, please see Appendix D.)

Furthermore, note that the new system has a column of zeros and that the reason for this is the fact that the matrix which was responsible for the imposition of restrictions necessary in order to expand the Arbitrage-free class of models of Duffie and Kan (1996) to include the dynamic Nelson-Seigel (1987)

[5]The Σ^R, the rotated version of the Σ matrix, is rotated by the matrix, O, where O is
$O = \begin{bmatrix} \delta_{x_1} & \delta_{x_2} & 0 \\ 0 & -\delta_{x_2}\lambda & \delta_{x_2}\lambda \\ 0 & \delta_{x_2}\lambda^2 & -2\delta_{x_2}\lambda^2 \end{bmatrix}$
O must be an orthogonal matrix, and so its usage introduces some constraints. Please see Appendix D for more details on the mechanics of this computation.

framework, also had the same independence of the state variables from the "level" factor. (Please see Appendix B for more details.) 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). To further facilitate the comparison of this model framework to that proposed by CDGJ (2008), it is worth noting that the state variables in the system of stochastic differential equations above are directly observable since:

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

where $Y^n \equiv \frac{\partial Y(\tau)}{\partial \tau^n}|_{\tau=0}$

and the state variables which are r(t), its risk-neutral drift and the drift of the drift are tied to that expression, as an application of Ito's lemma (see Appendix F for full derivation), by the following series of expressions

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

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

$Y^2(t) = \frac{1}{3}[\mu_2(t) - V_{00}(t)]$

Thus, a time series of state variable realizations can be obtained by exploiting the Taylor Series expansion at very short maturities (Collin-Dufresne, Goldstein, & Jones, 2008). Note that $V_{00}(t)$ is the variance of the short rate.

More conspicuously, CDGJ(2008) claim that these state variables are tied directly to the yield curve. Hence, they argue that there is no ambiguity regarding their interpretation and this plays a very important role in improving the numerical tractability of the maximum likelihood estimation procedure.

8.2 Appendix B: Proof that the Arbitrage-free Dynamic Nelson-Seigel (AFDNS) class of models falls within the $A_0(3)$ class

This appendix is essentially the same as that in CDR (2008), in that it has no new or unique insight. The motivation for a more detailed exposition of the derivation of CDR(2008) is to demonstrate the lack of dependence of the factors in my rotated model, with just observable variables, on the state variable describing the "level" factor, $r(t)$. 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:

$$\frac{dB(t,T)}{dt} = \rho_1 + (\kappa^Q)^T B(t,T),\ B(T,T) = 0$$

$$\frac{dC(t,T)}{dt} = \rho_0 - B(t,T)^T (\kappa^Q)^T \theta^Q - \frac{1}{2}\sum_{j=1}^n (\Sigma^T B(t,T) B(t,T)^T \Sigma)_{j,j},\ C(T,T) = 0$$

Now, since I want to find the state variables $B(t,T)$ that impose the dynamic Nelson-Seigel (1987) structure onto the affine-class of models of Duffie and Kan (1996), I will only focus on the top equation restated here for purposes of clarity.

$\frac{dB(t,T)}{dt} = \rho_1 + (\kappa^Q)^T B(t,T),\ B(T,T) = 0$

In order to solve this equation for $B(t,T)$, I make a note of the following relation which will help me to isolate $B(t,T)$.

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

Plugging in the expression above for $\frac{dB(t,T)}{dt}$,

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

Collecting terms and simplifying the expression yields,

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

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),

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

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

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:

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

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

So,

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

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

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

In order to exponentiate the matrix, I must use the matrix version of the power series expansion for the exponential function. This requires me to compute the matrix A^2 and A^3 in order to develop the pattern and to obtain the correct matrix.

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

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

We have that the above is

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

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

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

Noting the series definition of the exponential function,

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

Finally,

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

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

8.3 Appendix C: Zero Coupon Bond Yield Extrapolation Procedure

It is the purpose of this section to detail how I backed out the zero coupon yields from the swap rate quotes. I started from the time-series of Libor rate quotes, and used that data to back out forward rates at the short-end of the yield curve, i.e. out 1 year. To extrapolate the 2 year spot rate, I made the assumption following Fama and Bliss (1987) that the forward rate remains constant during the year. 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, while the notional principal was $10,000,000$. 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,

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

where x is the set of forward rates. It was a singleton set in obtaining the 2,3,4, and 5 year rates, while it contained 2 forward rate elements for the 5 to 7 year swap rate quotes and 3 forward rate elements for the 7 to 10 year swap rate quotes.

8.4 Appendix D: Rotation of Sigma: Orthogonalization Process

In order to carry out this transformation, I will need to find a orthogonal transformation for the A matrix that I will then apply that to the Brownian motion term so that it remains unchanged. In order to begin to consider the specific form for this matrix, lets recap, very generally, what has been done until this point. I started 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, I will need to find T, where T is a matrix, s.t.,

$T(A\Sigma_Y) = O$

where O is an orthogonal matrix. Thus, we will have orthogonally transformed the Brownian motion, without affecting the drift term.

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

In order to implement the Brownian motion rotation, I need to 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, I would want 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 my parameters, in order to establish an orthogonal transformation of the diffusion term.

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

$\Sigma_{X_{12}}:(\delta_1\sigma_{11}+\delta_2\sigma_{21})(\delta_2\sigma_{22})+(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})(-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})+(\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})=0$

$\Sigma_{X_{13}}:(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})(\delta_2\lambda\sigma_{33})+(\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})(-2\delta_2\lambda^2\sigma_{33})=0$

$\Sigma_{X_{21}}:(\delta_2\sigma_{22})(\delta_1\sigma_{11}+\delta_2\sigma_{21})+(-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})+(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})(\sigma_{21}\lambda^2\delta_2-2\delta_2\lambda^2\sigma_{31})=0$

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

$\Sigma_{X_{23}}:(-\delta_2\lambda\sigma_{22}+\delta_2\lambda\sigma_{32})(\delta_2\lambda\sigma_{33})+(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})(-2\delta_2\lambda^2\sigma_{33})=0$

$\Sigma_{X_{31}}:(\delta_2\lambda\sigma_{33})(-\delta_2\lambda\sigma_{21}+\delta_2\lambda\sigma_{31})+(-2\delta_2\lambda^2\sigma_{33})(\delta_2\lambda^2\sigma_{21}-2\delta_2\lambda^2\sigma_{31})=0$

$\Sigma_{X_{32}}:(\delta_2\lambda\sigma_{33})(-\delta_2\lambda\sigma_{22}+\delta_2\lambda^2\sigma_{32})+(-2\delta_2\lambda^2\sigma_{33})(\delta_2\lambda^2\sigma_{22}-2\delta_2\lambda^2\sigma_{32})=0$

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

Since there are several constraints which do not seem to have any immediate intuitive interpretation, below I attempt to simplify these constraints so that I can potentially uncover more information about my parameter space.

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

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

$\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2 + 4\lambda^4) = 0$

$\Sigma_{X_{13}}: \quad -\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{31}\sigma_{33}(\lambda^2 + 4\lambda^4) = 0$

$\Sigma_{X_{21}}: \quad \delta_2^2\sigma_{21}\sigma_{22}(\lambda^4+\lambda^2+1)+\delta_1\delta_2\sigma_{11}\sigma_{22}-(\delta_2^2\sigma_{21}\sigma_{32}+\delta_2^2\sigma_{31}\sigma_{22})(\lambda^2+2\lambda^4)+$
$\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2 + 4\lambda^4) = 0$

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

$\Sigma_{X_{23}}: \quad -\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{33}\sigma_{32}(\lambda^2 + 4\lambda^4) = 0$

$\Sigma_{X_{31}}: \quad -\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{31}\sigma_{33}(\lambda^2 + 4\lambda^4) = 0$

$\Sigma_{X_{32}}: \quad -\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{32}\sigma_{33}(\lambda^2 + 4\lambda^4) = 0$

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

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,

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

$$\Sigma_{X_{12}}: \quad \delta_2^2\sigma_{21}\sigma_{22}(\lambda^4+\lambda^2+1)+\delta_1\delta_2\sigma_{11}\sigma_{22}-(\delta_2^2\sigma_{21}\sigma_{32}+\delta_2^2\sigma_{31}\sigma_{22})(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{31}\sigma_{32}(\lambda^2+4\lambda^4) = 0 \tag{2}$$

$$\Sigma_{X_{13}}: \quad -\delta_2^2\sigma_{21}\sigma_{33}(\lambda^2 + 2\lambda^4) + \delta_2^2\sigma_{31}\sigma_{33}(\lambda^2 + 4\lambda^4) = 0 \tag{3}$$

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

$$\Sigma_{X_{23}}: \quad -\delta_2^2\sigma_{22}\sigma_{33}(\lambda^2+2\lambda^4)+\delta_2^2\sigma_{33}\sigma_{32}(\lambda^2+4\lambda^4)=0 \tag{5}$$

$$\Sigma_{X_{33}}: \quad \sigma_{33}^2\delta_2^2(\lambda^2+4\lambda^4)=1 \tag{6}$$

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 me the elimination of three of the constraints and the combination of two constraints into 1, so that in the end I am trying to estimate 10 parameters with 2 constraints. The details of this results are outlined as follows. I start with equation (6) above and isolate the λ related variables. Hence,

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

By plugging (6), in its rearranged form, into (3), I have 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 solution to this system of non-linear equations in the parameter space gives me λ. 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 $\lambda^2=\sqrt{\lambda^4}$ from a numerical standpoint.

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 in (4) yields an expression for δ_2^2 as follows:

From (4),

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

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 in equations (3), (4) and (6) 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 (3), (5), and (6)

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 (1) and (2), Please note that 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. The interpretations, however, remain the same. δ_1 and δ_2 are parameters associated with the Brownian motion rotation of the diffusion process, while the interpretation λ obtains from the AFDNS model. Please note that $\{\delta_1, \delta_2\}$, are different from the parameters that are traditionally associated with the short rate process in affine term structure models and are treated accordingly.

8.5 Appendix E: Analytic Computation of the first and second moments of the observable $A_0(3)$ system

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

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

$$sr_t = \delta_o + \delta_r r_t + \delta_{\mu_1}\mu_t^{(1,Q)} + \delta_{\mu_2}\mu_t^{(2,Q)}$$

In order to preserve some degree of generality, I 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

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

These allow me to re-write the system of observable variables as shown below:

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

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, I 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,

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

Now, I turn to identifying the exact functional form of the integrating factor, note that

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

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

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, I used a Simpson-method adapted quadrature after carrying out some simple algebraic manipulations.

8.6 Appendix F: Description of the CDGJ (2008) method for estimating state variables

In this section, I review the method that Collin-Dufresne, Goldstein, and Jones (2008) (henceforth, 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, 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:

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

where

$P_k(t) = k^{th} principal component factor$

$f_k(t) = factorloadingfunction\ (i.e. polynomial expansion for the k^{th} factor)$

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_i 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 and second order cross-partial derivatives with respect to pairs of state variables allows me to connect those derivatives to the bond pricing equation

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

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, after carrying out the above steps and collecting terms, the first three state variables are

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

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

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

Since I use the 3 factor sub family of arbitrage-free affine term structure models, these are my set of state variables.

9 Appendix G: Monte Carlo Simulation and the Robustness of the CDGJ(2008) 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.

In general, Monte Carlo simulations are useful for assessing the level of accuracy and bias in parameter estimates. This type of exercise can be very handy when carrying out estimation procedures that are rather non-ubiquitous within any literature.

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) (Collin-Dufresne, Goldstein, & Jones, 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:

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

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

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

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

Please note that 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. The basic summary statistics, across all simulations, of these estimates are shown in the table below.

Table V about here

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 Collin-Dufresne, Goldstein, and Jones (2008), who also find some bias in the third state variable.

References

[1] Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant (2002). Quadratic Term Structure Models: Theory and Evidence. *Review of Financial Studies*, 15, 243-288.

[2] Almeida, Caio Ibsen Rodrigues de and J. Vicente (2008). The Role of No-Arbitrage on Forecasting: Lessons from a Parametric Term Structure Model. *Journal of Banking and Finance*, 32, 2695-2705.

[3] 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*.

[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] 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.

[6] Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A. B. Sanders (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *Journal of Finance*, 47, 1209-1227.

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

[8] Christensen,J., F.Diebold, and G.Rudebusch (2007). An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model., forthcoming, *The Econometrics Journal*

[9] 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.

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

[11] 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

[12] Coroneo, L., K. Nyholm, and R. Vidova-Koleva (2008). How Arbitrage-free is the Nelson-Seigel Model? European Central Bank Working Paper Series No. 874.

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

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

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

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

[17] Dixit, A. (1991). A Simplified Treatment of the theory of optimal regulation of Brownian Motion. *Journal of Economic Dynamics and Control*.

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

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

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

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

[22] Fama, E. and R. Bliss (1987). The information in Long-Maturity Forward Rates. *The American Economic Review*, 77, 680-692.

[23] Farnsworth, H. & Bass, R. (2003). Term Structure with Semi-credible Targeting. *Journal of Finance*, 58, pp. 839–866.

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

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

[26] Heath D., R. Jarrow, and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60, 77-105.

[27] 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.

[28] Joslin, S, K. Singleton, and H. Zhu (2009). A New Perspective on Gaussian DTSM's. Working Paper Stanford University.

[29] 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.

[30] 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*

[31] Krippner, L. (2006). A Theoretically Consistent Version of the Nelson and Seigel Class of Yield Curve Models. *Applied Mathematical Finance*, 13, 39-59.

[32] Krugman, P. (1988). Target Zones and Exchange Rate Dynamics. NBER Working Paper 2481. National Bureau of Economic Research, Inc.

[33] 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.

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

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

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

[37] Moler, C. and C. Van Loan (2003). Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *Society for Industrial and Applied Mathematics*, 45, 3-49.

[38] Pastorello, S., V. Patilea, and E. Renault (2003). Iterative and recursive estimation in structural nonadaptive models. *Journal of Business and Economic Statistics*, 21, 449-482.

[39] 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.

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

[41] Pitt, M. and Neil Shephard (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446), 590-599.

[42] Polson, N., J. Stroud, and P. Muller (2001). Affine state-dependent variance models. Working Paper, University of Chicago and Wharton.

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

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

Table 1: Objective Function Summary Statistics I

$SumStat$	$Obj1$	$Obj2$	$Obj3$	$Obj4$	$Obj5$	$Obj6$
Max	9.93382E-07	9.91444E-07	9.99883E-07	9.99675E-07	3.82864E-06	6.16223E-06
$Mean$	5.7768E-08	4.18824E-08	1.00482E-07	2.59315E-08	5.10897E-08	9.5245E-08
Min	0	0	0	0	0	0

Table 2: Spot Rate Summary Statistics I

$SumStat$	$1mo$	$3mo$	$6mo$	$9mo$	$12mo$	$2yr$	$3yr$
$Mean$	.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 3: Spot Rate Summary Statistics II

$4yr$	$5yr$	$6yr$	$7yr$	$8yr$	$9yr$	$10yr$
0.048923095	0.050629149	0.050956625	0.052466774	0.054328929	0.05479061	0.054736418
0.013608364	0.012469005	0.0127204	0.011306438	0.009487977	0.009178017	0.010248859
0.076653834	0.077071259	0.07714431	0.077500499	0.078732693	0.077706258	0.078018169
0.017390459	0.0210382	0.017783384	0.025563812	0.022162332	0.025846701	0.013836525

Table 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 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 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)

Note: The elements of the measurement error matrix and their standard errors were multiplied by 10000

Table 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}	$9.05e^{-4}$ $(1.596e^{-6})$
$\Sigma_{me_{11}}$	0.005091 (0.044415)
$\Sigma_{me_{21}}$	-0.00350 (0.027970)
$\Sigma_{me_{22}}$	0.01010 (0.4994276)
$\Sigma_{me_{31}}$	0.02280 (0.1447076)
$\Sigma_{me_{32}}$	-0.0294 (0.631626)
$\Sigma_{me_{33}}$	0.1266 (1.3336224)

Note: The elements of the measurement error matrix and their standard errors were multiplied by 10000

Table 8: 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}	$8.039e^{-4}$ ($5.732e^{-7}$)
$\Sigma_{me_{11}}$	0.005091 (0.000001)
$\Sigma_{me_{21}}$	−0.00350 (0.002235)
$\Sigma_{me_{22}}$	0.01010 (0.000922)
$\Sigma_{me_{31}}$	0.02280 (0.003385)
$\Sigma_{me_{32}}$	−0.0294 (0.000008)
$\Sigma_{me_{33}}$	0.1266 (0.002004)
$\Sigma_{me_{41}}$	0.01593 (0.0000028)
$\Sigma_{me_{42}}$	$4.8318e^{-7}$ (0.0000005)
$\Sigma_{me_{43}}$	0.03513 (0.0000007)
$\Sigma_{me_{44}}$	0.01695 (0.00000083)
$\Sigma_{me_{45}}$	−0.007953 (0.00000268)
$\Sigma_{me_{51}}$	0.09995 (0.0000376)
$\Sigma_{me_{52}}$	0.04486 (0.0000013)
$\Sigma_{me_{53}}$	−0.04503 (0.00000291)
$\Sigma_{me_{55}}$	0.03022 (0.000000761)
$\Sigma_{me_{56}}$	−0.4043 (0.000000751)
$\Sigma_{me_{61}}$	0.01693 (0.0000026)
$\Sigma_{me_{62}}$	−0.1295 (0.0000356)
$\Sigma_{me_{63}}$	0.09999 (0.0000226)
$\Sigma_{me_{64}}$	−0.09053 (0.000001105)
$\Sigma_{me_{66}}$	0.49618 (0.000002216)

Table 9: $A_0(3)$ Yield Errors: This table contains root mean squared errors computed from yields over the 1996-2008 sample for the $A_0(3)$ model. For the parameter vectors in the table above, 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. The inversion method inverts the state vector $X(t)$ from the 3 month, 2 year, and 4 year yields, while 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, $X(t)$.

Yield	Inversion Method	CDGJ(2008) Method	Kalman Filter
$1 - month$	3.2866	3.2855	3.2856
$3 - month$	0	1.479	5.749
$6 - month$	5.1813	15.0795	20.665
$9 - month$	1.7615	5.3561	12.3892
$12 - month$	5.901	4.202	7.966
$2year$	0	19.002	11.168
$3year$	3.7600	2.0834	11.941
$4year$	0	1.771	14.841
$5year$	0.7301	3.1625	13.669
$6year$	3.6998	4.7819	13.815
$7year$	8.2837	6.938	14.080
$8year$	9.5827	8.841	13.344
$9year$	12.3431	10.1150	13.955
$10year$	15.1858	11.1312	14.953