Parameter Uncertainty in Kalman Filter Estimation of the CIR Term Structure Model

Conall O’Sullivan*

Abstract

The Cox, Ingersoll and Ross (1985) term structure model describes the stochastic evolution of government bond yield curves over time using a square root Ornstein-Uhlenbeck diffusion process, whilst imposing cross-sectional no-arbitrage restrictions between yields of different maturities. A Kalman filter approach can be used to estimate the parameters of the CIR model from panel data consisting of a time series of bonds of different maturities. The parameters are estimated by optimising a quasi log-likelihood function that results from the prediction error decomposition of the Kalman filter. The quasi log-likelihood function is usually optimised with a deterministic gradient based optimisation technique such as a quadratic hill climbing optimiser. This paper uses an evolutionary optimiser known as differential evolution (DE) to optimise over the parameter space. The DE optimiser is more likely to find the global maximum than a deterministic optimiser in the presence of a non-convex objective function which may be the case in multifactor term structure models with non-negativity constraints and parameter constraints. The method is applied to estimate parameters from a one and two-factor Cox, Ingersoll and Ross (1985) model. It is shown that in the two factor model the problem of local maxima arises whereby a number of different parameter vectors perform equally well in the estimation procedure. Fixed income derivative prices are particular sensitive to term structure parameters such as the volatility, the rate of mean reversion, and the market price of risk of each factor. The effect of different optimal parameter vectors on fixed income derivatives is examined and is found to be significant.

1 Introduction

Dynamic term structure models describe the stochastic evolution of government bond yield curves (or swap market yield curves) over time whilst imposing cross-sectional no-arbitrage restrictions between yields of different maturities. The classic dynamic term structure models are usually assumed to be driven by a small number of latent factors motivated by the empirical results of Steeley (1990) and Litterman and Scheinkman (1991) who found that three factors can explain up to 99% of the variability of the yield curve. However modelling yield curves using a low dimensional factor model will lead to measurement error between the theoretical implied yields (or bond prices) and market yields (or bond prices). If the dynamic term structure model (DTSM) can be formulated into a state space representation a Kalman filter approach can be used and this measurement error can be accounted for explicitly. In the Kalman filter approach model parameters are estimated from panel data consisting of a time series of bonds with different maturity dates. Parameter estimation is carried out by optimising a quasi

*University College Dublin, email: conall.osullivan@ucd.ie
log-likelihood function that results from the prediction error decomposition of the Kalman filter. The quasi log-likelihood function is usually optimised with a deterministic gradient based optimisation technique such as a quadratic hill climbing optimiser. This paper uses an evolutionary based optimiser known as differential evolution (DE) to optimise over the parameter space. The DE optimiser is more likely to find the global maximum than deterministic optimisers in the presence of a non-convex objective function which may be the case in term structure models with non-negativity constraints and parameter constraints such as the Cox, Ingersoll and Ross (hereafter CIR) model (1985). The term structure models used in this study are restricted to the one and two-factor versions of the CIR (1985) model for simplicity. However DE could potentially be even more useful than deterministic optimisers in more complex DTSMs such as those preferred by Dai and Singleton (2000). The data set consists of Fama-Bliss US government bond yields of 3, 6, 12 and 60 months maturities sampled monthly from April 1964 to December 1997. Duan and Simonato (1999) have used the same data to estimate a one-factor Vasicek model and one and two-factor CIR models using a deterministic optimiser. We find that the optimal parameter vectors using the DE optimiser agree with Duan and Simonato for the one-factor CIR case, however for the two-factor CIR case we get a number of different parameter estimates with the same maximum value for the log-likelihood function. The effect of the different approximately optimal parameter vectors on fixed income derivatives is examined and is found to be more pronounced than on the underlying bonds themselves. This leads to the conclusion that fixed income derivatives should also be considered or even included in the estimation of multi-factor term-structure models as this may reduce the problem of local maxima, whereby a number of different parameter vectors perform equally well in the estimation procedure.

The remainder of this paper is organised as follows: first we consider the motivation for the study. In the next section we consider the related literature in term structure model estimation and evolutionary computational techniques in finance. Section two introduces the one and two-factor CIR term structure models to be estimated. This section then outlines the use of the Kalman filter in term structure modelling. The third section will contain a description of differential evolution. The fourth section contains the results and compares them with the well known paper of Duan and Simonato (1999). The effect of local maxima on derivatives prices is examined. The final section concludes and discusses some possible future research avenues.

1.1 Motivation

As the models used in finance become less parsimonious and increasingly complicated to better account for the complexities of financial markets the problem of estimating and calibrating these high-dimensional models is no longer a straightforward step as many of these models contain local maxima. This means that a number of different parameter vectors that describe a model can result in virtually identical estimation or calibration performance. We will refer to these different parameter vectors as locally optimal parameter vectors. If the prices and hedge ratios from these locally optimal parameter vectors are virtually identical then having a number of locally optimal parameter vectors might not be a serious problem if these models are only used in the pricing and hedging of similar securities to those used in the estimation or calibration. However in many cases these models are used in a wider framework, such as in the pricing of exotic securities or in asset allocation decisions that involve many other asset classes. In these cases the locally optimal parameter vectors may cause large differences in exotic prices or may result in a different decision making process. Recent literature contains some examples of the occurrence of local maxima in financial models. For example Ben Hamida and Cont

---

1The parameters for one and two-factor Vasicek models were also estimated using DE. It was found they do not exhibit local maxima. Results are not reported but are available upon request.
Ayache et al (2004) calibrate a regime switching model to a set of option prices and find many possible parameter vector solutions that calibrate to the data equally well. They find that the introduction of exotic options in the calibration step helps reduce this problem. Thus in estimating or calibrating high-dimensional financial models, global optimisers may need to be used to provide information on the objective function as to whether it is well behaved or not and, if it is not well behaved, to reduce the probability of the optimiser returning a parameter vector that is simply a local maximum of the objective function.

Term structure modelling is crucial in fixed income modelling and the proprietary trading desks of many banks and hedge funds use two and three-factor dynamic term structure models in their statistical arbitrage modelling of fixed income markets. However estimating these models based on panel data can be a difficult task for multifactor models and these models can display locally optimal parameter vectors. The Kalman filter approach to estimating term structure models is by now well known and very popular, see for example Duan and Simonato (1999), Babbs and Nowman (1999) among many others. Parameters are estimated by maximising the log-likelihood function which is usually optimised w.r.t. the parameter vector using a deterministic gradient based optimiser. It is recommended to optimise over the log-likelihood function for a number of different parameter vector initialisations, thereby reducing the chance of the deterministic optimiser converging to a local maximum. However this is a rather ad-hoc way to proceed especially if the model is complicated and high dimensional such as Feldhutter and Lando’s (2005) five factor affine model with parameter constraints. The contribution of this paper is in the use of an optimiser known as differential evolution (DE), see Storn and Price (1997), to estimate term structure models. The objective function used in this paper is the quasi log-likelihood function (CIR) resulting from the prediction-error decomposition of a Kalman filter. The objective function is calculated from the prediction-error decomposition that is obtained by running a Kalman filter on a data set based on a single choice of the parameter vector. There is no guarantee that the maximisation of the objective function by appropriate choice of a parameter vector is a well posed problem especially as the complexity of the term structure model increases. For example in a two-factor CIR model the factors should always be positive. For certain choices of parameters the factors may become negative, e.g. if $2\kappa\theta < \sigma^2$ where $\kappa$, is the mean reversion rate, $\theta$ is the long run mean and $\sigma$ is the volatility of the factor, there is a positive probability that this factor will reach zero. In continuous time the process will be reflected away from zero; however in a discrete time setting finite time steps are used and the factor can become negative. To prevent this we can insure the factor does not become negative by replacing a negative value of the factor with a value of zero (the approach adopted in this paper) or by applying a constraint on the parameter space that $2\kappa\theta \geq \sigma^2$. Both of these approaches can result in a non-convex objective function. As the term structure model becomes more complex the number of such parameters and constraints increases thus the use of a global optimiser, such as DE, becomes even more crucial.

1.2 Related Literature

No arbitrage dynamic term structure modelling began with Vasicek (1977), Dothan (1982), Courtadon (1982) and Cox, Ingersoll and Ross (1985) (hereafter CIR). These models posit a diffusion process for the short rate of interest under the physical probability measure and combining this with an assumption on investors’ risk preferences results in a process for the short rate under the equivalent martingale measure in which bond prices and yields can be calculated. Vasicek and CIR are affine dynamic term structure models (DTSM) since bond yields are affine functions of the short rate and as a result
they are tractable and widely-used term structure models. Recognising the need for the use of more than one factor to better explain the correlation structure of yields at different maturities, Brennan and Schwartz (1979), Schaefer and Schwartz (1984), Longstaff and Schwartz (1992) and Chen and Scott (1993) among others, began examining multifactor term structure models. This culminated in the Duffie and Kan (1996) paper where a general affine term structure model was derived that encompasses multifactor versions of the Vasicek and CIR models. More recently Dai and Singleton (2000) have examined maximally flexible canonical affine term structure models. All other affine term structure models are specific examples of their maximally flexible models with certain parameters set to zero. These models allow for non-zero correlation between the latent factors however in their most general form they do not admit closed form solutions for bond prices, so they must be solved by numerically solving a system of ordinary differential equations. In this paper we focus on one and two-factor CIR models for simplicity and clarity. However the methodology of this paper is applicable to any term structure model that can be formulated into a state space model.

Parameter estimation in DTSMs can be done using a number of different methods including the maximum likelihood method, Lo (1986, 1988), the generalised method of moments (GMM), Hansen (1982), and the efficient method of moments (EMM), Gallant and Tauchen (1996). The Kalman filter can also be used to estimate the parameters of a DTSM if it can be formulated into a state space model, see Pennachi (1991), Duan and Simonato (1999), Babbs and Newman (1999) among many others. The KF is a natural approach to use when the underlying state is unobserved and follows a diffusion process. The KF allows for measurement error between the model yield and market yield which naturally arise when using a small number of factors to model the yield curve. Thus the KF can account for model misspecification and market imperfections such as bid-ask spread and illiquidity. Parameter estimates are obtained by optimising a log-likelihood function in Gaussian DTSMs or a quasi log-likelihood function in non-Gaussian DTSMs that results from the prediction error decomposition of the Kalman filter. Affine DTSMs are particularly suited to estimation using the KF because of their tractability and linear nature. However outside the Gaussian class of affine TSMs, including for example the affine CIR model, parameter estimates are inconsistent because the distribution of the shocks to the latent factors is non-Gaussian. However the degree of inconsistency is found to be of little importance in practice according to Lund (1997), Duan and Simonato (1999) and De Jong (2000) who conducted Monte Carlo experiments to check this. Chen and Scott (2003) conducted similar Monte Carlo experiments on the multifactor CIR model and found that whilst some parameter estimates were biased those combinations of parameters that are important for asset pricing were unbiased. Extended KFs or unscented KFs can be combined with quasi-maximum likelihood to estimate models using data that are non-linear functions of the latent factors such as swap data. Recently Duffee and Stanton (2004) suggest that the KF is a robust method to use for parameter estimation in dynamic term structure models and they recommend the use of the KF when maximum likelihood estimation is not feasible. However many prefer the use of KF even when maximum likelihood is feasible as is the case for Vasicek and CIR models. The KF method does not require the assumption that certain market observable rates are observed without error as is done in ML methods. The KF updates the first two moments of a system of latent variables and historical information about the unobserved system is embedded in these two moments. The KF is the optimal filtering technique to use amongst the class of linear filters. The KF is also very suited to out-of-sample tests given that the factors are updated in the filter with the arrival of each new observation date.
2 Dynamic Term Structure Models

In this section we will introduce the two term structure models used in the paper and describe the Kalman filter and how can be used to estimate parameters of dynamic term structure models. We assume that there are \( n \) state variables, denoted \( x_t \equiv (x_{1t}, \ldots, x_{nt})' \) (in this paper we only deal with \( n = 1, 2 \)). Uncertainty is generated by \( n \) independent Brownian motions, where we assume independence for simplicity. The independence assumption can be relaxed and factor correlations can be crucial in modeling certain phenomenon such as the term structure of volatility, see Dai and Singleton (2000).

Under the equivalent martingale measure these Brownian motions are denoted \( \tilde{z}_t \equiv (\tilde{z}_{1t}, \ldots, \tilde{z}_{nt})' \); the corresponding Brownian motions under the physical measure are denoted without the tildes. The instantaneous interest rate, denoted \( r_t \) is affine in the state and given by

\[
r_t = 1' \cdot x_t,
\]

where \( 1 \) is the unit vector with \( n \) elements.

2.1 CIR Model

In the CIR model the state dynamics under the physical measure are given by a square root diffusion process

\[
dx_t = \kappa_i (\theta_i - x_{it}) \, dt + \sigma_i \sqrt{x_{it}} \, d\tilde{z}_{it}.
\]

The equivalent martingale dynamics determine bonds prices. Assume the market price of risk of each factor \( \lambda_i \) is proportional to the factor. The dynamics under the equivalent martingale measure are given by

\[
dx_t = (\kappa_i \theta_i - (\kappa_i + \lambda_i) x_{it}) \, dt + \sigma_i \sqrt{x_{it}} \, d\tilde{z}_{it}.
\]

Zero-coupon bonds maturing at time \( t + \tau \) have prices and yields given by

\[
P(x_t, \tau) = \exp \left[ A(\tau) - B(\tau)' x_t \right], \tag{1}
\]

\[
Y(x_t, \tau) = \left( 1/\tau \right) \left[ -A(\tau) + B(\tau)' x_t \right]. \tag{2}
\]

where functions \( A(\tau) \) and \( B(\tau) \) are given by

\[
A(\tau) = A_1(\tau) + A_2(\tau),
\]

\[
B(\tau) = [B_1(\tau), B_2(\tau)],
\]

\[
A_i(\tau) = \frac{2\kappa_i \theta_i}{\sigma_i^2} \ln \left( \frac{2\gamma_i e^{(\kappa_i + \lambda_i + \gamma_i) \tau}}{(\kappa_i + \lambda_i + \gamma_i) (e^{\kappa_i \tau} - 1) + 2\gamma_i} \right),
\]

\[
B_i(\tau) = \frac{2 (e^{\kappa_i \tau} - 1)}{(\kappa_i + \lambda_i + \gamma_i) (e^{\kappa_i \tau} - 1) + 2\gamma_i},
\]

\[
\gamma_i = \sqrt{(\kappa_i + \lambda_i)^2 + 2\sigma_i^2}.
\]

A negative value for \( \lambda_i \) means the risk premium for holding longer term bonds is positive.
2.2 Kalman Filter

The Kalman filter is a powerful linear filtering technique introduced by Kalman (1960). In its original form it is assumed that a system is driven by an unobservable state that experiences additive noise and there are observables that are linear functions of this unobservable state however the observables themselves are measured with noise. The KF is a filter method that allows the estimation of the unobservable state and its covariance matrix at each point in time using only knowledge on the noisy measurements. It operates iteratively on the data so only the current estimate of the state and its covariance matrix are needed for the prediction of the future state thus it is computationally efficient. The reader is referred to Harvey (1989) for a thorough explanation of KFs and to Duan and Simonato (1999) and various other papers cited in the literature review for example of KFs in interest rate modelling. A brief explanation is outlined here for completeness. The state transition and measurement equations are as follows

\[ x_t = \Phi_0 + \Phi_1 x_{t-h} + \eta_t, \quad \text{var}[\eta_t] = Q_t, \quad (3) \]

\[ Y_t = H_0 + H_1 x_t + e_t, \quad \text{var}[e_t] = R, \quad (4) \]

where \( Q_t \) and \( R \) are the covariance matrices of the unobservable state innovation and the measurement error. Initial conditions are \( \hat{x}_0 = E(x_0) \) and \( \hat{P}_0 = \text{var}(x_0) \). The KF consists of three steps. The prediction step is given by

\[ \hat{x}^-_t = \Phi_0 + \Phi_1 \hat{x}^-_{t-h}, \]

\[ P^-_t = \Phi_1 P^-_{t-h} \Phi_1' + Q_t. \]

The predicted measurement, the prediction error and its covariance are given by

\[ \hat{Y}^-_t = H_0 + H_1 \hat{x}^-_t, \]

\[ u_t = Y_t - \hat{Y}^-_t, \]

\[ \text{var}(P^-_{yy,t}) = H_1 P^-_t H_1' + R. \]

The filtered updates are given by:

\[ K_t = P^-_t H_1' P^{-1}_{yy,t}, \]

\[ \hat{x}_t = \hat{x}^-_t + K_t u_t, \]

\[ P_t = (I - K_t H_1) P^-_t. \]

The log-likelihood function is derived from the prediction error decomposition of the KF and is given by

\[ \ln L(\Psi) = -\frac{Nm}{2} \ln 2\pi - \frac{1}{2} \sum_{t=0}^{N} \ln |P^-_{yy,t}| - \frac{1}{2} \sum_{t=0}^{N} u_t P^{-1}_{yy,t} u_t', \]

where \( N \) is the number of time steps with \( t = \{ h, 2h, ..., T-h, T \} \), \( N = T/h \), \( m \) is the number of different maturity bonds used and \( \Psi \) is the parameter vector.

To apply the KF in the one-factor CIR models initialise the algorithm with the unconditional mean and variance of the state

\[ \hat{x}_0 = \theta, \quad P_0 = \frac{\sigma^2}{2\kappa}, \]

\[ \text{(5)} \]
and choose the elements of the KF as follows:

\[
\Phi_0 = \left(1 - e^{-\kappa h}\right) \theta, \quad \Phi_1 = e^{-\kappa h},
\]

\[
Q_t = \theta \frac{\sigma^2}{2\kappa} \left(1 - e^{-\kappa h}\right)^2 + \hat{r}_{t-h} \frac{\sigma^2}{\kappa} \left(e^{-\kappa h} - e^{-2\kappa h}\right),
\]

\[
H_0 = -\frac{1}{\tau} A(\tau), \quad H_1 = \frac{1}{\tau} B(\tau), \text{ and } R = \text{diag}\left(\sigma^2_{e_i}\right),
\]

where \(\sigma^2\) is the variance of the \(i^{th}\) measurement error and \(\text{diag}\left(\sigma^2_{e_i}\right)\) is a diagonal matrix with diagonal elements \(\sigma^2_{e_i}\). The measurement error covariance matrix can be a non-diagonal however it is chosen to be diagonal in this study so the covariance structure of bond yields is represented only by the model itself and not the measurement error covariance matrix. The application of the KF in a multifactor model is a straightforward extension of the above one-factor case.

3 Differential Evolution

Evolutionary computational techniques have been used in finance applications for quite some time now. They first came to prominence in the form of genetic algorithms and genetic programming and are used in forecasting, classification and trading applications. However they are also used as a complimentary tool in optimisation problems involving complicated high-dimensional and possibly non-convex objective functions that may have many local maxima. Differential evolution (DE), Storn and Price (1997), is a population-based search algorithm that draws inspiration from the field of evolutionary computation. DE embeds concepts of mutation, recombination and fitness-based selection to evolve good solutions. There are a number of worthwhile gains to using DE including: it is more likely to reach the global optimum than deterministic based algorithm in the presence of non-convex objective functions; DE can easily accommodate non-negativity constraints and parameter constraints that might result in a non-convex objective function causing deterministic optimisers to converge to a local optimum; DE can return a population of parameter vectors that are approximately optimal thus giving insight into parameter or model uncertainty that is perhaps more informative than standard asymptotic diagnostic tests; DE can handle large scale optimisation problems thus the inclusion of more parameters does not pose any problems; and DE does not require the gradient of the objective function which can be cumbersome to calculate for certain models. Of course in certain situations, such as the Vasicek term structure model, DE will not offer any more insight into the problem above and beyond a deterministic based optimiser. DE may also take longer to converge than a deterministic optimiser and one can only use informal diagnostics in determining the convergence of the optimiser. That said this paper is not advocating the replacement of deterministic gradient based optimisers with global evolutionary optimisers, but rather encouraging the use of DE or other evolutionary optimisers in certain situations where deterministic optimisers may run into problems or even using a combination of deterministic and evolutionary optimisers.

Although several DE algorithms exist I will describe one version of the algorithm. See Storn and Price (1997) for more detail on the various types of algorithms, Brabazon and O’Neill (2006) for a complete analysis of evolutionary algorithms in the context of financial modelling and Ben Hamida and Cont (2005) for an example of a derivative pricing model calibration method using evolutionary algorithms. The DE algorithm used in this paper is based on the \(DE/rand/1/bin\) scheme. The different variants of the DE algorithm are described using the shorthand \(DE/x/y/z\), where \(x\) specifies how the base vector is chosen, \((rand\) if it is randomly selected, and \(best\) if the best individual in the population
interest rates are expressed on an annualised continuously compounded basis. The data covers the debt securities with maturities: 3, 6, 12 and 60 months taken from the Fama-Bliss data file. All which is kindly posted on their website. It consists of four monthly yield series for the U.S. Treasury In this study the data set used is the same as that used by Duan and Simonato (1999) (hereafter DS).

4 Results

The DE algorithm has three key parameters: the population size \( N \), the crossover rate \( CR \) and the scaling factor \( F \). Higher values of \( CR \) tends to produce faster convergence of the population of solutions. Typical values for these parameters are in the ranges, \( N = 50 - 100 \) or five to ten times the number of dimensions in a solution vector, \( CR = 0.4 - 0.7 \) and \( F = 0.4 - 0.9 \).

Under the mutation operator, for each vector \( \Psi_j(t) \) a variant vector \( V_j(t+1) \) is obtained:

\[
V_j(t+1) = \Psi_m(t) + F(\Psi_k(t) - \Psi_l(t)),
\]

where \( k, l, m \) \( \in \) \( 1, \ldots, N \) are mutually distinct, randomly selected indices, and all the indices \( \neq j \) \( (\Psi_m(t) \) is referred to the base vector and \( \Psi_k(t) - \Psi_l(t) \) is referred to as a difference vector). \( F \) is a scaling parameter and typically \( F \in (0,2] \). The scaling parameter controls the amplification of the difference between \( \Psi_k \) and \( \Psi_l \), and is used to avoid stagnation of the search process. Following the creation of the variant vector, a trial vector \( U_j(t+1) \) is obtained:

\[
U_{jk}(t+1) = \begin{cases} 
V_{jk}(t+1), & \text{if} \ (\text{rand} \leq CR) \ or \ (j = \text{rnbr}(\text{ind})); \\
\Psi_{jk}(t), & \text{if} \ (\text{rand} \geq CR) \ or \ (j \neq \text{rnbr}(\text{ind})).
\end{cases}
\]

where \( k = 1,2,\ldots,d \), \( \text{rand} \) is a random number generated in the range \( (0,1) \), \( \text{CR} \) is the user specified crossover constant from the range \( (0,1) \), and \( \text{rnbr}(\text{ind}) \) is a randomly chosen index chosen from the range \( (1,2,\ldots,d) \). The random index is used to ensure that the trial solution differs by at least one component from \( \Psi_j(t) \). The resulting trial (child) solution replaces its parent if it has higher fitness, otherwise the parent survives unchanged into the next generation.

The DE algorithm has three key parameters: the population size \( N \), the crossover rate \( CR \) and the scaling factor \( F \). Higher values of \( CR \) tends to produce faster convergence of the population of solutions. Typical values for these parameters are in the ranges, \( N = 50 - 100 \) or five to ten times the number of dimensions in a solution vector, \( CR = 0.4 - 0.7 \) and \( F = 0.4 - 0.9 \).
period from April 1964 to December 1997, totalling 405 time series observations. The data is out of date, however given that DS have estimated term structure models on this same data set it was thought appropriate to use it for comparison purposes. As with DS the unit of time is set to one year so that in the Kalman filter $h = \frac{1}{12}$. The following sections report the results for one and two-factor CIR models. The sensitivity of the DE optimiser solutions to the relevant DE parameters is examined in the one factor case.

4.1 One-factor CIR model

The one-factor CIR models were estimated using the DE optimiser and the results were very similar to those reported in DS. Table 1 reports the results for the one-factor CIR model from 100 runs of the DE optimiser. As with all stochastic optimisers, the algorithm should be run a number of times to ensure it has converged. The DE optimiser was run with the following parameters for the optimiser: $NP = 20, F = 0.8, CR = 0.8$ and setting the number of iterations to 300. The results are very similar to the results in DS which are reported in the final column of Table 1. Sensitivity analysis was conducted on the optimiser parameters. The optimiser was run 100 times for each of the following parameter settings: \{NP, F, CR\} = \{20, 0.8, 0.8\}, \{20, 0.6, 0.8\}, \{20, 0.4, 0.8\}, \{20, 0.8, 0.6\} and \{20, 0.8, 0.4\}. The parameter estimates and the parameter standard errors were not sensitive to the different settings. The convergence speed of the algorithm is not sensitive to the crossover rate, $CR$, but does improve for smaller values of $F$. However this faster convergence is achieved at the expense of a less extensive parameter space search. If the 5 worst performing parameter vectors (with the lowest log-likelihood values) were dropped the dispersion of the remaining 95 runs reduces dramatically across all the different settings. Table 2 contains results using the same optimiser parameters as Table 1 but with the 5 worst performing parameter vectors removed from the analysis. These results on the one-factor CIR model indicate that the DE optimiser is a useful optimiser however it must be run a number of times to ensure it has not converged too quickly. The results also indicate that the one-factor CIR model does not suffer from the problem of local maxima.

4.2 Two-factor CIR model

The two-factor CIR model was also estimated and if a factor was negative it was replaced with zero in the optimisation routine. Table 3 reports the results for the two-factor model from 100 runs of the DE optimiser. Most of the parameters estimated in this study are not very different from the parameters estimated by DS; however, the mean and median of the mean reversion for the second factor, $\kappa_2$, and the parameter estimate associated with the overall maximum likelihood value for $\kappa_2$ is quite different to the value of 0.0007 as estimated by DS. This suggests that the second factor might not be as close to a non-stationary process as DS have suggested. However this study is not claiming that parameter estimates should be very different from those of DS but is rather emphasising the local maxima problem that arises in the estimation of the two-factor CIR term structure model. Table 4 contains the same results as Table 3, however the 40 worst performing parameter vectors were removed from the analysis. More parameter vectors are removed in the two-factor case due to a higher proportion that converge too quickly to suboptimal solutions.

The analysis when examined in more detail seems to suggest that the Kalman filter is a double edged sword when used to estimate parameters. The Kalman filter returns unobservable factors for each of the 100 optimal parameters from the DE optimiser that are almost identical from one run to another. This is quite a promising result and it suggests that the Kalman filter is capable of retrieving

---

2These results are not reported to save space but are available from the author upon request.
the underlying factors that are driving the term structure for this particular data set. Figure 1(b) shows a plot of the two factors for the 100 optimal DE parameter vectors, the factors from each parameter vector are so close making it hard to tell them apart on the graph. In fact the two unobservable factors from this analysis are > 99.9% correlated with the corresponding unobservable factors from the other runs of the DE optimiser.

However any two parameter vectors $\Psi_i$ and $\Psi_j$, for $i, j \in \{1, \ldots, 100\}$, are capable of returning almost identical log-likehood values and unobservable factors yet the parameter vectors themselves can differ substantially. To examine this point further figures 2 and 3 contain plots of the parameter values from the 100 DE optimal parameters for the first and second factors respectively versus the negative log-likelihood value. Figure 3(a) plots the long run mean for factor 1, $\theta_1$, versus the negative log-likelihood function and it seems to be converging to its lower bound of zero. Figure 3(b) shows the mean reversion of the first factor, $\kappa_1$, and it seems to be converging to a global optimal value. Figure 3(c) suggest that there are a number of different optimal parameter vectors whose value for $\sigma_1$ can differ (from 0.11 to 0.12) yet whose log-likelihood value is almost identical. However most of the values for $\sigma_1$ are within one standard error of it mean value. Figure 3(d) suggests that the market price of risk of the first factor, $\lambda_1$, also seems to be converging to a global optimal value. The situation for the parameters of the second factor is different as illustrated in figure 4. The long run mean of the second factor, $\theta_2$, seems to be converging to its lower limit of zero. However there are a large number of different optimal parameter vectors whose value for $\kappa_2$ can differ substantially yet all have similar log-likelihood values and the same can be said for $\sigma_2$ and $\lambda_2$. This suggests the two-factor CIR model is misspecified, a conclusion that agrees with that of DS. However many firms still use misspecified term structure models due to their tractibility and ability to explain certain statistical features of the term structure. Thus the problem of locally optimal parameter vectors may be relevant for many term structure models that are used extensively in academia and industry today.

We will now examine the affect of these different approximately optimal parameter vectors on bond prices and bond derivatives. To do this we take each optimal parameter vector from the 100 DE runs and the corresponding unobservable factors $x_1$ and $x_2$ on the last month of the time series data. We then use these to price a 3-year zero coupon bond. For a bond of face value $100 the mean, median and standard deviation of the bond prices are $84.4875, $84.4842 and $0.0144 respectively. The locally optimal parameter vectors do not result in very different bond prices. Figure 4(a) depicts a histogram of the 100 different bond prices for each parameter vector $\Psi$. This means the problem of the local maxima is very minor when taken in the context of bond pricing. However if we price an option on the same bond using the 100 parameter vectors the situation becomes more problematic. Taking the same bond as above we price an at-the-money forward bond option with a strike price of $89.19 and a maturity of 1 year using the two-factor CIR option pricing formula given in Chen and Scott (1992). The mean, median and standard deviation of the bond option prices are $0.6403, $0.6371 and $0.0188. Figure 4(b) depicts a histogram of the 100 option prices for each parameter vector $\Psi$. As can be seen the effect on the option prices is proportionally much larger. The standard deviation of the bond prices is 0.017% of the mean bond price whereas the standard deviation of the option prices is approximately 3% of the mean option price. This problem does not disappear as we cut out the parameter vectors with lower log-likelihoods, i.e. those parameter estimates for which the optimiser converged too quickly. If we consider the 50 parameter vectors with the largest log-likelihoods the problem still persists with the standard deviation being 2% of the mean bond price. This is not surprising given the flatness of the log-likelihood plots where we can see that many different parameter vectors produce the same maximum log-likelihood value. These differences in option prices are magnified when considering out-of-the money options and longer maturity options which are typical of the fixed income market. All though a 2%-3% difference may not seem that large relative to the bid-ask spread in these markets,
Table 1: Parameter estimates for the one-factor CIR model. Column 1 denotes the parameter. Column two is the mean parameter estimate and standard error, three is the median, four is the standard deviation, and five and six are the minimum and maximum from the 100 DE optimal parameter vectors. Column seven is the parameter vector with maximum LL value and column eight contains DS parameter estimates and standard errors.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (s.e.)</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Max LL</th>
<th>DS (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>0.0611 (0.0157)</td>
<td>0.0610</td>
<td>0.0011</td>
<td>0.0607</td>
<td>0.0717</td>
<td>0.0610</td>
<td>0.0613 (0.0123)</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.2247 (0.0580)</td>
<td>0.2251</td>
<td>0.0035</td>
<td>0.1906</td>
<td>0.2267</td>
<td>0.2251</td>
<td>0.2249 (0.0457)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.0702 (0.0051)</td>
<td>0.0702</td>
<td>0.0000</td>
<td>0.0700</td>
<td>0.0703</td>
<td>0.0702</td>
<td>0.0700 (0.0045)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-0.1115 (0.0587)</td>
<td>-0.1119</td>
<td>0.0034</td>
<td>-0.1131</td>
<td>-0.0785</td>
<td>-0.1119</td>
<td>-0.1110 (0.0454)</td>
</tr>
<tr>
<td>( \sigma_{e_1} )</td>
<td>0.0028 (0.0002)</td>
<td>0.0028</td>
<td>0.0000</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0028 (0.0002)</td>
</tr>
<tr>
<td>( \sigma_{e_2} )</td>
<td>0.0000 (0.2883)</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 (7.6255)</td>
</tr>
<tr>
<td>( \sigma_{e_3} )</td>
<td>0.0030 (0.0002)</td>
<td>0.0030</td>
<td>0.0000</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0030 (0.0002)</td>
</tr>
<tr>
<td>( \sigma_{e_4} )</td>
<td>0.0099 (0.0003)</td>
<td>0.0099</td>
<td>0.0000</td>
<td>0.0099</td>
<td>0.0100</td>
<td>0.0099</td>
<td>0.0099 (0.0003)</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimates for the one-factor CIR model. The columns are the same as Table 1 however only the 95 best performing DE optimal parameter vectors are used.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Max LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>0.0610</td>
<td>0.0610</td>
<td>0.0000</td>
<td>0.0608</td>
<td>0.0610</td>
<td>0.0610</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.2252</td>
<td>0.2251</td>
<td>0.0001</td>
<td>0.2248</td>
<td>0.2258</td>
<td>0.2251</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.0702</td>
<td>0.0702</td>
<td>0.0000</td>
<td>0.0701</td>
<td>0.0702</td>
<td>0.0702</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-0.1119</td>
<td>-0.1119</td>
<td>0.0000</td>
<td>-0.1126</td>
<td>-0.1116</td>
<td>-0.1119</td>
</tr>
<tr>
<td>( \sigma_{e_1} )</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0000</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.0028</td>
</tr>
<tr>
<td>( \sigma_{e_2} )</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \sigma_{e_3} )</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0000</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0030</td>
</tr>
<tr>
<td>( \sigma_{e_4} )</td>
<td>0.0099</td>
<td>0.0099</td>
<td>0.0000</td>
<td>0.0099</td>
<td>0.0099</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

when these options are combined to form a typical fixed income instrument such as a cap (given a cap can be decomposed into a portfolio of bond options) they will quickly accumulate and result in large option price differences. The suggested remedies are to include more than four cross-sectional bond maturities in the estimation and perhaps even include a very liquid benchmark bond derivative in the estimation. If non-linear securities such as derivatives are used in the estimation it means that non-linear Kalman filter techniques such as the extended or unscented Kalman filter will have to be used in place of the standard Kalman filter. This research is currently being undertaken by the author.

5 Conclusion

In this paper the problem of local maxima in the context of parameter estimation for dynamic term structure models was highlighted using an evolutionary optimiser known as differential evolution. The affect of the local maxima on bond pricing was shown to be relatively minor however the affect on bond derivatives was shown to be a reasonable proportion of the derivative price. The suggested
Table 3: Parameter estimates for the two-factor CIR model. Column one denotes the parameter. Column two is the mean parameter estimate and standard error, three is the median, four is the standard deviation, and five and six are the minimum and maximum from the 100 DE optimal parameter vectors. Column seven is the parameter vector with maximum LL value and column eight contains DS parameter estimates and standard errors.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>MaxLL</th>
<th>Duan and Simonato (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.0703</td>
<td>0.0294</td>
<td>0.1080</td>
<td>0.0257</td>
<td>0.6852</td>
<td>0.0271</td>
<td>0.0303 (0.0031)</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.9509</td>
<td>1.1413</td>
<td>0.4045</td>
<td>0.0412</td>
<td>1.3144</td>
<td>1.2689</td>
<td>1.1627 (0.1508)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1146</td>
<td>0.1146</td>
<td>0.0034</td>
<td>0.1070</td>
<td>0.1221</td>
<td>0.1178</td>
<td>0.1202 (0.0079)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.1632</td>
<td>-0.3320</td>
<td>0.3562</td>
<td>-0.4959</td>
<td>0.6371</td>
<td>-0.4395</td>
<td>-0.3139 (0.1222)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0011</td>
<td>0.0000</td>
<td>0.0000 (0.0000)</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.0314</td>
<td>0.0295</td>
<td>0.0171</td>
<td>0.0016</td>
<td>0.0946</td>
<td>0.0297</td>
<td>0.0007 (0.0614)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0618</td>
<td>0.0621</td>
<td>0.0018</td>
<td>0.0558</td>
<td>0.0653</td>
<td>0.0622</td>
<td>0.0681 (0.0035)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.0367</td>
<td>-0.0356</td>
<td>0.0176</td>
<td>-0.1038</td>
<td>-0.0075</td>
<td>-0.0362</td>
<td>-0.0266 (0.0636)</td>
</tr>
<tr>
<td>$\sigma_{e1}$</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0000</td>
<td>0.0025</td>
<td>0.0027</td>
<td>0.0025</td>
<td>0.0027 (0.0001)</td>
</tr>
<tr>
<td>$\sigma_{e2}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0000</td>
<td>0.0000 (0.0985)</td>
</tr>
<tr>
<td>$\sigma_{e3}$</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0000</td>
<td>0.0020</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0021 (0.0001)</td>
</tr>
<tr>
<td>$\sigma_{e4}$</td>
<td>0.0018</td>
<td>0.0018</td>
<td>0.0001</td>
<td>0.0017</td>
<td>0.0020</td>
<td>0.0017</td>
<td>0.0012 (0.0001)</td>
</tr>
</tbody>
</table>

Table 4: Parameter estimates for the two-factor CIR model. The columns are the same as Table 3 however only the 60 best performing DE optimal parameter vectors are used.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>MaxLL</th>
<th>Duan and Simonato (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.0284</td>
<td>0.0282</td>
<td>0.0014</td>
<td>0.0257</td>
<td>0.0318</td>
<td>0.0271</td>
<td>0.0303 (0.0031)</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>1.1993</td>
<td>1.2051</td>
<td>0.0675</td>
<td>1.0415</td>
<td>1.3144</td>
<td>1.2689</td>
<td>1.1627 (0.1508)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.1156</td>
<td>0.1154</td>
<td>0.0025</td>
<td>0.1105</td>
<td>0.1196</td>
<td>0.1178</td>
<td>0.1202 (0.0079)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.3812</td>
<td>-0.3930</td>
<td>0.0606</td>
<td>-0.4959</td>
<td>-0.2436</td>
<td>-0.4395</td>
<td>-0.3139 (0.1222)</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0011</td>
<td>0.0000</td>
<td>0.0000 (0.0000)</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.0308</td>
<td>0.0297</td>
<td>0.0152</td>
<td>0.0016</td>
<td>0.0624</td>
<td>0.0297</td>
<td>0.0007 (0.0614)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0625</td>
<td>0.0624</td>
<td>0.0011</td>
<td>0.0610</td>
<td>0.0653</td>
<td>0.0622</td>
<td>0.0681 (0.0035)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>-0.0368</td>
<td>-0.0367</td>
<td>0.0156</td>
<td>-0.0690</td>
<td>-0.0077</td>
<td>-0.0362</td>
<td>-0.0266 (0.0636)</td>
</tr>
<tr>
<td>$\sigma_{e1}$</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0000</td>
<td>0.0025</td>
<td>0.0027</td>
<td>0.0025</td>
<td>0.0027 (0.0001)</td>
</tr>
<tr>
<td>$\sigma_{e2}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000 (0.0985)</td>
</tr>
<tr>
<td>$\sigma_{e3}$</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0000</td>
<td>0.0020</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0021 (0.0001)</td>
</tr>
<tr>
<td>$\sigma_{e4}$</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0000</td>
<td>0.0017</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0012 (0.0001)</td>
</tr>
</tbody>
</table>
Figure 1: Fama Bliss data set of four US Treasury yields with maturities: 3, 6, 12 and 60 months and the two unobservable factors for the 100 runs of the DE optimiser

Figure 2: Factor 1 parameter estimates versus the (negative) log-likelihood function
Figure 3: Factor 2 parameter estimates versus the (negative) log-likelihood function

Figure 4: Histograms of bond and option prices for each optimal parameter vector $\Psi$
remedies are to use a better specified dynamic term structure models than the CIR for the data set in this study or to use more cross-sectional bond maturities and even a liquid benchmark derivative in the estimation procedure. Future work should include simulation studies to determine whether the local maxima are a result of the dynamic term structure model itself or a result of some specific structure in the data set such as near unit root behaviour or perhaps a combination of the two.

References


