Centre for Computational Finance and Economic Agents

Working Paper Series

WP024-08

Jenny B. Castellanos Pinzon

Estimating Multifactor Uncorrelated Vasicek and CIR Model for the UK Term Structure

September 2008

University of Essex

CCFEA

www.ccfea.net

Estimating Multifactor Uncorrelated Vasicek and CIR Model for the UK Term Structure [*]

Jenny B. Castellanos Pinzon [‡]

Abstract

The objective of this paper is to examine a multifactor version of the Cox-Ingersoll-Ross (CIR) and Vasicek model to characterize the UK nominal term structure of interest rates. The estimation of the model parameters was achieved using a panel data approach that takes into account the characteristics of the short rate model and uses the available information contained in the observable term structure. The approach was implemented using the Kalman filter combined with a (quasi) maximum likelihood estimation, which not only provided the model parameters but the individual state variables and an estimated time series of the short rate. The main conclusion drawn is that more than one factor is necessary to fit accurately the observed yields. It is demonstrated that for the UK term structure two factors are adequate for both the Vasicek and CIR models.

1. Introduction

For many decades now, researchers and finance practitioners have modelled the term structure. The main goal has been to try to understand the dynamics of the term structure and to create an accurate model that fits the observed yield curve both at the long and short end assuming the usual no arbitrage requirement.

Within the literature, two of the most famous models that characterize the short interest rate are the Vasicek [21] and Cox, Ingersoll, and Ross models (1985, CIR hereafter) [7]. The former is a straight forward Gaussian model, however it allows for negative interest rates, while the latter is a more complex model with non Gaussian distribution but with only positive rates[1]. Other models were later developed such as Ho and Lee [15], Hull and White [16] and Longstaff and Schwartz [18] to introduce stochastic volatility, multifactor models, and time dependence among other features. However the Vasicek and CIR models are still

[*] I would like to acknowledge valuable comments by Dr. Kyriakos Chourdakis, Dr. Nick Constantinou and Dr. Steven Simon.
[‡] Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK, jbcast@essex.ac.uk
[1] As described by Cox-Ingersoll and Ross [6], the short interest rate can reach zero if $\sigma^2 > 2\theta\kappa$.

popular as both models have the advantage of being affine and tractable as well as having a close form solution for the bond price.

Within the framework of a short rate model, different approaches have been used to model and calibrate the term structure. The cross section approach uses bond yields for different maturities at a specific point in time. It estimates at each time period the parameters that fit more closely the model to the actual observed yields. One of the main drawbacks of this method is that there are different values for the same parameters for each time period instead of one set of parameters. Furthermore this approach does not guarantee that the structure of the specific short rate model is followed.

A second approach is the time series method, which fits the stochastic differential equation of the factor (the short rate) to observable proxies (e.g. short term rates). The shortcoming of this approach is that depending on the proxy used, the parameter values differ and the information from the observed yields is not used, making it inconsistent with the no arbitrage argument [see Chatterjee [4]).

As highlighted by authors like Canabarro [3] one factor models when estimated empirically do not fit accurately the real data of observed yield curves. This finding encouraged new research into how to use simple models to fit more closely the observed yields. A new approach was introduced by using multiple factors and a panel approach as implemented by Chen and Scott [5]. This approach avoids the disadvantages of the cross section and time series approaches by taking into account both the dynamics of the model (e.g. CIR or Vasicek) and the observed yields. The actual implementation was done representing the problem in a state space form and by using the Kalman Filter (described later) combined with a maximum likelihood (in the case of CIR a quasi maximum likelihood) approach.

Alternative methodologies have been used in the literature to estimate or model the term structure based on the CIR model. Some examples include the Efficient Method of Moments (Dai and Singleton [10]), the Markov Chain Monte Carlo or the Maximum Likelihood Method (Chen and Scott [6]). All methods used have different disadvantages, some use proxies, others are less efficient or are computationally expensive.

It has been noted in the literature that for the Gaussian case the Kalman Filter combined with a maximum likelihood estimator is tractable, consistent and unbiased. However with non Gaussian problems like the CIR model, the Kalman filter can not give an exact maximum likelihood but instead an approximation is done to generate a quasi maximum likelihood. Lund [19] and Duan and Simonato [11] validated the methodology by showing in Monte Carlo experiments that although there is a bias in the estimators, it is negligible.

Chen and Scott [5] also carried out Monte Carlo simulations to study the properties of the quasi maximum likelihood estimators. They first tested the properties of the

Kalman filter, the results suggested a very small conditional bias. They also checked the quasi maximum likelihood estimators. The results show that there are clear biases in the parameters of speed of mean reversion, long run mean and market price of risk, but no significant bias in the estimate of the volatility and the measurement errors. Furthermore they show that there is no bias in the relevant parameter combinations used in asset pricing. These results imply that the quasi maximum likelihood is an appropriate approach to use.

The Kalman filter has been used in both Vasicek and CIR models estimation, but mainly with US data. De Jong [8], Duan and Simonato [11] and Babbs and Nowman [1], estimated the Vasicek Model. Their implementation differs in the form they gave to the covariance matrix of the measurement errors, for some it is a diagonal matrix while for De Jong it is not restricted to a diagonal matrix. All these studies demonstrated that a single factor model is inadequate to fit to the observable yields. This is shown in the standard deviation of the measurement errors that in some cases went up to 73 basis points. A two factor model was shown to reduce significantly the measurement errors.

Babbs and Nowman [1] also considered a three factor model. They found that although the improvement is not as significant as when passing from the one factor to the two factor, the three factor model is preferred over the two factor model when carrying out formal test like the BIC criterion or the Likelihood ratio test. However they conclude that the measurement errors are quite similar. Furthermore these same authors estimated a generalized Vasicek term structure model for the UK gilted edged market for the period 1982 -1996. Their main conclusion was that a two factor model is required. However they only estimated one and two factor models so there is no evidence if a three factor model will be preferred over the two factor one.

Among the authors that have implemented the CIR multifactor model, there is Chen and Scott [5], De Jong [8], Duan and Simonato [11], Geyer and Pichler [12] and De Jong and Santa Clara [9], mainly US data is used. Their findings show that for the one factor CIR model the fit is poor with measurement errors of up to 100 basis points in Geyer and Pichler [12]. The two factor model has much lower errors. Geyer and Pichler estimated up to 5 factors, however they concluded that the fourth and fifth factors do not significantly improve the fit, and a 3 factor model is sufficient. This is the approach taken in this paper.

Finally a different modelling strategy was employed by Gong and Remolona [14], they used a two factor model, but instead of taking the whole yield curve, it was divided into a long and a short end. Their results show that the two factors differ in the mean reversion speed. This result is consistent with the findings of the above mentioned authors in Vasicek and CIR implementations.

The aim of this paper is to examine the effect of the inclusion of up to three factors in the CIR uncorrelated model for UK gilt bond yields data. In addition, and for comparison purposes, the multifactor uncorrelated Vasicek model will also be estimated. The question is whether the inclusion of more factors actually gives a better fit to the UK yield curve, as all of the studies have been made in US data, except Nath and Nowman [20], who only consider up to two factors without a comparison between the uncorrelated versions of the CIR and Vasicek models. The paper will examine whether the effort of using a more complex model, the CIR over Vasicek and the multifactor over the one factor model really gives a better fit to the observed UK yield curve data from January 1985 to August 2007.

The rest of this paper is organized as follows: Section 2 presents the CIR and Vasicek models. In Section 3 the state space formulation and the parameter estimation is described. Section 4 consists of the empirical results, and the last section concludes. The appendix gives a detailed description of the steps for the implementation of the Kalman filter.

2. Theoretical Framework

2.1 Term Structure

It can be said that there are mainly four general theories that try to explain the term structure. The expectations, liquidity preference, market segmentation and the arbitrage free pricing theory (see Cairns [2]). The latter one is the more commonly used as bonds are priced such that there is no room for making a self-financed risk-less profit.

With the fourth theory in mind a common group of models used to characterize the term structure are the so-called affine models. Bond prices are linearly related to underlying state variables. Different states create different bond prices. Therefore the dynamic of the term structure depends on the evolution of the state variables. Given that the instantaneous rate is known it is sufficient information to be able to characterize the whole term structure as it is shown in the following relationship. The price $P(t,T)$ at time *t* of a discount bond that matures at T is generally expressed as:

$$P(t,T) = e^{A(t,T) - B(t,T)r(t)}, \quad (1)$$

where $r(t)$ represents the short rate and $A(t,T)$ and $B(t,T)$ have different functional forms for the different models (Vasicek and CIR) as will be shown in the following section.

2.2 Vasicek and CIR Models

As many other key variables, interest rates are modelled in continuous time as stochastic processes that are characterized by a deterministic and a random part. As this paper focuses in the Vasicek and CIR Models this section shows their formulations. Both models are represented in the following SDEs. The first corresponds to the Vasicek [21] model while the second represents the CIR [6] model.

$$dr = \kappa(\theta - r)dt + \sigma dW(t), \quad (2)$$

$$dr = \kappa(\theta - r)dt + \sigma \sqrt{r} dW(t), \quad (3)$$

κ = the speed at which the short rate reverts back to it long run mean.

θ = risk neutral long run mean

σ = volatility of the short interest rate

W = is a Wiener process

As can be seen the two models only differ in the diffusion term. The form of the CIR diffusion term restricts the resulting rates to be positive and only when $2 * \kappa * \theta > \sigma^2$ does not hold rates can become zero. On the contrary the Vasicek model can produce negative rates. Additionally both models are mean reverting and have a simple formula to calculate bond prices.

As previously stated both models belong to the affine type and the functional forms for $A(t,T)$ and $B(t,T)$ are shown below. Throughout the paper the first expression corresponds to the Vasicek model while the second corresponds to the CIR model. The market price of risk λ can also be estimated using the Kalman Filter. Including this variable in the model for the short rate changes equation (3) to the following expression which denotes the risk neutral dynamics for the CIR model:

$$dr = (\kappa\theta - (\kappa + \lambda)r)dt + \sigma \sqrt{r} dW \quad (4)$$

Including the market price of risk λ , the functional forms for the pricing of the bonds is shown in the following equations for both the Vasicek and CIR models. For the Vasicek model the following formulas are such that λ is negative. For simplicity in the notation *(t,T)* used in previous equations is change to (τ) denoting time to maturity *(T- t)*.

$$A(\tau) = \begin{cases} \dfrac{\gamma\left(B(\tau) - \tau\right)}{\kappa^2} - \dfrac{\sigma^2 B(\tau)^2}{4\kappa} \\ \ln\left[\dfrac{2\gamma e^{\frac{1}{2}(\kappa + \lambda + \gamma)(\tau)}}{2\gamma + (\kappa + \lambda + \gamma)(e^{\gamma(\tau)} - 1)}\right]^{\frac{2\kappa\theta}{\sigma^2}} \end{cases} \tag{5}$$

$$B(\tau) = \begin{cases} \dfrac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right) \\ \dfrac{2(e^{\gamma(\tau)} - 1)}{2\gamma + (\kappa + \lambda + \gamma)(e^{\gamma(\tau)} - 1)} \end{cases} \tag{6}$$

$$\gamma = \begin{cases} \kappa^2\left(\theta - \dfrac{\sigma\lambda}{\kappa}\right) - \dfrac{\sigma^2}{2} \\ \sqrt{(\kappa + \lambda)^2 + 2\sigma^2} \end{cases} \tag{7}$$

3. Multifactor Interest Rate Model

3.1 State Space Model formulation

The Kalman Filter methodology requires a transition and measurement equation. To incorporate this into the Vasicek and CIR models it is necessary to put them in a state space form. The following section outlines this for the CIR model (the Vasicek model has an analogous structure) following the formulation of the multifactor model employed by Chen and Scott [5].

The first assumption made is that the instantaneous nominal interest rate is the sum of *J* state variables

$$r = \sum_{j=1}^{J} y_j \tag{8}$$

Each state variable is assumed to be independent, there is no correlation between the individual Wiener processes. Each state follows the same process form, in the case of the CIR model a square root diffusion process,

$$dy_j = \kappa_j(\theta_j - y_j)dt + \sigma_j\sqrt{y_j}\,dW_j \quad \text{for } j = 1,\ldots,J \tag{9}$$

and in the Vasicek model an Ornstein-Uhlenbeck process. The only difference between equation (4) and (10) is the *j* subscript that accounts for the multiple factors.

As more factors are included the price at time *t* of a risk free bond with maturity *T* from equation (1) is extended to the following general form as shown by Jamieson [13].

$$P(\tau) = \exp\left\{ \sum_{j=1}^{J} \left(A_j(\tau) - B_j(\tau)y_j\right) \right\} \tag{10}$$

The forms of $A_j(\tau)$ and $B_j(\tau)$ are the same as in the previous section, the only difference with equation (6), (7) and (8) is that every parameter has a subscript denoting the factor, as each factor or state variable has its own set of parameters. For example the form of $A_j(\tau)$ is as follows for the CIR process.

$$A_j(\tau) = \ln\left[\frac{2\gamma_j e^{\frac{1}{2}(\kappa_j + \lambda_j + \gamma_j)(\tau)}}{2\gamma_j + (\kappa_j + \lambda_j + \gamma_j)(e^{\gamma_j(\tau)} - 1)} \right]^{\frac{2\kappa_j\theta_j}{\sigma_j^2}} \tag{11}$$

The yield is observed (yield curve) and after calibrating the short rate model, the actual unobserved short rate can be calculated. To be able to use the Kalman filter a transition and measurement equations are required as shown in the next section.

3.2 Transition and Measurement Equation

Both Vasicek and CIR models are continuous time models. However for the state variable estimation they need to be discretized. The transition equation defines how the state variable evolves through time, it is a recursive expression for the factor in terms of its previous value. As shown in Chen and Scott [5] it can be expressed in the following way

$$y_t = C + Hy_{t-1} + \varepsilon_t \tag{12}$$

In the above equation, y_t represents the state variables while C and H follow from the mean of the transition density of *r(t)* over a discrete time interval. The details of the derivation can be seen in Jameison [13]. For both Vasicek and CIR models the expressions are the same.

$$C = \theta_j (1 - e^{-\kappa_j \Delta t})$$
$$H = e^{-\kappa_j \Delta t} \tag{12.1}$$

To calculate the values of C and H as shown in the above equations one must specify the time step of the discretization Δ*t*. If the data observed is collected weekly for example, then the time step will be 1/52. The noise term in the transition equation follows a certain distribution given the information set given at time *t-1*.

$$\varepsilon_{jt} \sim N(0, Q_t) \qquad Q_t = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix} \tag{13}$$

.
The diagonal Q matrix has different forms for the Vasicek and CIR models. The elements of Q represent the variances of the transition densities of the factors. For Vasicek these are constant. On the contrary, for the CIR model these values depend on the previous estimate of the factor. The two different forms of Q are shown below, the first corresponding to the Vasicek model and the second to the CIR model.

$$q_j = \begin{cases} \dfrac{\sigma_j^2}{2\kappa_j}(1 - e^{-\kappa_j \Delta t}) \\ \dfrac{\theta_j \sigma_j^2}{2\kappa_j}(1 - e^{-\kappa_j \Delta t})^2 + \dfrac{\sigma_j^2}{\kappa_j}(e^{-\kappa_j \Delta t} - e^{-2\kappa_j \Delta t})\hat{y}_{jt-1} \end{cases} \tag{13.1}$$

The second equation necessary for the implementation of the Kalman filter is the measurement equation. It is build from the relation expressed in equation (2). It describes the interaction between the observable variable (the yields) and the unobservable variable (the short rate).

$$Z_t = A + By_t + V_t \tag{14}$$

Z_t represents the continuously compounded yields extracted from the yield curve. A and B are constructed from the bond pricing formula discussed before. The additional term V_t is a noise term. This term reflects small measurement errors in the bonds prices. The addition of this term has been justified by many authors. Chen and Scott [5] that have identified several sources of errors such as the calculation of bond prices from averages of bid and ask prices or coupon bonds, non synchronous trading and rounding of prices. For the purpose of this paper the error terms are assumed to be uncorrelated across different maturities as in Chen and Scott [5].

As such, $V_t \sim N(0, R)$, where R is a diagonal matrix with as many diagonal elements as maturities. These elements are part of the set of parameters that are estimated with the Kalman filter and the maximum likelihood method. The values of the errors are an indication of the goodness of fit of the model. If the model gives a perfect fit with the observed yields then these errors should be zero.

For illustration purposes the transition equation for four different maturities and three factors, is as follows:

$$
\begin{bmatrix} z(\tau_1) \\ z(\tau_2) \\ z(\tau_3) \\ z(\tau_4) \end{bmatrix} = \underbrace{\begin{bmatrix} \sum_{i=1}^{3} \frac{-A_i(\tau_1)}{\tau_1} \\ \sum_{i=1}^{3} \frac{-A_i(\tau_2)}{\tau_2} \\ \sum_{i=1}^{3} \frac{-A_i(\tau_3)}{\tau_3} \\ \sum_{i=1}^{3} \frac{-A_i(\tau_4)}{\tau_4} \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} & \frac{B_2(\tau_1)}{\tau_1} & \frac{B_3(\tau_1)}{\tau_1} \\ \frac{B_1(\tau_2)}{\tau_2} & \frac{B_2(\tau_2)}{\tau_2} & \frac{B_3(\tau_2)}{\tau_2} \\ \frac{B_1(\tau_3)}{\tau_3} & \frac{B_2(\tau_3)}{\tau_3} & \frac{B_3(\tau_3)}{\tau_3} \\ \frac{B_1(\tau_4)}{\tau_4} & \frac{B_2(\tau_4)}{\tau_4} & \frac{B_3(\tau_4)}{\tau_4} \end{bmatrix}}_{B} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \\ v_{4t} \end{bmatrix} \tag{15}
$$

3.3 Parameter estimation using Genetic Algorithm for the Kalman Filter

The parameters for the multifactor models were estimated following Chen and Scott [5] using the Kalman filter methodology with (quasi) maximum likelihood estimation. The appendix reviews in detail the different steps to implement the Kalman filter and the likelihood function.

The implementation of the Kalman Filter is straight forward, however the optimization step to find the maximum likelihood estimators, needs to be discussed further. Initial values for the different parameters are required for the optimization. There are several approaches to this issue. Geyer and Pichler [12] for example based their initial values on 500*J (J the number of factors) random samples, for each parameter, taken from a reasonable range. Their starting values were those that initially maximize the log likelihood. Another possibility is to simply choose what could be a good initial guess. However this requires that the program is run several times with different starting points to assure the global maximum is achieved.

In the case of this implementation there is no need to put any initial values as they were generated using a Genetic Algorithm (GA). Briefly GA is a heuristic method used to search a space. The idea of the GA is that as new generations are created there is evolution towards better solutions until the best is reached. Generally the process is as follows. It starts with the random generation of an initial population of candidate solutions. All different possible solutions are evaluated using the fitness function. The generation of new solutions is done by selecting part of the population to “breed”, usually among this are the fittest solutions (there are different methods to generate this selection). The selected population “mates” using crossover and mutation. The new potential population has therefore characteristics of its ‘parents’ plus some small modifications. This is done generation after generation. Generally the average fitness value increases as mainly only fitter solutions are selected to mate. For this paper the fitness function is the likelihood function. The GA finishes when a halting criterion is reached. There can be many different criteria. In this paper the criterion was that the fitness function did not have a significant increase.

The values given by the GA were then used as the initial values of the parameters for the final optimization. The process was done several times for each model to assure the maximum was found. From the different results it was very obvious that the surface in which the search takes place is very rough and there is an additional complication as different combinations of parameters could give very similar likelihood values especially for the CIR case. This makes the search problem even tougher, the use of a GA initially is a very good approach as it approximates the solution to a more smooth area were traditional search methods can be used. It also has to be noted that the optimization restricted the values of some of the parameters. For example the volatility and long run mean were limited to be positive values.

4. Empirical Results and Analysis

4.1 Data

The data used for this paper consists of zero coupon yield curves obtained from the Bank of England website from their Statistics Section. The Bank calculates the zero coupon yields for the government bonds daily. For this analysis, weekly data from 16 January 1985 to 16 August 2007 was used. Yields for 4 maturities were collected weekly making the time step needed for the transition equation equal to 1/52. To avoid the weekday effects or bank holidays, Wednesdays were chosen.

The four maturities were chosen so that the term structure could be characterized in a shorter and longer end. The complete data set was not available for certain maturities, so the closest maturities to the desire ones were chosen (1.5, 5, 10 and 19 years) to ensure a sufficient number of observations. The total number of observations was 1179 per maturity. The following table shows the statistical properties of the data.

UK Data Summary Statistics January 1985 - August 2007

Maturity (years)	Mean (%)	Standard Deviation
1.5	7.166	2.639
5.0	7.276	2.380
10	7.277	2.349
19	6.912	2.083

Table 1: Main statistics for the zero coupon UK yields.

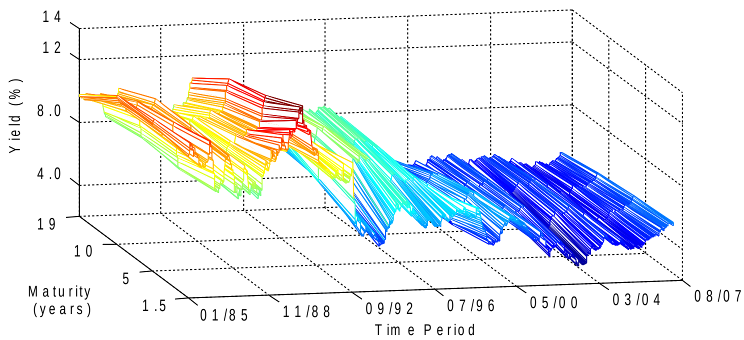


Figure 1: UK Term Structure from January 1985 to August 2007.

4.2 CIR Results

The results for the CIR model are shown in Table (2). It not only includes the estimated parameters and its standard errors (in brackets)[2], but also the likelihood value, the root mean square error (RMSE), the mean error (ME), the mean absolute error (MAE) and the factor loadings.

The mentioned errors were calculated using the filtered yields and the actual observed ones. The filtered yields were estimated by using the parameters given by the maximum likelihood estimation and using the filtered factors in the measurement equation. The factor loadings correspond to the coefficients of the state variables or to the elements of matrix B (see measurement equation (15)).

Several aspects can be noted from the results. The likelihood function increases as more factors are added, however the increase from one factor to two factors is much larger than when passing from two factors to three. In the one factor case the standard deviations of the measurement errors are as high as 71 basis points for the last maturity. However it seems that for the second maturity (5 years) the errors are very small. If the errors for all the maturities are combined they sum up to more than 100 basis points, which in terms of pricing derivatives is significant. Therefore, the one factor model fails to fit accurately the observed UK yields, and this is consistent with the various investigations on US data.

In general except for the 5 year maturity the errors decrease as more factors are added. In the two factor model the highest error is 30 basis points, with a significant decrease in the first and third maturity. As for the three factor model

[2] The standard errors were calculated using the Hessian Matrix as suggested in Jamieson Bolder D.(2001) [13]

CIR Model						
	One Factor	***Two***	***Factors***	***Three***	***Factors***	
θ	0.06814	0.02466	0.01341	0.02149	0.02146	0.03514
	[0.02885]	[0.00024]	[0.00024]	[0.00689]	[0.0006]	[0.00571]
σ	0.04184	0.07561	0.08299	0.04080	0.08476	0.15307
	[0.00082]	[0.00155]	[0.00136]	[0.00156]	[0.00253]	[0.00170]
κ	0.01889	0.60349	0.00014	0.00010	0.48973	0.00010
	[0.00807]	[0.00287]	[0.00027]	[0.00039]	[0.01234]	[0.00022]
λ	-0.00687	-0.22274	-0.05351	-0.06122	-0.05303	-0.16044
	[0.00803]	[0.00022]	[0.00046]	[0.00138]	[0.01094]	[0.00355]
r_1	0.00670	0.00048		0.00048		
	[0.00013]	[0.00008]		[0.00004]		
r_2	0.00001	0.00171		0.00165		
	[0.00007]	[0.00006]		[0.00004]		
r_3	0.00372	0.00034		0.00001		
	[0.00006]	[0.00008]		[0.00002]		
r_4	0.00717	0.00308		0.00021		
	[0.00015]	[0.00012]		[0.00002]		
Lik	19355	22989		24745		
BIC	-38642	-45876		-49355		
RMSE						
	0.00669	0.00023		0.00023		
	<0.00001	0.00169		0.00165		
	0.00372	0.00014		0.00002		
	0.00716	0.00307		0.00010		
ME						
	-0.00084	<0.00001		<0.00001		
	<0.00001	0.00036		0.00038		
	0.00041	-0.00001		<0.00001		
	-0.00086	-0.00028		<0.00001		
MAE						
	0.00500	0.00014		0.00012		
	<0.00001	0.00124		0.00128		
	0.00291	0.00010		<0.00001		
	0.00540	0.00242		0.00006		
Factor	0.99040	0.76062	1.03833	1.04659	0.73223	1.11943
Loading	0.96374	0.44281	1.10993	1.16026	0.40218	1.34057
	0.91720	0.25263	1.15213	1.32874	0.22236	1.35024
	0.81866	0.13553	1.05352	1.60418	0.11831	0.93739

Table 2: CIR Model results for the one, two and three factor model.

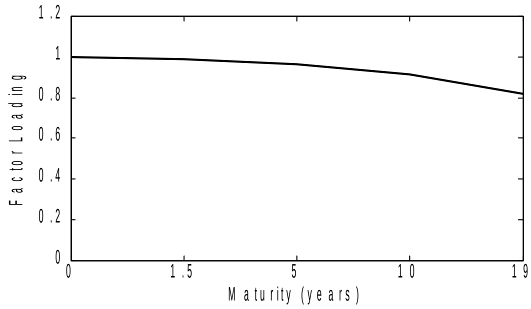
although the likelihood increased the errors are not dramatically better than the two factor model. For the first maturity they stay the same, for the second maturity they are of similar level, it is for the long end that the additional factor seems to have a bigger impact, specifically the last maturity as the error decreases from 30 to 2 basis points.

The above pattern can also be seen in the RMSE, ME and MAE estimations. In general they decrease as more factors are added and are quite small in magnitude, validating this methodology as quite successful in fitting the term structure, (in sample sense). The combined RMSE decrease from 0.017 in the one factor to 0.0051 and 0.0019 in the other two factors. It can be suggested given these changes, that the third factor contributes in a much lower magnitude than the second factor.

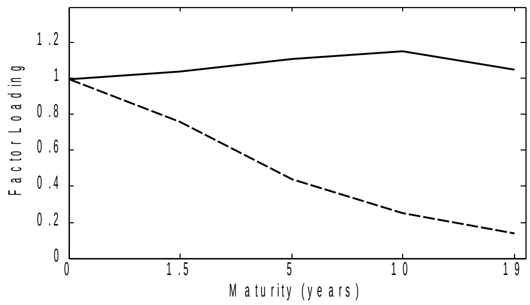
As for the estimated parameters, it can be seen that the long run mean is always around 7% as it would be expected given the statistics of the yields used (Table (1)). When more factors are added there is one factor that has a very predominant speed, as other authors have found. The remaining factors on the contrary have a very low speed. It can be seen from both the two and three factor kappa parameters that the speed of the additional factors is very slow. As for the market price of risk all estimates are negative.

At this point it is important to mention that in the estimation of the parameters there could be an identification problem. Different parameters can give very similar likelihood values. This is specifically the case for the long run mean, the reversion speed and the market price of risk. The volatility and measurement errors do not present this problem. Geyer and Pichler [12] suggested that it is possible to improve the parameter identification if more maturities are added; on the other hand Chen and Scott [4] found that this is not the case. However it has to be taken into account that the combinations of these parameters that are relevant for pricing are always the same and identifiable.

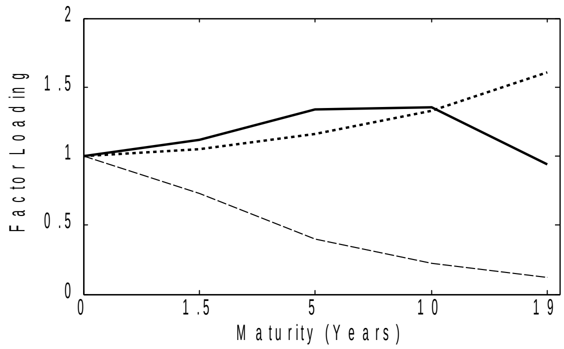
Several authors such as Litterman et al. [17] have named the different factor loadings as level, slope and curvature. To appreciate better the factor loadings from the estimation, they are plotted in the following graphs. As authors like Chen and Scott [5] and Pichler and Geyer [12] have found the loading factors are all positive.



(a) One Factor CIR



(b) Two Factor CIR



(c) Three Factor CIR

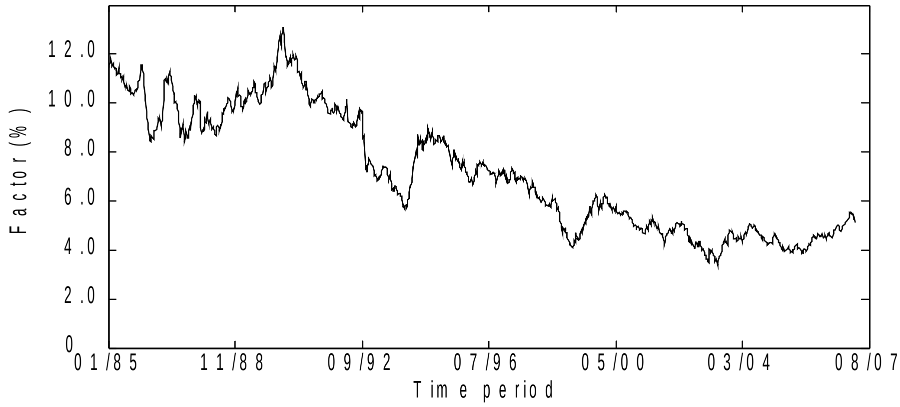
Figure 2: CIR Factor Loadings. The factor loadings are the elements of the B matrix in the measurement equation (15). The continuous line represents the first factor, the dashed line represents the second factor and the dotted line represents the third factor.

The above graphs are consistent with the results of one and two factor models of Chen and Scott [5] and Geyer and Pichler [12]. The first factor is associated with the level of the interest rate, while the second factor is associated with the slope. The third factor however differs from the standard definition (the curvature). Instead it seems to be a factor that affects mainly long term rates.

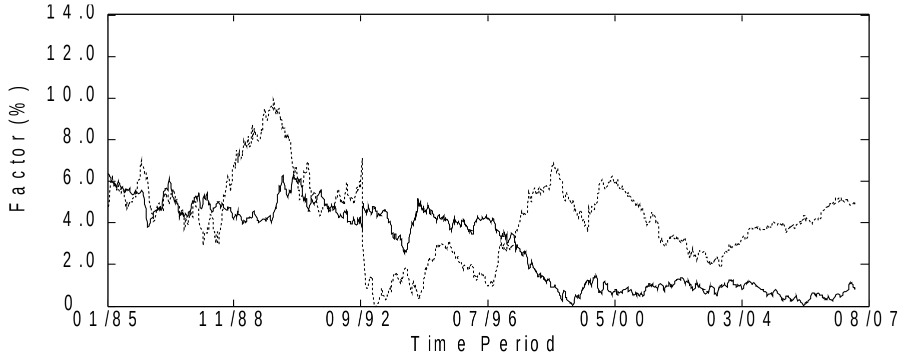
Combining the results for the measurement errors where the inclusion of a third factor did not significantly reduce the errors (it only did for the last maturity), it might be that the model is trying to fit the additional factor as an adjustment for the long end of the curve. If the first and third factor are averaged, the end result is a very similar estimate of the first factor in the two factor model, just that the long end has a higher loading. It might be that for the UK term structure two factors are sufficient and the third factor mainly affects the long end of the term structure.

The different estimated factors are plotted in figure 3. From panel (b) and (c) it can be observed that the second factor (dotted line) shows the highest volatility. Given that the factor loading for this factor is greater for the short end of the term structure, it can be said that the second factor accounts for most of the variation in this end.

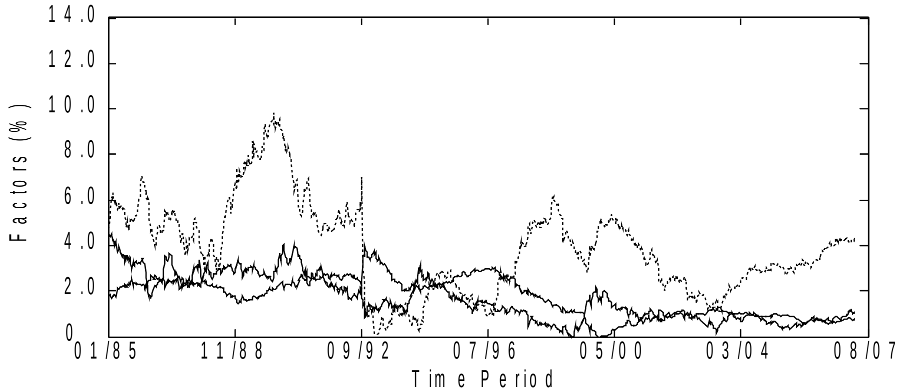
All the plots show the effect of Black Wednesday, this corresponds to September 16, 1992. It can be seen that on that day the interest rates increased as a response to the increase by the government of the interest rates from 10% to 12% and 15%, only to be lowered a day later back to 12%. This event affected mostly the short end of the curve and in the figure for the two and three factor models, it can be seen that this increase-fall period is reflected mainly in the second factor. The others do not react significantly while the second one has a sharp drop.



(a) One Factor - CIR model



(b) Two Factors - CIR Model



(c) Three factors – CIR Model

Figure 3: Individual Factors from the one, two and three CIR model

4.3Vasicek Results

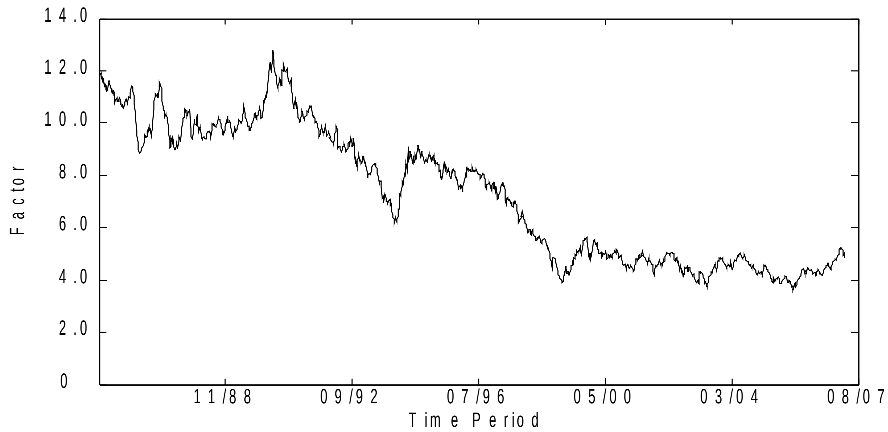
The following table shows the results for the Vasicek Model for one to three factors.

Vasicek Model						
	One Factor	Two	Factors	Three	Factors	
θ	0.08159	0.04416	0.01392	0.00010	0.02930	0.00278
	[0.09448]	[6.73131]	[6.65993]	[14.06975]	[29.72861]	[34.40225]
σ	0.01131	0.01509	0.01010	0.01757	0.02696	0.00988
	[0.00043]	[0.00040]	[0.00029]	[0.00055]	[0.00186]	[0.00047]
κ	0.01808	0.37354	0.02024	0.15838	1.81923	0.01631
	[0.00161]	[0.01261]	[0.00150]	[0.00587]	[0.14331]	[0.00322]
λ	-0.01646	-0.17876	-0.00001	-0.11638	-0.00010	-0.00010
	[0.15267]	[0.19340]	[0.16878]	[0.16942]	[0.21134]	[0.09692]
r_1	0.01001	0.00001		0.00001		
	[0.00021]	[0.00018]		[0.00005]		
r_2	0.00389	0.00181		0.00022		
	[0.00009]	[0.00005]		[0.00002]		
r_3	0.00000	0.00059		0.00173		
	[0.00006]	[0.00004]		[0.00005]		
r_4	0.00452	0.00314		0.00048		
	[0.00011]	[0.00009]		[0.00003]		
Lik	19438	22973		24766		
BIC	-38808	-45844		-49397		
RMSE						
	0.01001	<0.00001		<0.00001		
	0.00389	0.00178		0.00010		
	<0.00001	0.00042		0.00171		
	0.00452	0.00314		0.00037		
ME						
	-0.00157	<0.00001		<0.00001		
	-0.00059	-0.00034		<0.00001		
	<0.00001	<0.00001		0.00001		
	-0.00080	-0.00029		<0.00001		
MAEZ						
	0.00776	<0.00001		<0.00001		
	0.00301	0.00140		0.00007		
	<0.00001	0.00030		0.00142		
	0.00337	0.00245		0.00023		

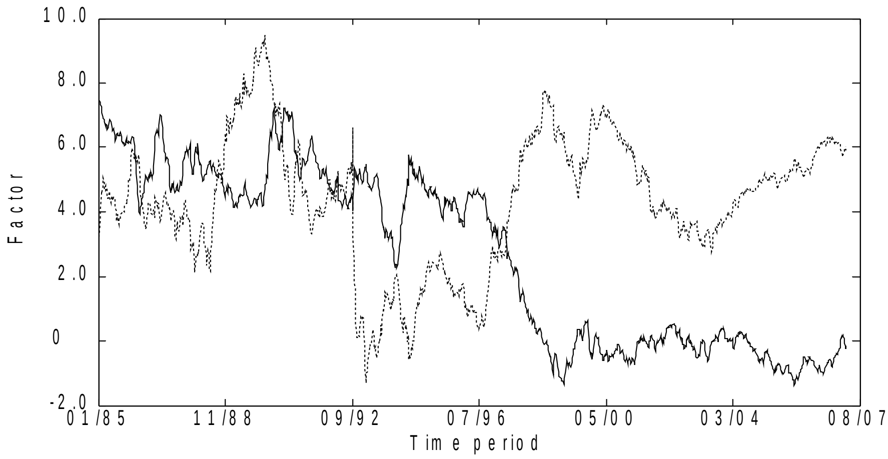
Table 3. Vasicek Model Estimated Parameters

The individual factors that correspond to the parameters shown in table 3 are plotted in figure 4. It can be seen that as the restriction of non-negativity does not apply for the Vasicek model, the rates become negative (especially in the three factor model). However when adding up the individual factors in the two and three factor model, the final process does not have negative values as shown in panel (c). From Table 3 it can be seen that the one factor is not enough to characterize the term structure. The measurement errors are up to 100 basis points, the inclusion of more factors reduced these errors.

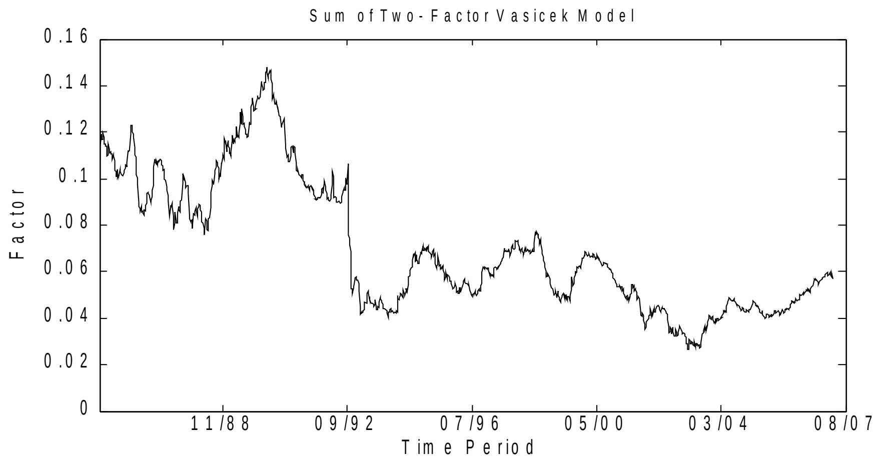
As for the comparison with the CIR model the likelihood values are very similar for the two models. Looking at the errors it seems that the CIR does slightly better. Undertaking the same estimation as in the CIR model, the combined RMSE errors for each factor are as follows 0.018, 0.0053 and 0.0021. It can be seen that these values are very similar to the CIR ones. In this sense is not clear what model is better. However with the Vasicek model more parameters are not significant.



(a) One Factor – Vasicek Model



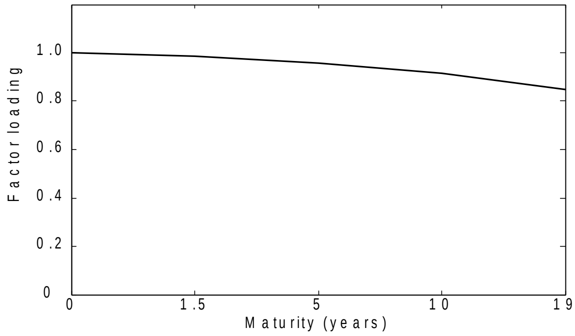
(b) Two Factors – Vasicek Model



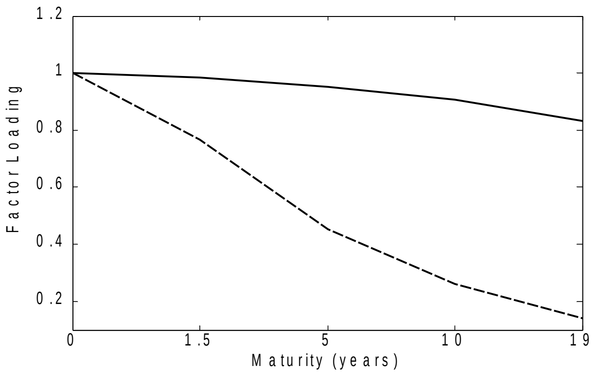
(c) Sum of Two Factor Vasicek Model

Figure 4. Vasicek Model Individual Factors. Figure (a) and (b) show the individual factors while figure (c) shows the sum of the Two factor model.

The factor loadings for the Vasicek Model, shown in figure 5 are in line with the findings for the CIR model. The first factor is related to the level while the second factor corresponds to the slope and affects mainly the short end of the curve.



(a) One Factor loading – Vasicek Model



(b) Two Factor loading – Vasicek Model

Figure 5. Factor Loadings for the Vasicek Model. The continuous line denotes the first factor while the dashed line corresponds to the second factor loading.

4.4 Monte Carlo Simulation

As part of the methodology validation a Monte Carlo simulation for the CIR one factor model was carried out. Using an approximation to the chi-square distribution, the yields were simulated by first simulating the short rate and then using the measurement equation (see equation (15)) to transform the rate into yields. The time step used for the simulation was 1 day, 30000 days were simulated each time, 50 simulations, however the sampling for the Kalman filter was done weekly, this was done so that the chi square distribution could be better approximated. The Kalman filter was used to recover the parameter values used in the simulation. The following table shows the parameter values given for the simulation and the mean and standard deviation of the recovered parameters from the quasi maximum likelihood estimation.

Monte Carlo CIR 1 factor Results										
	Theta	Sigma	Kappa	Lambda	r1	r2	r3	r4	κ*θ	κ+λ
Parameter Value	0.062	0.04	0.02	-0.01	0.0002	0.0001	0.0001	0.0002	0.00124	0.01
Mean	0.1120	0.0400	0.0462	-0.0362	0.000197	2.64E-05	2.64E-05	0.00011	0.00124	0.00999
StDev	0.2053	0.0001	0.0406	0.0405	3.02E-06	3.64E-05	3.64E-05	5.9E-06	6.3E-06	3E-05

Table 4. Monte Carlo One Factor CIR Results

The results coincide with Chen and Scott [4] results. They also did a Monte Carlo study with a two factor CIR model. The conclusions were the same that can be drawn from the above table. The theta, kappa and lambda estimators are noisy, it can be seen for example that theta has a standard deviation of 20% how ever the relevant parameters for asset pricing are quite accurately estimated as well as sigma and the measurement errors. This results support once again the use of the present tool.

5. Conclusions and Further research

The results from this study show that one factor models are insufficient to characterize the UK term structure to a high level of accuracy, the errors are larger than 100 basis points. The accuracy of the fit can be improved by using more factors. The inclusion of a second factor improves the fit greatly, much more than the inclusion of a third factor. The errors decreased by more than 110 basis points while the improvement of the third factor was of 30 basis points.

The first and third factor relate to the medium and long end of the yield curve while the second factor clearly accounts for the changes in the shorter end. The difference in interpretation of the third factor with the standard findings in US data may suggest that for the UK case two factors are sufficient and the third factor mainly improves the fit for the longer end of the yield curve.

As for the comparison of the Vasicek and CIR uncorrelated case, the results show that both models are comparable. The likelihood values were very similar as the measurement errors and other errors calculated. The clear disadvantage with the Vasicek model is the negative rates of the individual factors for the two and three factor model. However when the factors were added no negative rates were present. The latter suggest that both models can be used for the purpose of fitting the term structure how ever with the Vasicek model there is no guarantee that the sum of the factors will not result in a negative rate.

Finally the Kalman Filter methodology is a good approach to this type of problem as the short rate is treated as an unobservable variable and there is no need to use any proxies. As well the Kalman filter has proved to be a powerful tool in identifying the underlying factors and one of its advantages is that it is relatively simple to implement. The quasi maximum estimators have been proven by different authors to generate a good approximation and have acceptable biases especially as the relevant combinations for asset pricing seem not to show these biases. How ever care must be taken in the optimization, there is an identification problem and the search space is quite rough with several parameter combinations outputting very similar likelihood values.

For future research the same methodology can be applied to different problems for example credit spreads. This paper uses government bonds, that are considered default free instruments, but the same methodology can be applied to identify the default probabilities which are unobservable using the spreads. As well a natural extension is to allow for correlation in the models, this may imply an even better fit to the term structure.

6. References

[1] Babbs, S. H., and K. B. Nowman (1999). Kalman Filtering of Generalized Vasicek Term Structure Models. Journal of Financial and Quantitative Analysis, pp. 115-130.

[2] Cairns Andrew J.G. Interest Rate Models An Introduction. Princeton University Press, 2004.

[3] Canabarro, E., (1995). Where do one factor interest rate models fail?, The Journal of Fixed Income, September 1995, 31-52

[4] Chatterjee, S. (2005). Application of the Kalman for Estimating Continuous Time Term Structure Models: The Case of UK and Germany. Working Papers. University of Glasgow.

[5] Chen, R., and L. Scott (1995). Multi-factor Cox-Ingersoll-Ross models of the term-structure: estimates and tests from a Kalman Filter model, Working paper, University of Georgia.

[6] Chen, R., and L. Scott (1993). Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. Journal of Fixed Income, pp. 14 – 31.

[7] Cox J.C., J.E. Ingersoll and S.A. Ross, (1985). A Theory of the Term Structure of Interest Rates. Econometrica, 53 pp. 385-407.

[8] De Jong, F. (1998). Time-Series and Cross-Section Information in Affine Term Structure Models. Journal of Business & Economic Statistics, Vol. 18, No. 3. (Jul., 2000), pp. 300-314.

[9] De Jong, F. and Santa-Clara, P. The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables. The Journal of Financial and Quantitative Analysis, Vol. 34, No. 1. (Mar., 1999), pp. 131-157.

[10] Dai, Q., and K.J. Singleton (2000). Specification Analysis of Affine Term Structure Models. Journal of Finance 55,pp 1943-78.

[11] Duan, J.-C., and J.-G. Simonato (1995). Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter. Discussion paper, Centre universitaire de recherche et analyze des organizations (CIRANO).

[12] Geyer, A. L. J., and S. Pichler (1998). A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates, Department of Operations Research, University of Economics Vienna.

[13] Jamieson Bolder D. (2001). Affine Term-Structure Models: Theory and Implementation. Bank of Canada Working Paper 2001-15.

[14] Gong F.F and Remolona E. M. (1997b). Two Factors along the yield curve. Manchester School supplement.1-31

[15] Ho, T.S.Y, and Lee, S.B. (1986). Term Structure Movements and Pricing Interest Rate Contingent Claims. Journal of Finance, 41,1011-1029.

[16] Hull, J and White, A. (1990). Pricing Interest Rate Derivative Securities. Review of financial studies, 3,573-92.

[17] Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns, Journal of Fixed Income (June): 54-61

[18] Longstaff, F. A., and E. S. Schwartz (1992a). Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, The Journal of Finance, XLVII, 1259-1282.

[19] Lund, J. (1997). "Econometric Analysis of Continuous-Time Arbitrage-Free models of the Term Structure of Interest Rates", Department of Finance, The Aarhus School of Business.

[20] Nath, P. and K. Ben Nowman, (2001). "Estimates of the continuous time Cox-Ingersoll-Ross term structure model: further results for the UK gilt-edged market", Applied Economic Letters, 2001, 8, 85-88.

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

7. Appendix

The Kalman Filter in Detail

The Kalman Filter is a tool that was originated in the field of engineering, however it has proven to have useful applications in finance such as the one described in this paper. The Kalman filter is an optimum estimator in the minimum mean square error sense. It estimates the state variable based on the dynamics and relationship of the unobserved variable with an observable variable.

The Kalman filter does not estimate the parameters of the models, but gives all the necessary out puts to estimate them through quasi or maximum likelihood estimation. Using this methodology has several advantages. First it treats the short rate as an unobserved variable which indeed it is, avoiding the use of proxies that usually alter the results. Secondly not only the parameters are estimated but the factors can as well be estimated.

The filter works as a recursion of equations. The initial estimator of the state variable is optimal in the sense that it uses the previous estimated values and the conditional distribution of the unobserved variable. As new information is available the optimal estimate is updated as to minimize the mean square error, its covariance matrix (in this paper denoted by P) is also updated,.

To better understand the Kalman filter the following diagram shows the recursive equations as specified in many different papers such as Lund [19].

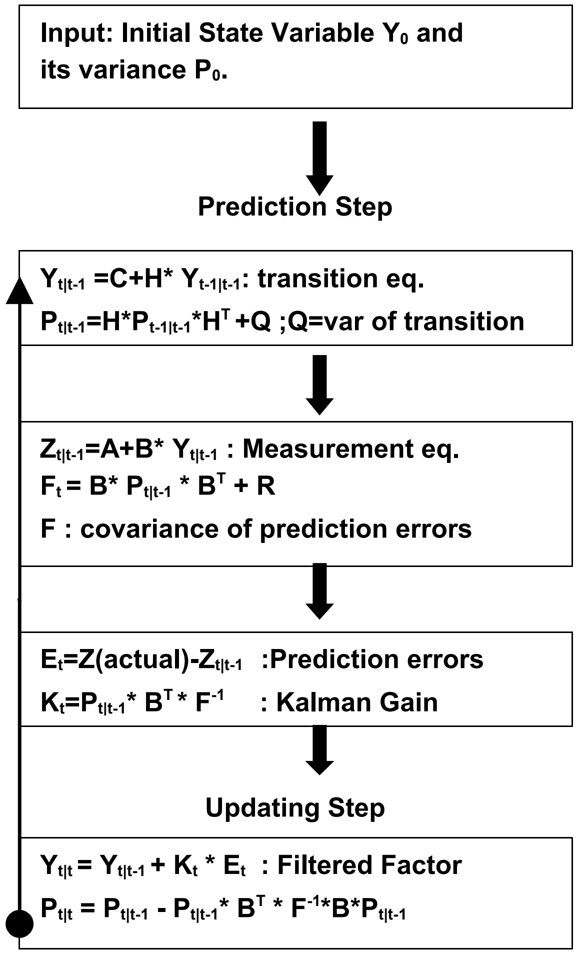


Figure 7. Kalman Filter Flow Chart

Step 1 Initialize

To start the Kalman Filter algorithm apart from the parameters of the model and the elements of R, it is necessary to give initial values to the state variables and to its variances which are denoted as *P*. For the initial values of the factors the long run mean are used as they are the unconditional means. The unconditional variance for the CIR model is shown below for a three factor model.

CIR Vasicek

$$y_0 = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \quad P_0 = \begin{bmatrix} \frac{\sigma_1^2 \theta_1}{2\kappa_1} & 0 & 0 \\ 0 & \frac{\sigma_2^2 \theta_2}{2\kappa_2} & 0 \\ 0 & 0 & \frac{\sigma_3^2 \theta_3}{2\kappa_{31}} \end{bmatrix} \qquad P_0 = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1} & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2} & 0 \\ 0 & 0 & \frac{\sigma_3^2}{2\kappa_{31}} \end{bmatrix}$$

<u>Step 2 Forecasting</u>

This step consists of forecasting the state variables and its variances for the next period. This is done using the transition equation. For both Models this is the same. The following matrix form shows the implementation for three factors.

$$\begin{bmatrix} \hat{y}_{1t|t-1} \\ \hat{y}_{2t|t-1} \\ \hat{y}_{3t|t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} \theta_1(1-e^{-\kappa_1 \Delta t}) \\ \theta_2(1-e^{-\kappa_2 \Delta t}) \\ \theta_3(1-e^{-\kappa_3 \Delta t}) \end{bmatrix}}_{C} + \underbrace{\begin{bmatrix} e^{-\kappa_1 \Delta t} & 0 & 0 \\ 0 & e^{-\kappa_2 \Delta t} & 0 \\ 0 & 0 & e^{-\kappa_3 \Delta t} \end{bmatrix}}_{H} \begin{bmatrix} \hat{y}_{1t-1|t-1} \\ \hat{y}_{2t-1|t-1} \\ \hat{y}_{3t-1|t-1} \end{bmatrix}$$

The Covariance matrix is calculated using the previous value, H and adding the conditional variance of the transition system. This Q differs for both Models as in CIR it depends on the previous values of the state variable which is not the case in Vasicek. The specific form of the Q matrix is shown in section 3 in equations (13.1).

$$P_{t|t-1} = H_t * P_{t-1|t-1} * H_t^T + Q_t \qquad (16)$$

Having calculated the values of the *Q* matrix, the next step is to forecast *Z,* the continuously compounded yield, this is done using the measurement equation. The forecast is conditional on the information available up to time
t-1.

$$\begin{bmatrix} \hat{z}_{1t} \\ \hat{z}_{2t} \\ \hat{z}_{3t} \\ \hat{z}_{4t} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} + \begin{bmatrix} & & \\ & B & \\ & & \end{bmatrix} \begin{bmatrix} \hat{y}_{1t|t-1} \\ \hat{y}_{2t|t-1} \\ \hat{y}_{3t|t-1} \end{bmatrix} \quad (17)$$

At this point the variance of *Z* denoted by *F* can also be calculated as shown in equation (19). The dimension of the covariance matrix depends on the number of maturities, for example as this paper uses four maturities the covariance matrix was dimensions 4x4. This matrix is an element of the likelihood function that is used to estimate the parameters of the models.

$$F_t = B_t * P_{t|t-1} * B_t^T + R_t \quad (18)$$

<u>Step 3 Updating</u>

In this step the estimate for the factors is updated as new information becomes available. For the update two more elements are required, the Kalman Gain denoted *K* which is a 3x4 matrix and the estimation errors for *Z*, denoted ζ . The Kalman Gain can be think of as the weights to be given to the prediction errors to constitute the new filtered estimated

$$\begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix} = \begin{bmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \\ z_{4t} \end{bmatrix} - \begin{bmatrix} \hat{z}_{1t} \\ \hat{z}_{2t} \\ \hat{z}_{3t} \\ \hat{z}_{4t} \end{bmatrix} \quad (19)$$

$$K_t = P_{t|t-1} * B_t^T * F_t^{-1} \quad (20)$$

.

The updated estimate of the factor is calculated as follows and it is commonly referred to as the filtered variable.

$$\begin{bmatrix} \hat{y}_{1t|t} \\ \hat{y}_{2t|t} \\ \hat{y}_{3t|t} \end{bmatrix} = \begin{bmatrix} \hat{y}_{1t|t-1} \\ \hat{y}_{2t|t-1} \\ \hat{y}_{3t|t-1} \end{bmatrix} + \begin{bmatrix} & K & \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix} \tag{21}$$

The state variance, *P,* also needs to be updated as new information is available

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} * B_t^T * F^{-1} * B * P_{t|t-1} \tag{22}$$

All steps in the Kalman Filter are the same for both models, however there is an additional restriction mentioned in Chen and Scott that has to be imposed for the CIR process. Given the non negativity constraint that the state variables have, in the updating state if the estimate for any of the state variables is negative it is replaced by zero as suggested by Chen and Scott [4] or the update step is jumped and the estimate is replaced by the previous period filtered variable as suggested by Geyer and Pichler [12].

Step 4 Likelihood Function

As mentioned before the Kalman filter outputs the sufficient elements to generate a maximum likelihood estimation to identify the best parameters. Different authors suggest slightly different likelihood functions, how ever all of them give the same parameters. For this paper the likelihood function showed in Lund [19] is used and it is as follows:

$$LogL = \sum_{t=1}^{n} -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|F_t| - \frac{1}{2}\zeta_t^T F_t^{-1} \zeta_t \tag{23}$$