Kalman Filtering and the Estimation of Multi-factor Affine Term Structure Models

Thabo Tokwe

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

November 23, 2016

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of the Cape Town. It has not been submitted before for any degree or examination in any other University.

Signed by candidate

November 23, 2016

Abstract

When optimising the likelihood function one often encounters various stationary points and sometimes discontinuities in the parameter space (Gupta and Mehra, 1974). This is certainly true for a majority of multi-factor affine term structure models. Practitioners often recover different parameter optimisations depending on the initial parameters. If these parameters result in different option prices, the implications would be severe. This paper examines these implications through numerical experiments on the three-factor Vasicek and Arbitrage-free Nelson-Siegel (AFNS) models. The numerical experiments involve Kalman filtering as well as likelihood optimisation for parameter estimation. It was found that the parameter sets lead to the same short rate process and thus the same model. Moreover, likelihood optimisation in the AFNS does not result in different parameter sets irrespective of the starting point.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor doctor Peter Ouwehand. His unfailing support and guidance made it possible to complete this dissertation. I am also thankful to my parents and friends. The motivation and love they provided was invaluable.

Contents

List of Figures

vi

List of Tables

Chapter 1

Introduction

The understanding of what moves bond yields has been and continues to be a subject of profound interest for market analysts, investors and policy makers. Term structure modelling in particular has produced immense literature over the years: not only does it give insight into interest-rate risk, it also facilitates the simulation of interest rate scenarios and the valuation of financial instruments– all of which are essential to these economic agents (Piazzesi, 2009).

Interest rate term structures define a relationship between the yield of a fixed income investment and the time to maturity of its cash flows. Models of term structure describe these interest rate dynamics in two dimensions. The first is essentially a cross-sectional fitting of a zero-coupon yield curve to a set of zero-coupon bond prices and the other describes term structure inter-temporal dynamics, showing the evolution of interest rates with time (Chatterjee, 2005).

As a result of the substantial correlations among bond yields of different maturities, researchers have endeavoured to construct parsimonious models using state variables that drive the entire yield curve for a given market while describing their correlation structure. Maes (2000) and Dai and Singleton (2003) delineate three requirements for the construction of any continuous-time yield curve model: a particular diffusion for the underlying factors, risk premia and a functional relationship between the risk-free rate and the underlying state variables.

Pioneering studies (Vasicek, 1977; Cox et al., 1985) were based on a single factor. However, the significant disparities between empirical data and model predictions necessitated the development and implementation of multi-factor models. Contemporary research is primarily focused on two classes of such models viz. Nelson-Siegel Models (NSMs) and Affine Term Structure Models (ATSMs) (Gasha et al., 2010). Whereas the former tend to incorporate both observable and unobservable macroeconomic factors, they need not be consistent with the absence of arbitrage opportunities. ATSMs on the other hand may depend on the absence of arbitrage and assume that the underlying unobservable factors follow stochastic

processes. This study only considers the latter class of TSMs.

The affine class of models is relatively tractable and there are extensive studies on it (Duffee and Stanton, 2012; Duffie and Kan, 1996; Christensen *et al.*, 2011). Despite this the models rely on imperfect observations of bond data. Naturally, therefore, some assumptions about the distributions of these errors have to be made (state variable dynamics are also stochastic). Regardless of the chosen structure of the measurement error it is often the case that one cannot simply invert the model to find unknown factors. Hence more sophisticated filtering techniques are essential (Chatterjee, 2005). One of these techniques is the Kalman filter.

Following the work of Duan and Simonato (1999), Lund (1997), De Jong (2000), Geyer and Pichler (1999) and Babbs and Nowman (1999), the Kalman filter has become widely used for parameter estimation in term structure modelling. It has been observed however, that when using the Kalman filter in conjunction with quasi—maximum likelihood estimation the original parameters used to simulate state processes are rarely recovered in the case of multi-factor models. It is therefore the aim of this study to explain why this is the case and to make some inferences about the effects of this in option pricing.

Chapter 2

Affine Term Structure Models

2.1 Overview of Affine Term Structure Literature

The underlying state variables (factors that describe the state of a dynamic system at any given time) and their stochastic processes simultaneously determine the cross-sectional shape as well as the temporal evolution of the yield curve. Prior to the advent of Vasicek (1977), empirical and theoretical investigations such as those by Macaulay *et al.* (1938) tended to analyse only one of these implications. The first term structure model to use the panel data approach— where the dynamic and cross-sectional implications are accounted for simultaneously— was developed by Vasicek (1977) (Lund, 1995). This model treats the instantaneous spot interest rate as the single factor. It involves a specification of the underlying short rate diffusion and the use of no-arbitrage arguments to derive bond pricing formulae which then enable the construction of a term structure. Whereas Vasicek (1977) used Gaussian diffusion, Cox *et al.* (1985) proposed a one-dimensional square root diffusion process which ensures positive short rates in what is known as the CIR model (Cox *et al.*, 1985). Despite their popularity, the predictions of these pioneering models were vastly different to empirical data and this failure inspired the development of numerous other models including multi-state extensions of these models (Dai and Singleton, 2000).

The development of multi-factor models was further substantiated by the cogent argument that term structures are embedded in large macroeconomic systems. Logically therefore, there should be a multiplicity of economic factors which relate to the term structure of interest rates (Langetieg, 1980). Motivated by this argument, Langetieg (1980) extended the Vasicek model three years after its publication to incorporate a general number of economic factors. Multi-factor versions of the CIR model were then developed by inter alia Longstaff and Schwartz (1992) and Chen and Scott (1992).

With respect to the sufficient number of factors to adequately characterize the

term structure, principal components analysis can be used to decompose the motion of interest rates into independent factors. Wilson (1994) showed that 80-90% of the variation of the term structure is explained by parallel shifts of the yield curve, 5-10% by a twist (long term and short term rates move in opposing directions, centred around a point), and 1-2% by a butterfly (long and short term rates move in the same direction, while mid-term rates move in the opposite direction)— see Figure 2.1. This result has important consequences because researchers inevitably have to make some concession between richness of econometric representation of the underlying factors and the computational expense of running these models.
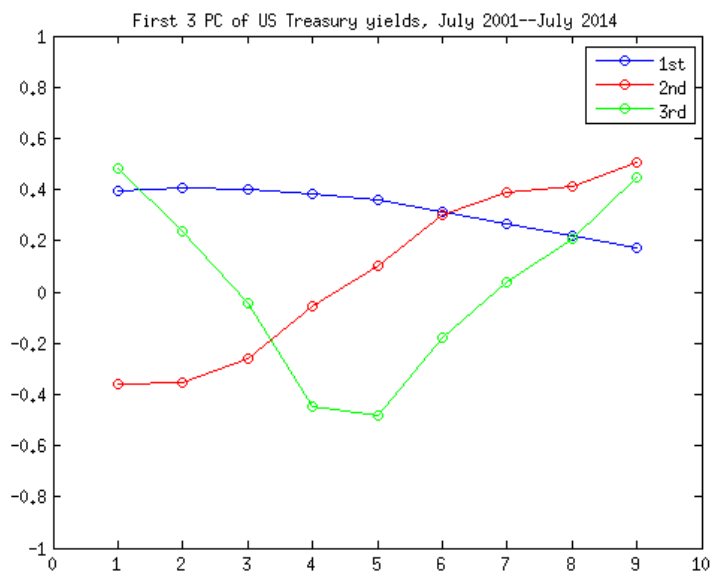
Fig. 2.1: The first three principal components of US Treasury yields for 9 maturities— to construct this figure we take the time series of bond yields for N maturities, and find the eigenvalues of the resulting $M \times N$ matrix (where M is the number of observations in each time series). Then we arrange these in decreasing order i.e. l_1, l_2, l_3, ..., l_N. Then the first principal component explains $l_1/(l_1 + .. + l_N)$ of the variance, the first two explain $(l_1 + l_2)/(l_1 + .. + l_N)$ of the variance, etc. Typically the first three will explain over 95%. The graph shows how the factors may be interpreted. The first one is typically flat, which means that a move in the direction of the first eigenvector results in a roughly equal move for all maturities – a parallel shift. The second is increasing, and thus a move in the direction of the second eigenvector will move big maturities more than small maturities — a slope. The third has a hump.

In light of this trade-off that confronts all practitioners, it is perhaps not surprising that empirical implementation of multi-factor models (those that consider the co-movement over time of short and long term bond yields) is primarily focused on ATSMs. This class of models was defined by Duffie and Kan (1996) and was so called because bond yields are affine functions of the underlying state variables that introduce stochasticity to the model. Equivalently, they are those models for which both the drift and the volatility-squared are affine functions of the state variables. These models greatly facilitate empirical implementation since they yield essentially closed-form expressions for zero-coupon-bond prices. The numerical and analytical tractability thus afforded by these models makes them a subject of much research.

Notwithstanding the notable progress heretofore, there are numerous limitations associated with this class of models particularly the first-generation interest rate models above. For a summary of the disparities between the models and empirical term structures as well as an elaborate evaluation of these limitations see Duffee (2002). Among other things it was found there that the standard class of affine models produces a poor forecast of future moves in long dated bonds. In addition— and of equal importance— parameter estimation in these models is known to be problematic due preponderantly to the existence of numerous local likelihood maxima that have a similar fit to the data while their economic implications may be substantially different. These difficulties seem to reflect a problem of model over-parametrisation (Christensen *et al.*, 2011). Many practitioners (e.g Duffee (2002) and Dai and Singleton (2000)) attempt to remedy this by restricting some parameters to zero thereby obtaining a more parsimonious model. This approach seems to assuage the estimation problem but it is a method improvised for reasons of expediency and is not well motivated theoretically or statistically (Christensen *et al.*, 2011). In spite of these short-comings however, Dai and Singleton (2003) insist that these models give appreciable insight into the expectation hypothesis and the term structure; therefore they cannot be discarded. This exemplifies the conundrum that plagues interest rate modelling incessantly i.e. theoretically rigorous models often yield disappointing empirical results while empirically successful models are not in harmony with accepted theory.

Recently Christensen *et al.* (2011) claimed to have overcome this dilemma by developing a new class of models— the Arbitrage-Free Nelson-Siegel (AFNS) models. They developed this class by augmenting the popular Dynamic Nelson-Siegel (DNS) model with properties of ATSMs. The DNS model is empirically successful but lacks a theoretical foundation, while ATSMs are empirically inadequate but theoretically sound. Furthermore, the DNS is empirically stable and simple to es-

timate; it is also relatively flexible and fits the cross-section as well as the time series of yields remarkably well for many grades of bonds and in many countries (Christensen *et al.*, 2011). These are some of the properties that justify its popularity among financial market practitioners. Nevertheless, DNS fails on a fundamental theoretical dimension— it does not impose the restriction necessary to eliminate the possibility of arbitrage opportunities. Thus, Christensen *et al.* (2011) imposed the absence of arbitrage on the DNS to reconcile it with theory— the result is what is now known as the AFNS.

The AFNS was formulated such that it is consistent with the affine class of models as defined by Duffie and Kan (1996). This model has the tractability and theoretical appeal of ATSMs without the parameter estimation difficulties and with much better empirical performance. It is more parsimonious than the three-factor Vasicek and CIR models and because it uses DNS, the latent state variables are easily identified as the level, slope and curvature. These facts make the model easier to optimise since it is not as over-specified as other ATSMs (Christensen *et al.*, 2011). Furthermore, bond prices have simple closed form expressions, making implementation simple.

Typically the number of bond maturities used is greater than the number of state variables, so if no error structure is specified, solution methods break down due to the singularity problem— which occurs when the number of bond tenors exceeds the number of state variables. This is unsurprising because there are inherent measurement errors in the observed data. These errors stem primarily from the fact that ATSMs assume frictionless markets, which is obviously untrue. Furthermore, there are exogenous sources of error e.g. non-synchronous trading, rounding of prices and bid-ask spreads (Lund, 1995). According to Maes (2000), irrespective of the measurement error structure; if the model cannot simply be inverted to solve for the state variables, a more sophisticated filtering technique is necessary— this is where the Kalman filter finds relevance.

The Kalman filter is a recursive algorithm that is used for parameter estimation. The fundamental idea and application of this approach was first put forward by Pennachi (1991) (Lund, 1997) and after subsequent work by *inter alia* Duan and Simonato (1999), Lund (1997), De Jong (2000), Geyer and Pichler (1999) and Babbs and Nowman (1999), there was a proliferation of papers using this methodology for estimating exponential affine term structure models, especially for multi-factor Gaussian and CIR models.

The filter uses maximum likelihood estimation (MLE) for parameter estimations. In the case of Gaussian term structure models, the exact likelihood function is obtained directly from the Kalman filter algorithm. However, when using

non-Gaussian models there is no exact closed form likelihood function. In such instances a quasi-maximum likelihood (QML) estimator can be constructed from the first and second conditional moments (Maes, 2000). Duan and Simonato (1999) showed that this results in a loss of efficiency as well as the consistency property but Bollerslev and Wooldridge (1992) showed that the QML is asymptotically efficient and consistent.

In their review of various numerical optimisation algorithms for the likelihood function, Gupta and Mehra (1974) acknowledge that maximum likelihood estimation leads to difficult non-linear programming problems in practice. One usually encounters numerous maxima, saddle points and sometimes even discontinuities in the parameter space. Moreover, Kim and Orphanides (2005) alluded to these difficulties, adding that the likelihood function can be flat in multiple directions around these maxima resulting in large standard errors of the estimates. The solutions that researchers have for such problems involve the use of arbitrary procedures as mentioned before. Needless to say, this is a precarious course especially when one does not understand the consequences of selecting one solution over another because the economic implications of each distinct set of parameters may be different. We therefore wish to consider the effects of these disparate parameter estimations on short rate dynamics and briefly on bond option pricing. We also wish to verify the claim that AFNS models do not have multiple maxima in the likelihood function.

2.2 Mathematical Framework

This section is explores some of the fundamental concepts that pertain to term structure modelling as treated by Piazzesi (2009) and Pitsillis (2015). We focus exclusively on the affine class, as suggested in the preceding text. At this stage it would be appropriate to note that there are numerous important assumptions that facilitate term structure modelling. The assumption of an arbitrage-free and complete market forms the framework in which bond prices are modelled.

The term structure relates bond yields to different maturities so its most basic building blocks are $P_t(\tau)$, the time-t zero coupon bond price with tenor $\tau = T - t$ (where T is the maturity) and the instantaneous short rate (risk free), r_t. Bond yields can then be calculated as

$$y_t(\tau) = -\frac{\log P_t(\tau)}{\tau}.$$

The short rate is simply the limit of the yield as maturity tends to zero i.e. $r_t = \lim_{\tau \downarrow 0} y_t(\tau)$. The relationship between the short rate and bond prices depends on

the assumptions made about the bond market. For the purposes of this study, only martingale pricing theory is considered.

We assume the absence of arbitrage in the following form: that there exists a measure $\mathbb{Q}$ (equivalent to the real world or data generating measure $\mathbb{P}$) under which discounted bond prices are martingale, and because we also assume the market is complete, such a measure is unique. Under the $\mathbb{Q}$-measure bond prices given by

$$P_t(\tau) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} r_u du\right)\bigg|\mathcal{F}_t\right]$$

where the filtration generated by the stochastic processes is given by $\mathcal{F}_t$. Here the numéraire is the riskless money market account therefore, $\mathbb{Q}$ is the risk-neutral measure. It has been assumed that the reader is familiar with the martingale pricing theory and thus much detail has been omitted for brevity— for further reading see Björk (2009).

Building a term structure model in this paradigm is well established. To construct a model, one has to assume:

- a particular diffusion for the unobservable states;

- specification of the price of interest rate risk;

- a functional relationship between the underlying factors and the short rate.

Formally this can presented as

$$\text{Assumption 1}: dX_t = \nu(X,t)dt + \sigma(X,t)dW_t$$
$$\text{Assumption 2}: \frac{d\pi_t}{\pi_t} = -r(X,t)dt - \Lambda(X,t)'dW_t$$
$$\text{Assumption 3}: r = f(X,t).$$

In the equations above, W_t is an $n \times 1$ vector of independent $\mathbb{P}$ standard Brownian motions that constitute the noise component of the model, X_t is an n-dimensional vector of state variables, $\nu(X,t)$ is an $n \times 1$ vector of instantaneous (conditional) means, $\sigma(X,t)$ is the covariance matrix of size $n \times n$, π_t is the price deflator i.e. $\pi_t = \exp\left(\int_t^T r_u du\right)$ for maturity T, and lastly, $\Lambda(X,t)$ is a vector of size $n \times 1$ representing the market price of risk. Every i^{th} element of $\Lambda(X,t)$ is the price of risk corresponding to $W_t^{(i)}$. Since pricing is done under risk neutral dynamics, a change of measure is required and to accomplish this, Girsanov's theorem can be used. This theorem relates the risk neutral and real dynamics in this equation:

$$d\widetilde{W}_t = dW + \Lambda(X,t)dt.$$

where $\widetilde{W}_t$ is the standard $\mathbb{Q}$ Brownian motion. Furthermore, under $\mathbb{Q}$

$$dX_t = (\nu(X,t) + \Lambda(X,t)\sigma(X,t))dt + \sigma(X,t)d\widetilde{W}_t.$$

Thus the dynamics are

$$dX_t = \mu(X,t)dt + \sigma(X,t)d\widetilde{W}_t.$$

Since the $\mathbb{Q}$ dynamics of the risk factors are specified, Itō's lemma can be used to facilitate the identification of arbitrage-free prices. The assumption that the underlying states are Itō diffusions (and thus Markovian) allows one to write the bond price as a function of X_t directly, i.e.

$$P_t^{(\tau)} = F(X_t, t, \tau).$$

Provided that $F(X, t, \tau)$ satisfies certain smoothness conditions Itō's lemma can be used to derive a differential equation for $P_t^{(t)}$. The SDE of the bond price is given by

$$dF(X,t,\tau) = \mu_F(X,t,\tau)dt + \sigma_F(X,t,\tau)d\widetilde{W}_t.$$

From Itō's lemma we get that

$$\mu_F(X,t,\tau) = F_t(X,t,\tau) + F_X(X,t,\tau)\mu(X,t) + \frac{1}{2}\text{tr}[\sigma(X,t)\sigma(X,t)'F_{XX}(X,t,\tau)].$$

Here $F_t(X,t,\tau) = \frac{\partial F(X,t,\tau)}{\partial t}$, $F_X(X,t,\tau)$ is a $1 \times n$ gradient vector and $F_{XX}(X,t,\tau)$ is an $n \times n$ Hessian matrix. Also

$$\sigma_F(X,t,\tau) = F_X(X,t,\tau)\sigma(X,t).$$

The risk neutral dynamics of $F(X,t,\tau)$ are known to follow

$$dF(X,t,\tau) = r_t F(X,t,\tau)dt + \sigma_F(X,t,\tau)d\widetilde{W}_t;$$

by definition of $\mathbb{Q}$. For no arbitrage to be admitted $\mu_F(X,t,\tau) = r_t F(X,t,\tau)$ and this yields the bond price PDE

$$F_t(X,t,\tau) + F_X(X,t,\tau)\mu(X,t) + \frac{1}{2}\text{tr}[\sigma(X,t)\sigma(X,t)'F_{XX}(X,t,\tau)] - r_t F(X,t,\tau) = 0[1].$$

2.3   Affine term structure models

A short rate model is said to possess an affine term structure (ATS) if bond prices are given by

$$P_t^{(\tau)} = F(X,t,T) = e^{A(\tau)+B(\tau)X_t}[2].$$

---

[1] Here tr[...] refers to a trace as opposed to a transpose.
[2] Generally one can have $A(t,\tau)$ and $B(t,\tau)$ but we ignore this for our purposes

Where $A(\tau)$ and $B(\tau)$ are sufficiently regular deterministic functions. In particular, $A(\tau)$ and $B(\tau)$ are at least twice differentiable functions of τ. Duffie and Kan (1996) illustrate under general circumstances that such an expression of bond prices implies that $\mu(X, t), \sigma(X, t)\sigma(X, t)'$ as well as r_t are all affine functions of X_t, i.e.

$$\mu(X, t) = \alpha_1(t)X_t + \alpha_0(t)$$
$$\sigma(X, t)\sigma(X, t)' = \beta_1(t)X_t + \beta_0(t)$$
$$r_t = \delta_X' X_t + \delta_0$$

where $\alpha_1(t)$ and $\beta_1(t)$ are real-valued $n \times n$ matrices while $\alpha_0(t), \beta_0(t)$ and δ_X are $n \times 1$ vectors with δ_0 being a scalar.

Following Dai and Singleton (2000) we note that X_t follows an affine diffusion given by

$$dX_t = \widetilde{\kappa}(\widetilde{\theta} - X_t)dt + \Sigma\sqrt{S(t)}d\widetilde{W}_t.$$

$\tilde{\kappa}$ and Σ are $n \times n$ matrices which may be non-diagonal and asymmetric, $\tilde{\theta}$ is an $n \times 1$ matrix while $S(t)$ is a diagonal matrix where the diagonal elements are given by $S(t)_{ii} = \rho_i + \beta_i' X_t$.

Under this parametrisation, $A(\tau)$ and $B(\tau)$ are solutions to these Ricatti equations

$$\frac{dA(\tau)}{d\tau} = -\widetilde{\theta}'\widetilde{\kappa}'B(\tau) + \frac{1}{2}\sum_{i=1}^{n}[\Sigma'B(\tau)]_i^2\rho_i - \delta_0,$$
$$\frac{dB(\tau)}{d\tau} = \widetilde{\kappa}'B(\tau) - \frac{1}{2}\sum_{i=1}^{n}[\Sigma'B(\tau)]_i^2\beta_i + \delta_X.$$

These ODEs can be solved through numerical integration using $A(T, T) = 0$ and $B(T, T) = 0_{n \times 1}$ as initial conditions since they are completely specified when the risk neutral dynamics of r_t are known.

Pricing bonds in empirical studies of ATSMs requires the distributions of X_t and $P(t, T)$ under the data-generating measure $\mathbb{P}$ and for this we consider the price of risk. Since the drift and the square diffusion are affine functions of the short rate, the price of risk is assumed to have the form

$$\Lambda_t = \sqrt{S(t)}\lambda,$$

where λ is an $n \times 1$ vector of constants. With this specification of the price of risk the $\mathbb{P}$ dynamics of X_t are

$$dX_t = [\widetilde{\kappa}(\widetilde{\theta} - X_t) + \Sigma S(t)\lambda]dt + \Sigma\sqrt{S(t)}dW_t,$$

which implies   (2.1)

$$dX_t = \kappa(\theta - X_t)dt + \Sigma\sqrt{S(t)}dW_t,$$

where $\kappa = \widetilde{\kappa} - \Sigma\Phi$ and $\theta = \kappa^{-1}(\widetilde{\kappa}\widetilde{\theta} + \Sigma\psi)$. Φ is an $n \times n$ matrix with the ith row given by $\lambda_i\beta_i'$ and and ψ is an $n \times 1$ vector elements $\lambda_i\rho_i$.

In order to ensure that these differential equations have strong solutions, a number of restrictions need to be invoked. For a detailed exposition of these conditions see Dai and Singleton (2000).

We consider two such models— the three-factor Vasicek model and AFNS model. The general Vasicek model has the following dynamics.

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

where all the parameters are constant. For an n-factor Vasicek model the τ-bond price is given by

$$P_t(\tau) = \exp\left(A(\tau) + \sum_{i=1}^{n} B_i(\tau)X_t^i\right),$$

where

$$B_i(\tau) = \frac{1}{\kappa_i}(1 - e^{-\kappa_i\tau})$$

and

$$A(\tau) = \sum_{i=1}^{n} \frac{\gamma_i(B_i(\tau) - \tau)}{\kappa_i^2} - \frac{\sigma_i^2 B_i^2(\tau)}{4\kappa_i} + \sum_{\{i,j:i\neq j\}} \frac{\sigma_{i,j}}{2\kappa_i\kappa_j}\left(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i + \kappa_j}\left(1 - e^{(\kappa_i+\kappa_j)(\tau)}\right)\right);$$

$$\gamma_i = \kappa_i^2\left(\theta_i - \frac{\sigma_i\lambda_i}{\kappa_i}\right) - \frac{\sigma_i^2}{2}$$

(Gasha *et al.*, 2010). In a three factor model we use $n = 3$. The AFNS is considered in the next section.

2.4   The Nelson-Siegel yield curve model

Since the AFNS model is more novel, it is given more consideration. In this section we provide a brief exposition of the Nelson-Siegel model. We commence with a simple derivation from first principles following which we accentuate some of its defining characteristics and thereafter we relate it to the AFNS model.

Nelson and Siegel (1987) developed the Nelson-Siegel model to fit static yield curves. They considered an instantaneous forward rate model at maturity τ, denoted $f(\tau)$ as the solution to a 2nd order ODE with real and unequal roots. The motivation behind this was the fact that functions that readily produce the typical term structure shapes are usually solutions to such ODEs. The derivation therefore begins with the differential equation in this generalised form

$$\frac{\partial^2 f(\tau)}{\partial\tau^2} + b\frac{\partial f(\tau)}{\partial\tau} + cf(\tau) = d.$$

Dividing through by d we get

$$\frac{1}{d}\frac{\partial^2 f(\tau)}{\partial \tau^2} + \frac{b}{d}\frac{\partial f(\tau)}{\partial \tau} + \frac{c}{d}f(\tau) = 1.$$

This is a non-homogeneous differential equation because the right-hand side is non-zero. To solve this we invoke the following theorem (Stewart, 2015).

Theorem 1. *The general solution of the second order non-homogeneous linear differential equation*

$$y'' + p(t)y' + q(t)y = g(t)$$

can be expressed in the form

$$y = y_c + y_p,$$

*where y_p is **any** specific function that satisfies the non-homogeneous equation and $y_c = C_1 y_1 + C_2 y_2$ is the general solution to the corresponding homogeneous equation*

$$y'' + p(t)y' + q(t)y = 0.$$

y_1 and y_2 are fundamental solutions of the corresponding homogeneous equation while C_1 and C_2 are arbitrary constants.

□

It should be obvious that the analogous homogeneous differential equation in our case is

$$\frac{1}{d}\frac{\partial^2 f(\tau)}{\partial \tau^2} + \frac{b}{d}\frac{\partial f(\tau)}{\partial \tau} + \frac{c}{d}f(\tau) = 0.$$

The general solution y_c of this homogeneous equation is

$$y_c = \frac{1}{d}e^{\alpha_1 \tau} + \frac{b}{d}e^{\alpha_2 \tau},$$

where α_1 and α_2 are the solutions to the quadratic equation

$$\frac{1}{d}\alpha^2 + \frac{b}{d}\alpha + \frac{c}{d} = 0.$$

Thus the solution of the non-homogeneous equation we need to find y_p s.t.

$$f(\tau) = \frac{1}{d}e^{\alpha_1 \tau} + \frac{b}{d}e^{\alpha_2 \tau} + y_p.$$

Solving this requires that some inference is made about the form we expect the solution to take. Since the right-hand side of the non-homogeneous equation is constant, we attempt to identify a solution (y_p) which is a constant i.e. we assume it satisfies

$$y_p = u.$$

This implies

$$\frac{\partial y_p}{\partial \tau} = 0 \text{ and } \frac{\partial^2 y_p}{\partial \tau^2} = 0.$$

We verify whether this is sufficient by substituting into the non-homogeneous equation

$$\frac{1}{d} \cdot 0 + \frac{b}{d} \cdot 0 + \frac{c}{d} \cdot u = 1 \Rightarrow u = \frac{d}{c}.$$

Therefore the solution to our differential equation is

$$f(\tau) = \frac{1}{d}e^{\alpha_1 \tau} + \frac{b}{d}e^{\alpha_2 \tau} + \frac{d}{c}.$$

This can be rewritten as

$$f(\tau) = \beta_0 + \beta_1 e^{-\tau/m_1} + \beta_2 e^{-\tau/m_2},$$

For $\beta_0 = \frac{d}{c}, \beta_1 = \frac{1}{d}, \beta_2 = \frac{b}{d}, m_1 = -\frac{1}{\alpha_1}, m_2 = -\frac{1}{\alpha_2}$.

The original Nelson-Siegel model for instantaneous forward rates at time τ took this form. The zero curve version of this was derived by integrating the forward curve from 0 to τ then dividing by τ i.e.

$$y(\tau) = 1/\tau \int_0^\tau f(x)dx.$$

which is easily solved to give the Nelson-Siegel yield curve model as

$$y(\tau) = \beta_L + \beta_S \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} \right) + \beta_C \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} - e^{-\lambda \tau} \right).$$

Here, $\beta_L = \beta_0$, $\beta_S = \beta_1$, $\beta_C = \beta_2$ and λ is a constant. In practice, rates are obtained as random variables with noise. Thus

$$Y_t(\tau) = y_t(\tau) + \epsilon_t(\tau), \ \epsilon(\tau) \sim N(0, \sigma_\epsilon).$$

Note that this is a cross-sectional model so the β terms are constant.

2.4.1 Characteristics of Nelson-Siegel Model

The NSM has been studied extensively heretofore, for an exhaustive analysis of the NSM see Diebold and Li (2005). We bring to light only some of the key findings of that work here.

It is clear from the above representation that the static model has one variable τ, corresponding to bond maturities and four independent parameters λ, β_L, β_S and β_C. The solution for these parameters occurs under a multivariate setting where there will be a series of noisy bond yield data corresponding to various tenors (τ).

In the model framework, the multiple yields are captured in $Y(\tau)$. The yields generally cluster around the short maturities instead of longer ones, reflecting that most trading activity in interest rate securities occurs with shorter tenors.

The fact that the model is derived as a solution of a second order differential equation implies that the numerous possible permutations of the β parameters allow the curve to take extremely different shapes. From the perspective of economics this is vital because the term structure is a construct of interest rate expectations. Thus the shape of the yield curve changes drastically over time as a result of multiple contributing elements including demand and supply of money. The term structure typically has an upward slope since longer term bonds usually offer higher yields. However, it is possible under certain economic conditions to have the contrary where longer term rates are lower than short term rates. Moreover, humps and troughs are also possible in the mid-rates depending on market expectations.

The β parameters are essential in determining the shape of the term structure. In literature they are known to be the level, slope and curvature of the term structure respectively. The coefficients of each of these terms gives some insight about their effect on the yield curve. The coefficient of β_L for instance, is 1 hence it is called the level or long-term factor. The coefficient of β_S is $\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} \right)$ however , and is 1 when $\tau = 0$ and rapidly decreases to 0 monotonically. Lastly, the coefficient of the third term, β_C is $\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right)$. This term begins at 0 when $\tau = 0$ and increases to a maximum before decaying to 0 for longer maturities (Veerhuis, 2011). These effects are illustrated graphically in Figure 2.2 below.

2.4.2 Arbitrage Free Nelson Siegel

In this section we provide a derivation of the three factor AFNS as outlined by Christensen *et al.* (2011). For convenience, we denote $A(\tau)$ and $B(\tau)$ as $A(t,T)$ and $B(t,T)$ respectively (recall that $\tau = T - t$).

Following from section 2.3, bond yields can be written as

$$y(t,T) = -\frac{B(t,T)}{T-t}X_t - \frac{A(t,T)}{T-t}.$$

Given a three-factor affine term structure model with $X_t = (X_t^1, X_t^2, X_t^3)$, the model most compatible (in terms of factor loadings in the yield equation) with the Nelson-Siegel yield function has a yield function of the form

$$y(t,T) = X_t^1 + \left(\frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right)X_t^2 + \left(\frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right)X_t^3 - \frac{A(T-t)}{T-t},$$
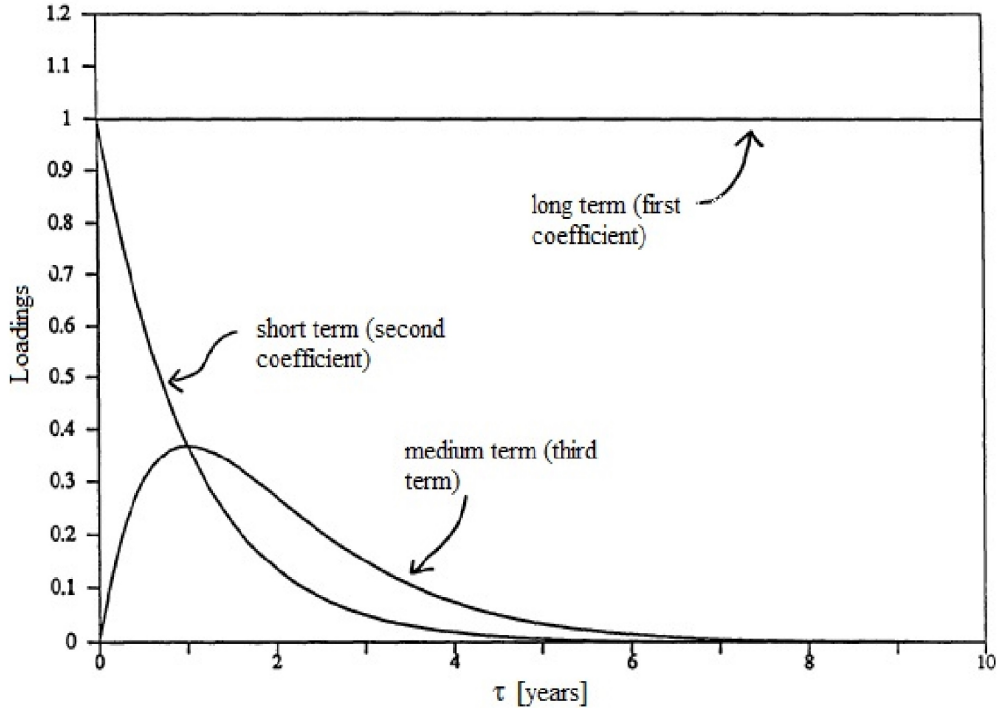
Fig. 2.2: Factor loadings for the three-factor NSM with fixed λ (Nelson and Siegel 1984)

with ODE solutions for the $B(t,T)$ matrix being

$$B^1(t,T) = -(T-t)$$
$$B^2(t,T) = -\frac{1-e^{-\lambda(T-t)}}{\lambda}$$
$$B^3(t,T) = -\frac{1-e^{-\lambda(T-t)}}{\lambda} + (T-t)e^{-\lambda(T-t)}.$$

At this point it is clear that the variable loadings exactly match Nelson-Siegel. However there is a the "yield-adjustment term" ($\frac{A(t,T)}{T-t}$) which is only a function of τ ($\tau = T - t$) and distinguishes this yield equation from Nelson-Siegel.

To show that the class of affine models that satisfies these ODEs exists, Christensen, Diebold and Rudebusch (2011) make the following argument.

Proposition 2.1. *Suppose that the short rate is*

$$r_t = X_t^1 + X_t^2,$$

where the risk-neutral dynamics of the underlying factors are given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 - X_t^1 \\ \tilde{\theta}_2 - X_t^2 \\ \tilde{\theta}_3 - X_t^3 \end{pmatrix} dt + \Sigma \begin{pmatrix} \widetilde{dW_t^1} \\ \widetilde{dW_t^2} \\ \widetilde{dW_t^3} \end{pmatrix}, \lambda > 0.$$

Then the zero-coupon bond prices are

$$P(t,T) = \exp(A(t,T) + B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3),$$

where $B^1(t,T)$, $B^2(t,T)$, $B^3(t,T)$ and $A(t,T)$ satisfy the following system of ODEs

$$\begin{aligned} dB(t,T) &= \left(\frac{dB^1(t,T)}{dt} \quad \frac{dB^2(t,T)}{dt} \quad \frac{dB^3(t,T)}{dt} \right) \\ &= \left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right]' \end{aligned} \tag{2.2}$$

and

$$\frac{dA(t,T)}{dt} = -B(t,T)\tilde{\kappa}\tilde{\theta} - \frac{1}{2}\sum_{i=1}^{n}[\Sigma'B(t,T)]_i^2$$

with boundary conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = A(T,T) = 0$. The solution for this system is

$$\begin{aligned} B^1(t,T) &= -(T-t) \\ B^2(t,T) &= -\frac{1 - e^{-\lambda(T-t)}}{\lambda} \\ B^3(t,T) &= -\frac{1 - e^{-\lambda(T-t)}}{\lambda} + (T-t)e^{-\lambda\tau}. \end{aligned}$$

and

$$A(t,T) = (\tilde{\kappa}\tilde{\theta})_2 \int_t^T B^2(s,T)ds + (\tilde{\kappa}\tilde{\theta})_3 \int_t^T B^3(s,T)ds + \frac{1}{2}\sum_{i=1}^{n}\int_t^T [\Sigma'B(s,T)]_i^2 ds.$$

Finally, we can represent bond yields as

$$y(t,T) = X_t^1 + \left(\frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} \right)X_t^2 + \left(\frac{1 - e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)} \right)X_t^3 - \frac{A(t,T)}{T-t},$$

Proof. See Appendix A.

Chapter 3

Estimating Multi-factor Affine Term Structure Models

The short rate process has significant autocorrelation therefore we use a vector autoregressive process to simulate the dynamics of the underlying factors. In this respect, the Kalman filter has enjoyed much use especially in Gaussian affine term structure models owing to their Gaussianity and their defining linear relationship between yields and risk factors (Pitsillis, 2015). Drawing from the discussions by Duffee and Stanton (2012), Pitsillis (2015) and Veerhuis (2011), our exposition of the Kalman filter begins with an overview of the filtering problem in the most general sense. This is merely an overture and is therefore a sparse section that serves to contextualise the Kalman Filter as an estimation tool. Following the presentation of the filtering problem, we derive the steps of the filter specifically for our purposes. To complete the discussion, the derivation of the likelihood function is also included.

In state space form, the dynamics of the underlying factors can be denoted as

$$X_{t+1} = \Phi X_t + C + \nu,$$

where

$$\nu \sim N(0, \Sigma).$$

This representation provides a useful way to model the time evolution of the system. Having employed a state space model, a process that enables the optimisation of unknown parameters is required. The optimisation is carried out through filtering, an approach which we discuss below.

3.1 The Filtering Problem

The filtering problem seeks to address the evolving state process X_t, $t \in \mathbb{N}$ given by

$$X_t = g_t(X_{t-1}, \eta_{t-1})$$

where $g_t : \mathbb{R}^{n_X} \times \mathbb{R}^{n_\eta} \to \mathbb{R}^{n_X}$ is a function of the state X_t and of η_{t-1}, an iid process. Given the context of this investigation we specifically wish to sequentially estimate the stochastic process

$$r_t = \sum_{i=1}^{N} X_t^i,$$

(i.e. the short rate) from the measurements given by

$$Y_t = h_t(X_t, \epsilon_t)$$

where $h_t : \mathbb{R}^{n_X} \times \mathbb{R}^{n_\epsilon} \to \mathbb{R}^{n_z}$, ϵ_{t-1} is also a function of the state X_t and the iid process ϵ_t.

The filter is designed to sequentially estimate X_t at some time s given measurements up to time t, i.e. given $Y_{0:t}$— here $0 : t$ indicates the measurements from time zero up to t. This requires that a probability density function describing the problem $p(X_t|Y_{0:t})$ be defined. The approach therefore makes two assumptions to this effect. The first being that the initial value $p(X_0|Y_0) \equiv p(X_0)$ is equal to Y_0. The second assumption is that X_t is Markov, making X_{t-1} a sufficient statistic of the history of X_t. Likewise Y_{t-1} possesses all the information required to calculate X_t and Y_t hence,

$$p(X_t|X_{0:t-1}, Y_{0:t-1}) = p(X_t|X_{t-1}).$$

Moreover,

$$p(Y_t|X_{0:t}, Y_{0:t-1}) = p(Y_t|X_t).$$

In essence, the filtering problem is solved by predicting the state at the next time step given our current knowledge and then updating our prediction once additional measurements are available. We therefore consider the prediction and update steps to preface the Kalman filter.

3.2 Prediction

To predict the state vector at time t given the information available up to time $t-1$ we employ the probability density function $p(X_{t-1}|Y_{t-1})$. The prediction density

function can then be obtained through the Chapman-Kolmogorov equation.

$$p(X_t|Y_{t-1}) = \int p(X_t, X_{t-1}|Y_{t-1})dX_{t-1}$$

$$= \int p(X_t|X_{t-1}, Y_{t-1})p(X_{t-1}|Y_{t-1})dX_{t-1}$$

$$= \int p(X_t|X_{t-1})p(X_{t-1}|Y_{t-1})dX_{t-1}$$

3.3 Update

To update we use the new measurements at time t i.e. Y_t. Using Bayes' rule we obtain posterior density of the current state

$$p(X_t|Y_t) = \frac{p(Y_t|X_t)p(X_t|Y_{t-1})}{p(Y_t|Y_{t-1})}$$

where the normalising constant

$$p(Y_t|Y_{t-1}) = \int p(Y_t|X_t)p(X_t|Y_{t-1})dX_t$$

depends on the likelihood function $p(Y_t|X_t)$ which is defined by the measurement equation h_t and known statistic ϵ_t.

3.4 The Kalman filter

For the Kalman filter the posterior density is assumed to be Gaussian and thus fully characterised by the mean and covariance. Moreover, since our scope extends only to Gaussian models, we note that it can be shown that $p(X_t|Y_t)$ is also Gaussian when the following are true:

- $p(X_{t-1}|Y_{t-1})$ is Gaussian;

- η_{t-1} and ϵ_t are both sampled from Gaussian; distributions with known mean and covariance;

- $g_t(X_{t-1}|\eta_{t-1})$ is a known linear function; and

- $h_t(X_t|\epsilon_t)$ is also known and linear.

Therefore, in state spaced form

$$Y_t = Z_t X_t + \epsilon_t$$
$$X_t = T_t X_{t-1} + \eta_t$$

Where Z_t and T_t are matrices defining the linear functions ϵ_t and η_t are assumed to be Gaussian noise.

In affine term structure modelling where the set of parameters ρ is known, we observe a set of m zero-coupon bond yields as the measurement in order to update predictions of n underlying factors. The linear function which relates the risk factors to the measurements is

$$Y_t = D_t(\rho) + Z_t(\rho)X_t + \epsilon_t.$$

The prediction step is

$$X_t = C_t(\rho) + F_t(\rho)X_{t-1} + \eta_t.$$

Usually ϵ_t is assumed to be a m dimensional serially independent Gaussian white noise with covariance matrix $H(\rho)$. This noise is the measurement error and represents the variation not captured by the model as well as liquidity effects and micro-structure noise. η_t with covariance matrix $Q(\rho)$, represents the innovations process governing the dynamics of the risk factors.

Note that there are no restrictions on the temporal dynamics of the model so that the model is time varying in all parameters. However, we concern ourselves only with the Vasicek and AFNS case where all the parameters ρ, are constant with respect to time. This specificity permits us to re-depict our model simply as

$$Y_t = D + ZX_t + \epsilon$$
$$X_t = C + FX_{t-1} + \eta_t.$$

Here D and Z depend on risk neutral dynamics and bond tenors while C and F depend on real world parameters and the observation time step.

The Kalman filter provides the minimum mean square error estimate of X_t. We denote the mean estimate of X_t by $X_{t|t}$ and the Variance by $P_{t|t}$. In light of our assumptions, X_t is multivariate normal and we assume without loss of generality that

$$X_{0|0} = \mathbb{E}[X_t]$$
$$P_{0|0} = \mathbb{V}ar[X_t].$$

The subsequent step-ahead predictions $X_{t+1|t}$ with covariance $P_{t+1|t}$ are obtained from the transition equation as

$$X_{t+1|t} = \mathbb{E}[X_{t+1}|X_t = X_{t|t}]$$
$$= C + FX_{t|t}$$

$$P_{t+1|t} = \mathbb{V}ar[X_{t+1}|X_t = X_{t|t}]$$
$$= FP_{t|t}F' + Q.$$

This in turn allows the forecast of the measurements as

$$Y_{t+1|t} = \mathbb{E}[Y_{t+1}|X_t = X_{t|t}]$$
$$= D + ZX_{t+1|t},$$

with covariance

$$V_{t+1|t} = \mathbb{V}ar[Y_{t+1}|X_t = X_{t|t}]$$
$$= ZP_{t+1|t}Z' + H.$$

Having calculated the forecasts, we now direct our focus to reconciling the predictions with the actual observation at $t + 1$. When new information is available at time $t + 1$ we can calculate the error in the prediction as given by

$$d_{t+1} = Y_{t+1} - Y_{t+1|t}.$$

Finally, we can then update our estimates of X_{t+1} as follows[1]

$$X_{t+1|t+1} = X_{t+1|t} + P_{t+1|t}Z'V_{t+1|t}^{-1}d_{t+1}$$
$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}Z'V_{t+1|t}^{-1}ZP_{t+1|t}.$$

3.5   Maximum Likelihood Estimation

When the model parameters are not known, the log-likelihood is used as an objective function that can be optimised numerically in order to estimate the parameters that are implied by the zero-coupon bond yield data. As evident from the discourse of Chapter 2, maximum likelihood estimation is perhaps the distillation of our study. Therefore this discussion would be incomplete without making pronounced some of the central elements that relate to optimising the likelihood function.

Owing to their limiting properties, maximum likelihood estimators (MLEs) are invoked continually in the theory and practice of statistics. A thorough study of the theory of statistical inference and MLEs is given in the book by Millar (2011). Under our assumptions the conditional density function corresponding to the likelihood is defined as

$$f(y_{t+1}|y_{t+1|t}) = \left[(2\pi)^N|V_{t+1|t}|\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}d_t'V_{t+1|t}^{-1}d_t\right].$$

---

[1] There are numerous equivalent formulations, see Lund (1997) for example

The associated log-likelihood is then

$$l(\rho) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|V_{t+1|t}| + d_t'V_{t+1|t}^{-1}d_t\right),$$

so that the parameter estimates are

$$\hat{\rho}(y) = \max_{\rho} l(\rho, y).$$

The parameter estimates $\hat{\rho}(y)$ are dependent on observations Y and are thus stochastic. If the distribution of the parameter estimates is sufficiently regular, MLEs show some desirable qualities as the sample size approaches infinity. One of the characteristics is the fact that $\hat{\rho}(y)$ converges to the true parameter set ρ_0 i.e. they are consistent. Another useful property is the Gaussian limiting distribution (Pitsillis, 2015) so that

$$\mathbb{P}(\sqrt{T}(\hat{\rho}(y) - \rho_0)) \to \mathbb{P}(\mathcal{N}(0, \mathcal{I}(\rho_0)^{-1}))$$

where $\mathcal{I}(\rho_0)$ is the Fischer information matrix. This matrix is defined as

$$\mathcal{I}(\rho_0)_{ij} = \mathbb{E}\left[\left(\frac{\partial}{\partial\rho_i}\log f(Y|\rho)\right)\left(\frac{\partial}{\partial\rho_j}\log f(Y|\rho)\right)\right]\Bigg|_{\rho=\rho_0}.$$

Under certain regularity conditions this becomes

$$\mathcal{I}(\rho_0)_{ij} = -\mathbb{E}\left[\left(\frac{\partial^2}{\partial\rho_i\partial\rho_j}\log f(Y|\rho)\right)\right]\Bigg|_{\rho=\rho_0}.$$

This is clearly the information matrix because its inverse is the covariance matrix of the limiting distribution as shown above. This means that with a large enough Fischer information matrix, this covariance matrix will be zero. Obviously this will only occur when we have all the information so that $\hat{\rho} = \rho_0$ with certainty. Furthermore, $\mathcal{I}(\rho_0)$ has the geometric interpretation of being the curvature. That is, apart from some pathological cases, the greater the (negative) curvature of the likelihood function around ρ_0, the more information there is about the true parameters making estimation more accurate.

The optimisation of the likelihood function requires the calculation of the first derivative and Fisher information matrix which is an arduous exercise. However Harvey (1989) (Veerhuis, 2011) illustrates that the information matrix is simply

$$\mathcal{I}_{ij} = \frac{1}{2}\sum_t\left[V_{t+1|t}^{-1}\frac{\partial V_{t+1|t}}{\partial\rho_i}V_{t+1|t}^{-1}\frac{\partial V_{t+1|t}}{\partial\rho_j}\right]' + \mathbb{E}\left[\sum_t\left(\frac{\partial d_t}{\partial\rho_i}\right)'V_{t+1|t}^{-1}\frac{\partial d_t}{\partial\rho_j}\right],$$

and we get an asymptotically equivalent result when we drop the expectation. We then maximise the likelihood function by setting the scores

$$U(\rho) = \frac{\partial l(\rho|Y)}{\partial \rho}$$

to zero. The score vector can be computed as

$$U(\rho) = \frac{1}{2}\sum_t \left\{ \left[V_{t+1|t}^{-1} \frac{\partial V_{t+1|t}}{\partial \rho_i} \right] (I - V_{t+1|t}^{-1} d_{t+1} d_{t+1}') + 2\frac{\partial d_{t+1}'}{\partial \rho_i} V_{t+1|t}^{-1} d_t \right\}.$$

Here the score vector and Fisher information matrix are represented explicitly, however the new terms $\dfrac{\partial V_{t+1|t}}{\partial \rho_i}$ and $\dfrac{\partial d_t}{\partial \rho_i}$ need further exploration. They are expressed as

$$\frac{\partial d_t}{\partial \rho_i} = -Z\frac{\partial X_{t+1|t}}{\partial \rho_i} - \frac{\partial Z}{\partial \rho_i}X_{t+1|t} - \frac{\partial D}{\partial \rho_i}$$

$$\frac{\partial V_{t+1|t}}{\partial \rho_i} = \frac{\partial Z}{\partial \rho_i}P_{t+1|t}Z' + Z\frac{\partial P_{t+1|t}}{\partial \rho_i}Z' + ZP_{t+1|t}\frac{\partial Z'}{\partial \rho_i} + \frac{\partial H}{\partial \rho_i}$$

Furthermore we need expressions for $\dfrac{\partial X_{t+1|t}}{\partial \rho_i}$ and $\dfrac{\partial P_{t+1|t}}{\partial \rho_i}$. These are simply

$$\frac{\partial X_{t+1|t}}{\partial \rho_i} = \frac{\partial C}{\partial \rho_i} + \frac{\partial F}{\partial \rho_i}X_{t|t} + F\frac{\partial X_{t|t}}{\partial \rho_i}$$

$$\frac{\partial P_{t+1|t}}{\partial \rho_i} = \frac{\partial F}{\partial \rho_i}P_{t|t}F' + F\frac{\partial P_{t|t}}{\partial \rho_i}F + FP_{t|t}\frac{\partial F'}{\partial \rho_i} + \frac{\partial Q}{\partial \rho_i}\frac{\partial X_t}{\rho_i}$$

$$\frac{\partial X_{t|t}}{\partial \rho_i} = \frac{\partial X_{t|t-1}}{\partial \rho_i} + \frac{\partial P_{t|t-1}}{\partial \rho_i}Z'V_{t|t-1}^{-1}d_t + P_{t|t-1}\frac{\partial Z'}{\partial \rho_i}V_{t|t-1}^{-1}d_t$$

$$= P_{t|t-1}Z'\frac{\partial V_{t|t-1}^{-1}}{\partial \rho_i}d_t + P_{t|t-1}Z'V_{t|t-1}^{-1}\frac{\partial d_t}{\partial \rho_i}$$

$$\frac{\partial P_{t|t}}{\partial \rho_i} = \frac{\partial P_{t|t-1}}{\partial \rho_i} - \frac{\partial P_{t|t-1}}{\partial \rho_i}Z'V_{t|t-1}^{-1}ZP_{t|t-1} - P_{t|t-1}\frac{\partial Z'}{\partial \rho_i}V_{t|t-1}^{-1}ZP_{t|t-1}$$

$$- P_{t|t-1}Z'\frac{\partial V_{t|t-1}^{-1}}{\partial \rho_i}ZP_{t|t-1} - P_{t|t-1}Z'V_{t|t-1}^{-1}\frac{\partial Z}{\partial \rho_i}P_{t|t-1}$$

$$- P_{t|t-1}Z'V_{t|t-1}^{-1}Z\frac{\partial P_{t|t-1}}{\partial \rho_i}.$$

This concludes the derivation of the Kalman filter in general form and we can now implement this for both the Vasicek and AFNS models. Obviously this requires looping through t where the step size is the frequency of sampling from the market (e.g. daily, weekly, monthly, etc.). We conclude this chapter by a presentation of the specific state-space formulations of the Vasicek and AFNS models below.

3.6   The State-space Fomulation

To reformulate or model to state-space form, the measurement and transition equations of our system have to be defined. Our development begins with a specification of m market yields, $y(\tau)$ (corresponding to bond tenors $\tau_1, \tau_2, ..., \tau_m$) which we will observe from market data as our measurement in each time step. In general, we require 3 bond tenors for a 3-factor model however, adding more maturities introduces more cross-sectional information about the term structure at each time step. This information is especially important in specifying the price of risk (Pitsillis, 2015). The measurement equation therefore relates the measured yields to the unobservable factors in the following:

$$y_t(\tau) = -\frac{\ln P_t(\tau)}{\tau} = \frac{-A(\tau) - \sum_{j=1}^{3} B_j(\tau)y_{j,t}}{\tau}.$$

This is a general equation and applies for continuous t, but for any practical application one has to discretize t since observations are discrete. Denoting the current time step as t_i and the next as $t_{i+1} = t_i + \triangle t$, where $\triangle t$ is the sampling frequency; the measurement equation for both the Vasicek and AFNS models is

$$\begin{bmatrix} y_i(\tau_1) \\ y_i(\tau_2) \\ \vdots \\ y_i(\tau_m) \end{bmatrix} = -\begin{bmatrix} \frac{A(\tau_1)}{\tau_1} \\ \frac{A(\tau_2)}{\tau_2} \\ \vdots \\ \frac{A(\tau_m)}{\tau_m} \end{bmatrix} - \begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} & \frac{B_2(\tau_1)}{\tau_1} & \frac{B_3(\tau_1)}{\tau_1} \\ \frac{B_1(\tau_2)}{\tau_1} & \frac{B_2(\tau_2)}{\tau_2} & \frac{B_3(\tau_2)}{\tau_2} \\ \vdots & \vdots & \vdots \\ \frac{B_1(\tau_m)}{\tau_m} & \frac{B_2(\tau_m)}{\tau_m} & \frac{B_3(\tau_m)}{\tau_m} \end{bmatrix} \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ x_{3,i} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

or equivalently

$$Y_t = D_t(\rho) + Z_t X_t + \epsilon_t$$

where

$$\epsilon \sim \mathcal{N}(0, H)$$

$$H = \begin{bmatrix} h_1^2 & 0 & \dots & 0 \\ 0 & h_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_m^2 \end{bmatrix}.$$

The transition equations are depend on the dynamics of the underlying factors which are different for both models. In our case, we considered the independent

Vasicek model with a diagonal $\bar{\kappa}$ therefore the transition equation for this model is

$$
\begin{bmatrix} x_{1,i+1} \\ x_{2,i+1} \\ x_{3,i+1} \end{bmatrix} = \begin{bmatrix} \theta_1(1 - e^{\kappa_1 \triangle t}) \\ \theta_2(1 - e^{\kappa_2 \triangle t}) \\ \theta_3(1 - e^{\kappa_3 \triangle t}) \end{bmatrix} + \begin{bmatrix} e^{\kappa_1 \triangle t} & 0 & 0 \\ 0 & e^{\kappa_2 \triangle t} & 0 \\ 0 & 0 & e^{\kappa_3 \triangle t} \end{bmatrix} \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ x_{3,i} \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}
$$

or equivalently

$$
X_{t+1} = C_{t+1}(\rho) + F_{t+1}(\rho)X_t + \eta_{t+1}
$$

where

$$
\eta \sim \mathcal{N}(0, Q)
$$

$$
Q = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1}\left(1 - e^{-2\kappa_1 \triangle t}\right) & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}\left(1 - e^{-2\kappa_2 \triangle t}\right) & 0 \\ 0 & 0 & \frac{\sigma_3^2}{2\kappa_3}\left(1 - e^{-2\kappa_3 \triangle t}\right) \end{bmatrix}.
$$

The form of the transition system for the AFNS model is similar to the Vasicek since both models are Gaussian. However, a more general representation of the transition equation of Gaussian models is required because although the covariance matrix is diagonal, the κ matrix is not. In general, the AFNS state transition equation is

$$
X_{t+1} = (I - \exp(-\bar{\kappa} \triangle t))\bar{\theta} + \exp(-\bar{\kappa} \triangle t)X_t + \eta_{t+1}
$$

or equivalently

$$
X_{t+1} = C_{t+1}(\rho) + F_{t+1}(\rho)X_t + \eta_{t+1}
$$

where I is the identity matrix with the same size as $\bar{\kappa}$ and

$$
\eta \sim \mathcal{N}(0, Q)
$$

$$
Q = \int_0^{\triangle t} e^{-\bar{\kappa}s} \Sigma\Sigma' e^{-\bar{\kappa}'s} ds.
$$

Since the Kalman filter is recursive, one needs to find an appropriate starting values for the recursion. We use the unconditional mean and variance or our system as initialisations. Thus

$$
X_{0|0} = \bar{\theta}
$$

for both models whereas

$$
P_{0|0} = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1} & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2} & 0 \\ 0 & 0 & \frac{\sigma_3^2}{2\kappa_3} \end{bmatrix}
$$

for the Vasicek model and

$$
P_{0|0} = \int_0^{\infty} e^{-\bar{\kappa}s} \Sigma\Sigma' e^{-\bar{\kappa}'s} ds
$$

for the AFNS model. This concludes the chapter.

Chapter 4

Simulation Based Study of Parameter Estimates

When simulation studies of the Kalman filter are done– where the underlying state process is simulated from a known set of parameters– the original parameters are rarely recovered for multi-factor models. The aim of this study therefore, is to perform numerical experiments in order to find out why, and whether this matters. The key questions which were considered in order to achieve this aim were as follows.

- Are there multiple parametrizations of a model that lead to essentially the same model?

- Does the quasi-log likelihood function that is maximised have many essentially equivalent local maxima?

- Given that we are unable to obtain the original parameters for the state process, how does this affect our ability to price options (bond options)?

To answer these questions a time series of zero-coupon bond yields of different maturities was simulated from a known set of parameters and these were considered as the measurement data excluding noise. A measurement noise matrix was then specified (also known) in order to account for the measurement error inherent in real data. The likelihood function was then optimised multiple times from a randomly generated set of starting parameters within the feasible parameter space.

The optimisation was successful for both of the considered models and it was found that the original parameters were not recovered (entirely). The AFNS model however recovers all but one of the parameters. Furthermore, the quasi-log likelihood function seems to have multiple maxima corresponding to each of the disparate parameters. Monte Carlo approximations of bond call options were done for each parameter set to illustrate the effects on option pricing. Since the primary

focus of the study was the effect of disparate parameter sets on option pricing, the price of risk was set to zero in each case.

4.1   Yields and the Kalman Filter

Simulations were performed using a time step of 0.01 years ($\triangle t = 1/100$ yrs) and the sampling period was 30 years (T = 30 yrs). In every case, the bond tenors that were employed were 3 months, 6 months, 1 year and 5 years. Using these tenors and a set of arbitrary but feasible parameters, paths of the underlying variables were simulated following which, bond yields were calculated. Then Gaussian white noise was added to these yields as measurement error. The noise standard deviation was chosen to be equal for all the yields. This results in some loss of generality. However, the yields would be "measured" from the same market. So it should be plausible to suppose that they would be subject to the same mechanisms that introduce the measurement error and therefore that this error is similar.

The resulting time series of bond yields with noise y_k, $k \, \epsilon \, \{1, ..., n\}$ ($n = T/\triangle t$) was passed into an optimisation routine to search the feasible parameter space for the original parameters that were used to simulate the yields. In each iteration the Kalman filter is called with the specified starting parameters (a feasible random guess) then the quasi-log likelihood value is calculated and the optimiser iterates through until the function maximising parameters are identified[1]. Ten sets of initial parameters were generated randomly and for each set, fifteen optimisations were conducted where the resulting parameters including the likelihood values corresponding to each of the fifteen runs were calculated. The average of these parameters as well as the average likelihood value was recorded as the final result. This means that for ten different starting values, ten sets of parameters with respective likelihood values result. The optimisation was performed using the sqp algorithm in the constrained non-linear programming solver fmincon in Matlab[2].

---

[1] Most optimisers minimise the function thus one may need to optimise the negative of the objective function.

[2] More information regarding this solver may be found in: http://www.mathworks.com/help/optim/ug/fmincon.html?requestedDomain=www.mathworks.com

4.2   Single Factor Vasicek

This methodology was applied first to the single factor Vasicek model. The results as displayed in Table 4.1 show that the original parameters have been identified successfully. The low standard deviations also show that these parameters were recovered with high precision. Furthermore, Figure 4.1 clearly shows that the filtered path approximates the simulated path satisfactorily for the one-factor Vasicek model. In this case there were no problems with recovering the original parameters. These results are hardly surprising because one factor models are not known to be susceptible to the problem of having non-unique parameter sets that maximise the likelihood function. The problem is endemic to multi-factor models.

Tab. 4.1: One factor Vasicek model parameter estimation results

Parameters	True value	Mean Est.	Std. Dev
κ	0.6000	0.6003	0.0007
θ	0.1000	0.1000	0.0000
σ	0.0200	0.0202	0.0009
ϵ	0.0001	0.0001	0.0000
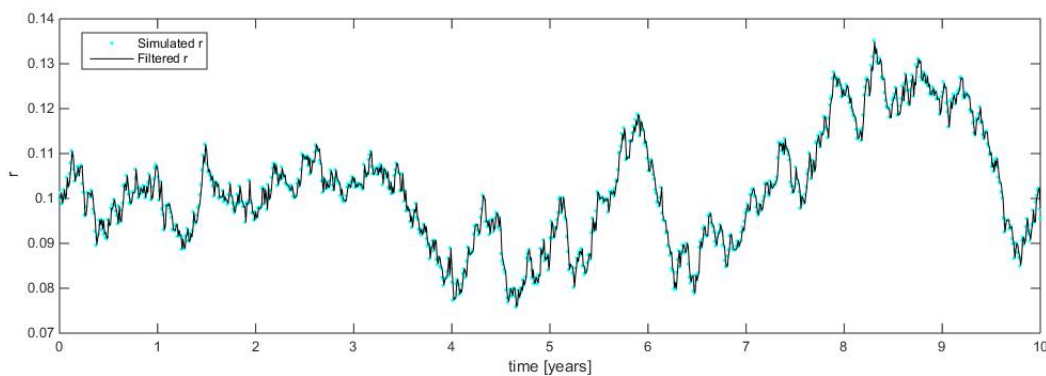
Fig. 4.1: Simulated and Filtered states for one factor Vasicek— this graph was constructed by simulating the short rate from a known set of parameters shown in Table 4.1 and then the mean estimates from the optimisation were used to generate the filtered paths from bond yields.

4.3 Three-Factor Vasicek Model

Many optimisations of the independent Vasicek model were performed but to avoid redundancy we present a summary of the results. Table 4.2 displays some of the permutations of recovered parameters from the optimisation. The disparities between the parameter sets in this table are immediately obvious. Clearly none of the recovered parameters match the original ones except for the measurement error. Furthermore, we see from these results that although the parameters are different, the value of the likelihood function is very similar in all cases particularly for Runs 1,4,6 and 7. This suggests one or a combination of two things. The first being that the likelihood function has local maxima that are similar to the global maximum and the second being that the global maximum is not given uniquely by one parameter set. The latter case may further insinuate that there are multiple parametrisations even for a single model. On the other hand, the former may mean that the although the likelihood function has a global maximum corresponding to a unique set of parameters, the numerical techniques used in the optimisation find a local maximum which may in some instances result in a model that differs in behaviour when compared to the actual parameters.

Moreover, Figures 4.3 and 4.4 show a comparison of simulated and filtered short rate and underlying factors respectively. We used the original parameters to simulate the underlying factors from which we then calculated the short rate as well as the corresponding zero-coupon bond yields with added white noise. Figure 4.3 was then constructed by filtering through these yields using the original parameters and the rest of the parameters in Table 4.2 for the short rate; Figure 4.4 shows the results of filtering through the same yields but for the unobservable factors. What is clear from these figures is that the Kalman filter does not capture the underlying processes (X_is) correctly but the short rate is captured well in all cases. This is simply because the filter is designed to approximate the short rate so the filtered underlying processes are not necessarily the same as the simulated ones but the short rate process always resembles the simulated short rate.

Since $r_{simulated} = X_{simulated}^1 + X_{simulated}^3 + X_{simulated}^2$, a filter that correctly captures all the underlying processes will surely capture the short rate process as well. So if the short rate can be correctly filtered out even when the underlying processes are not, there should be at least two ways of representing it. In fact the short rate r can be expressed as follows

$$r(t) = X_1 + X_2 + X_3$$

Where

$$X_i(t) \sim N(X_i(0)e^{-\kappa_i t} + \theta_i(1 - e^{-\kappa_i t}), \Phi_i(t))$$

$$\Rightarrow r(t) \sim N(\sum_{i=1}^{i=3}(X_i(0)e^{-\kappa_i t} + \theta_i(1 - e^{-\kappa_i t})), \sum_{i=1}^{i=3} \Phi_i(t))$$

$$\Rightarrow r(t) = \sum_{i=1}^{i=3}(X_i(0)e^{-\kappa_i t} + \theta_i(1 - e^{-\kappa_i t})) + \sqrt{\sum_{i=1}^{i=3} \Phi_i(t)}Z$$

Where $Z \sim N(0, 1)$ and $\Phi_i(t) = \frac{\sigma_i^2}{2\kappa_i}(1 - e^{2\kappa_i t})$.

As one can imagine, there will be a numerous ways of arriving at the same short rate i.e. whatever set of parameters yields the drift $\sum_{i=1}^{i=3}(X_i(0)e^{-\kappa_i t} + \theta_i(1 - e^{-\kappa_i t}))$ and volatility $\sqrt{\sum_{i=1}^{i=3} \Phi_i(t)}$ will suffice. Suppose for example that $X_i(0) = \theta_i$ then $r(t) = \sum_{i=1}^{i=3} \theta_i + \sqrt{\sum_{i=1}^{i=3} \Phi_i(t)}Z$ in which case $\theta_1 = 0.1$, $\theta_2 = 0.2$, $\theta_1 = 0.3$ would an short rate equivalent to $\theta_1 = 0.2$, $\theta_2 = 0.2$, $\theta_1 = 0.2$ if the volatilities are kept the same. Likewise $\Phi_1(t) = 0.02$, $\Phi_2(t) = 0.04$, $\Phi_3(t) = 0.02$ yields the same short rate as $\Phi_i(t) = (0.02 + 0.04 + 0.02)/3$.

Table 4.3 clearly shows that the parameters from Table 4.2 yield the same short rate, this holds for any feasible t. This is in harmony with the above analysis, i.e. there are many equivalent parameter sets that represent the same short rate. This means therefore that the global maximum is not determined uniquely by one parameter set. One should therefore not expect to consistently recover the original parameters when optimising– in fact the original parameters are rarely recovered.

The question about which are the "true" parameters is one that requires consideration. For this we remind the reader that the Kalman filter takes zero-coupon bond yields as the observable processes and such are expressed as

$$P_t^{(\tau)} = \mathbb{E}^{\mathbb{Q}}\left[exp\left(-\int_t^{t+\tau} r_u du\right) \cdot 1 \,\bigg|\, \mathcal{F}_t\right].$$

Hence given that the different parametrisations yield the same short rate process, the bond yields should also follow the same processes. The optimiser can therefore not ascertain which set of parameters was used to simulate the yields. One therefore gets the solution closest to the initial guess.

We are able to pose such a question here because this is a simulation based study where every parameter is known beforehand but in reality one should consider whether the "true" parameter set does indeed exist and if so, whether it can be recovered deliberately in the context of our model. Real data is not simulated so it is extremely difficult or even impossible to know the "data generating" parameter set. Thus it would be nonsensical to consider two parameter sets which yield the same short rate model with partiality.

We see furthermore in Figure 4.2 that for the most part, the prices are similar with only two parameter sets pricing without the error bound. Run 6 as seen in Table 4.3 has a slightly higher volatility in r_t compared to the rest and this is probably why it has a higher price. This shows that it is possible to get parameter sets

that result in different prices, depending on the precision in the optimisation. The tolerance and the maximum number of iterations used in the optimisation strongly affect the quality of the results. For obvious reasons one would prefer the highest tolerance and maximum number of iterations because this gives more accurate results. However, this was made impossible by the time constraints and the speed of the computers that could be used in this study so the optimisation is not perfect.

Tab. 4.2: Results from the three factor Vasicek model

	Original	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
Likelihood	2386322	2386322	2386318	2386316	2386321	2386216	2386322	2386321
κ_1	0.6	0.420892	0.401544	0.449136	0.368763	0.004846	0.658336	0.678350
κ_2	0.3	0.385128	0.180147	0.389024	0.411483	0.391150	0.396464	0.406237
κ_3	0.7	0.682629	0.676115	0.694639	0.683725	0.623222	0.659439	0.348437
θ_1	0.01	0.009316	0.031120	0.032404	0.007067	0.000192	0.021370	0.022176
θ_2	0.02	0.051577	0.010588	0.000281	0.014269	0.004163	0.021827	0.024996
θ_3	0.04	0.009174	0.028338	0.037380	0.048730	0.070893	0.027284	0.022883
σ_1	0.02	0.037241	0.050490	0.029934	0.021116	0.001014	0.019496	0.034857
σ_2	0.05	0.034889	0.001444	0.042366	0.046497	0.046837	0.050411	0.049291
σ_3	0.03	0.034432	0.035194	0.033131	0.034379	0.039969	0.030524	0.012095
ϵ	0.001	0.001000	0.001000	0.001000	0.001000	0.001000	0.00993	0.001000

Tab. 4.3: Drift and volatility of short rate using t = 0.01 years, $X_i(0) = \theta_i$

	Original	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
$\sum_{i=1}^{i=3}(X_i(0)e^{-\kappa_i t} + \theta_i(1 - e^{-\kappa_i t}))$	0.07000	0.07007	0.07005	0.07007	0.07007	0.07525	0.07048	0.07006
$\sqrt{\sum_{i=1}^{i=3}\Phi_i(t)}$	0.00615	0.00614	0.00614	0.00614	0.00614	0.00614	0.00619	0.00614
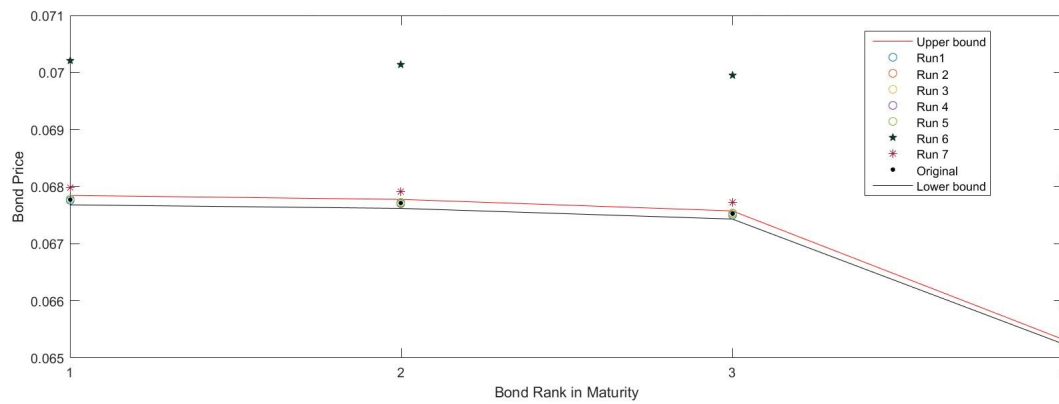
Fig. 4.2: Three-factor Vasicek bond call options— this graph shows Monte Carlo zero-coupon bond call option prices under the three-factor Vasicek model for each parameter set in Table 4.2. The horizontal axis corresponds to maturities 3m, 6m, 1yr and 5yrs as ranked in ascending maturity. 500 000 realisations were used to calculate each call price using an exercise price of $e^{-0.5(\theta_1 + \theta_2 + \theta_3)}$ where the $\boldsymbol{\theta}$ corresponds only to the original parameter set. The maturity was 6 months. Lastly, the error bounds are three standard deviations from the original parameter price.
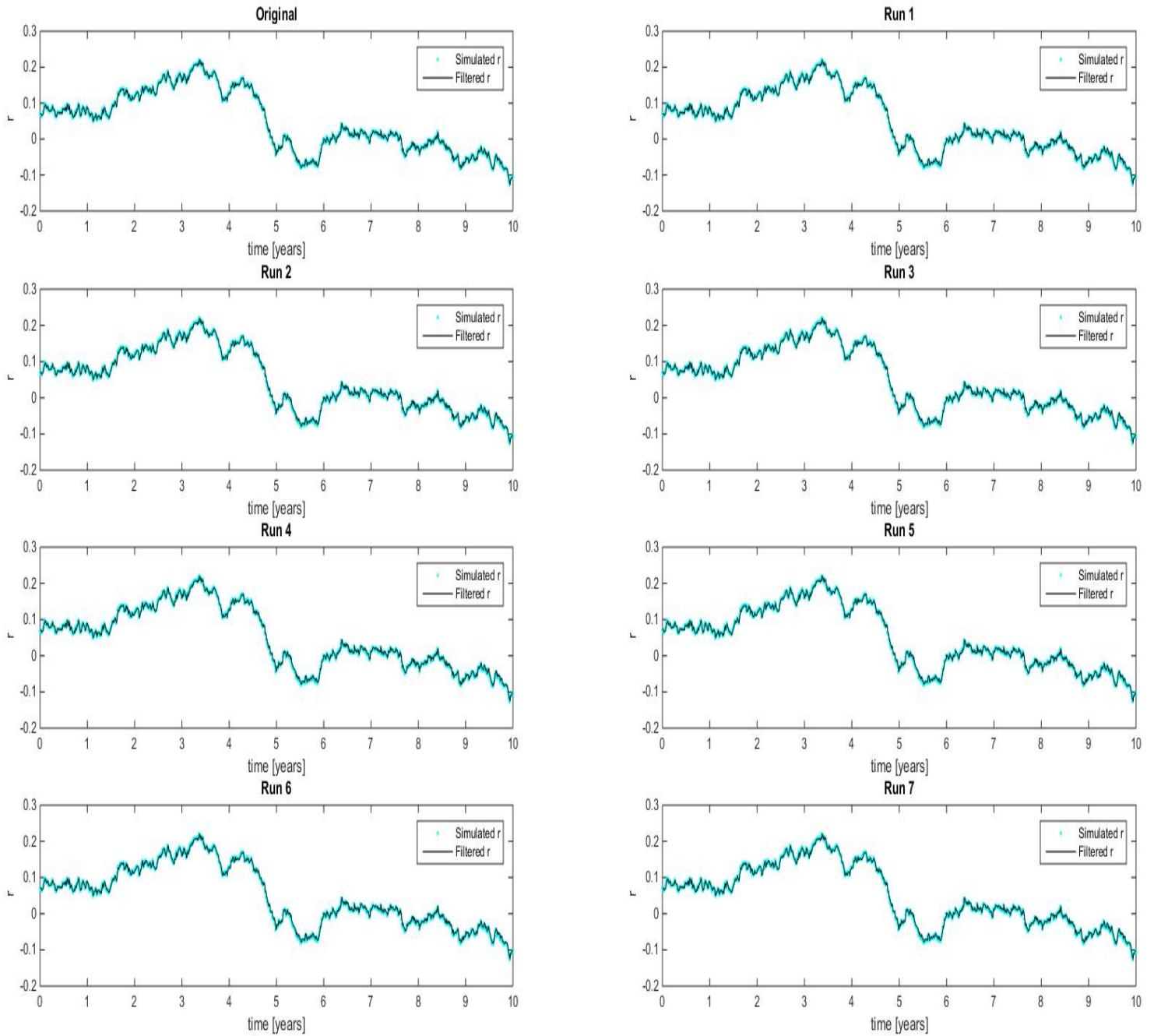
Fig. 4.3: Simulated and Filtered three-factor Vasicek Short rates.

Fig. 4.4: Simulated and Filtered three-factor Vasicek states.

4.4   The Independent Risk-Neutral AFNS

Upon examining the results of Table 4.4 one can easily notice that contrary to the Vasicek model, the AFNS optimisation has a reasonably good parameter recovery. All the parameters excluding θ_3 have been reproduced. This serves as confirmation that the AFNS model has a better recovery of parameters. This is because the model has fewer parameters therefore permutations of possible sets are sparse. Furthermore, the underlying processes of the AFNS have been identified in advance and the model is in fact constructed so that the underlying processes are exactly the level, slope and curvature. We know that X_1 is the level, X_2 is the slope and X_3 is the curvature. All of these factors are distinct and have specific and unique coefficients in the bond yield formula. To notice this, recall that for the AFNS

$$y(t, \tau) = X_1(t) + \left[\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right] X_2(t) + \left[\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right] X_3(t) - \frac{A(t, \tau)}{\tau},$$

whereas for the Vasicek

$$y(t, \tau) = \frac{B_1(t, \tau)}{\tau} X_1(t) + \frac{B_2(t, \tau)}{\tau} X_2(t) + \frac{B_3(t, \tau)}{\tau} X_3(t) - \frac{A(t, \tau)}{\tau}.$$

The $\frac{B_i(t,\tau)}{\tau}$ terms in the latter have the same form whereas the coefficients of the $X_1(t)$ terms in the former are significantly different. This has the implication that for the Vasicek model, one does not know at all what $X_1(t)$ is and how it is different to $X_2(t)$ for instance. Thus these models accommodate any set of processes as long as they model short rate and the bond prices correctly. The AFNS removes this ambiguity by having specific underlying factors so that the level does not take the place of the slope for example. The Kalman filter can distinguish between $X_1(t)$ and $X_1(t)$ because of their specific dynamics and unique relation both to the bond price and the short rate.

This is in harmony with the results shown in Figure 4.7 (*cf.* Figure 4.4) i.e. as opposed to the Vasicek model the simulated factor processes are well approximated by the filter in the AFNS. Figure 4.6 shows that the short rate is also approximated well. These figures were constructed by the same methodology used for Figures 4.3 and 4.4 above. A possible explanation to the failure in approximating $X_3(t)$ may be that the Kalman filter estimates the short rate which is only a function of $X_1(t)$ and $X_2(t)$ so that because $X_3(t)$ is only features in the bond yield equation, it is not as restricted and therefore it is not estimated well. Another likely possibility is that the optimiser found local maxima which do not have the right θ_3.

The specificity of the AFNS model allows one know that "true" parameters do indeed exist in its context. This is because two disparate sets of parameters have different economic meanings since all the factors have well defined implications

for the bond curve. The fact that all the level and slope are recovered well gives a good short rate approximation however, the mean reversion level of the curvature is poorly recovered. This inspires mixed expectations about how the bond prices generated from the optimisation parameters will compare to those of the original parameters because different curvatures can mean different yield curves while similar short rates should mean similar yield curves.

Lastly, the option prices in Figure 4.5 are all clearly within the error bound thus we conclude they are similar.

Tab. 4.4: Results from the AFNS model

	Original	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
Likelihood	84796.58	84796.41	84796.41	84796.41	84796.41	84796.41	84796.41	84796.41
λ	0.10000	0.09589	0.09577	0.09589	0.09589	0.09586	0.09589	0.09592
θ_1	0.17221	0.16743	0.17002	0.17270	0.17408	0.17460	0.17203	0.16169
θ_2	0.18587	0.18547	0.18421	0.18061	0.17992	0.18103	0.18151	0.19121
θ_3	0.06905	0.10142	0.07334	0.08765	0.041765	0.05579	0.06062	0.06276
σ_1	0.04154	0.04112	0.04111	0.04112	0.04112	0.04112	0.04112	0.04113
σ_1	0.03029	0.03066	0.03067	0.03065	0.03066	0.03066	0.03065	0.03065
σ_1	0.00890	0.01070	0.01075	0.01070	0.01070	0.01071	0.01070	0.01069
ϵ	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.00010
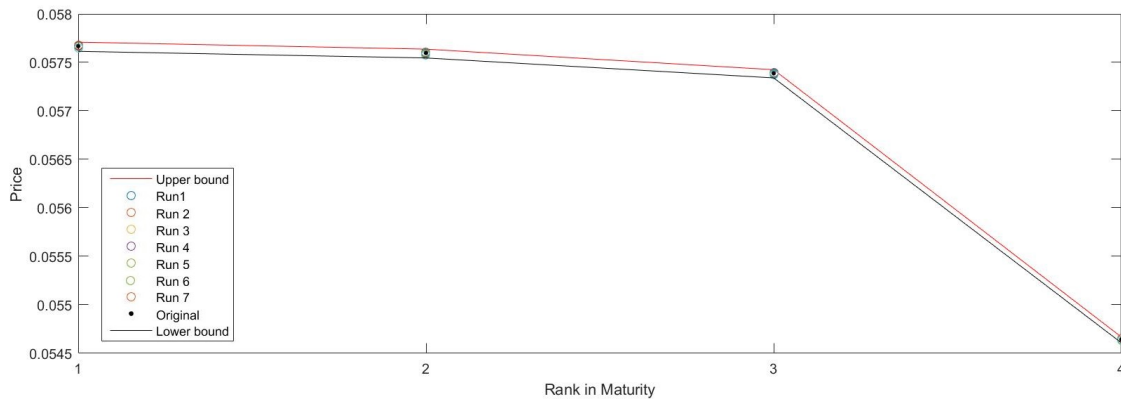
Fig. 4.5: Three-factor AFNS bond call options— this graph shows Monte Carlo zero-coupon bond call option prices under the three-factor AFNS model for each parameter set in Table 4.4. The horizontal axis corresponds to maturities 3m, 6m, 1yr and 5yrs as ranked in ascending maturity. 500 000 realisations were used to calculate each call price using an exercise price of $e^{-0.5(\theta_1+\theta_2+\theta_3)}$ where the θ corresponds only to the original parameter set. The maturity was 6 months. Finally, the error bounds are three standard deviations from the original parameter price.
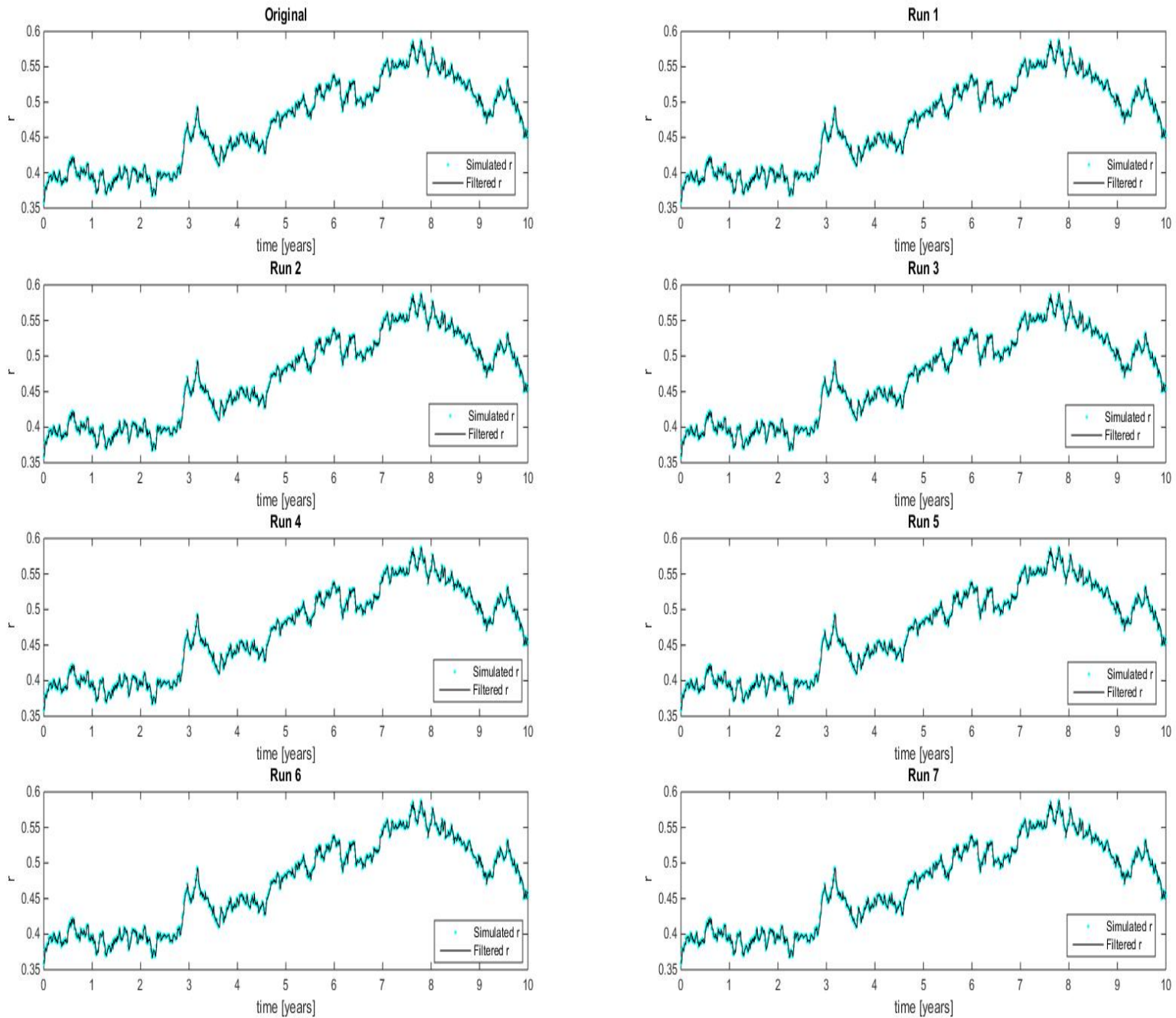
Fig. 4.6: Simulated and Filtered three-factor AFNS Short rates.
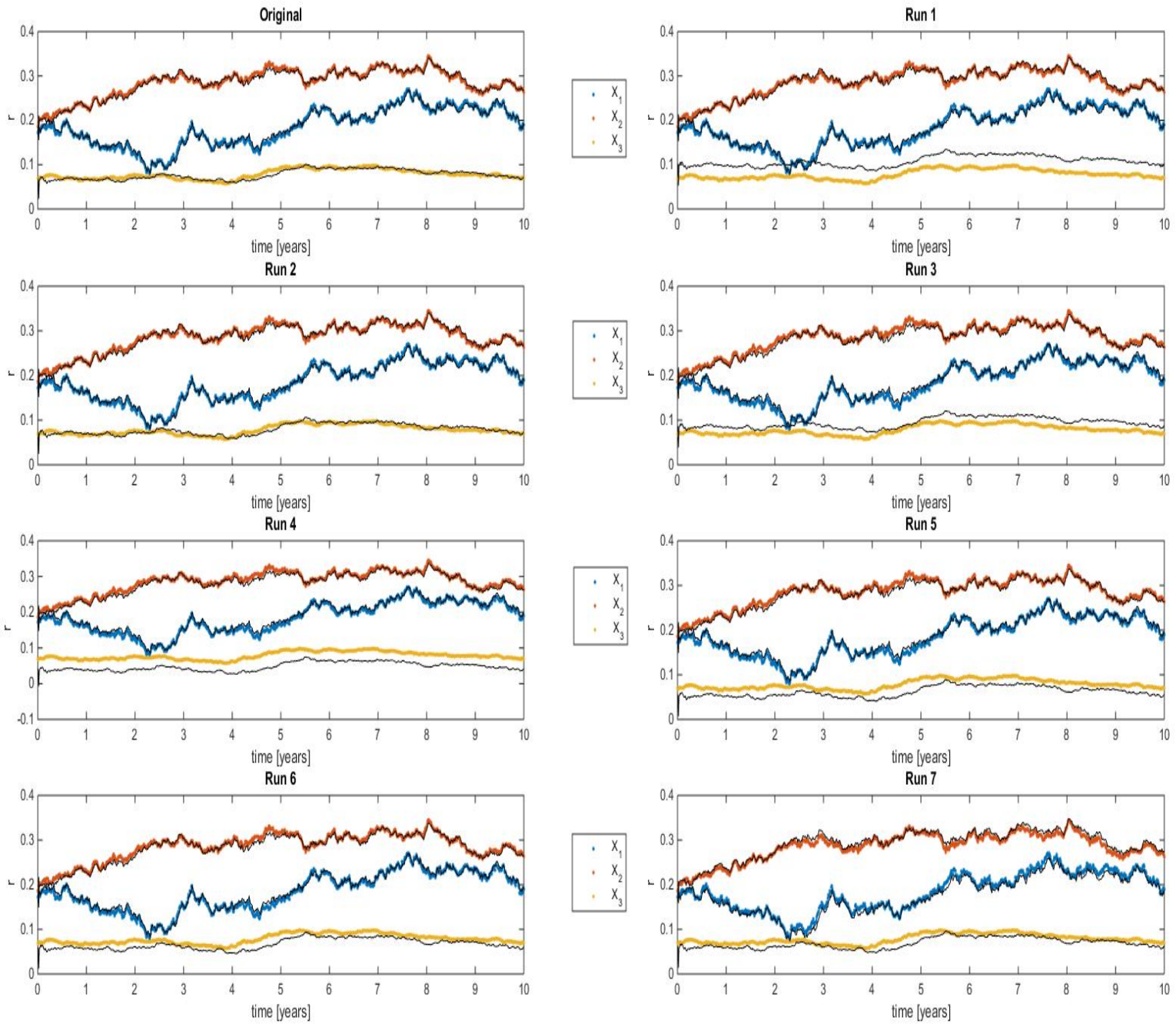
Fig. 4.7: Simulated and Filtered three-factor AFNS factors rates.

Chapter 5

Conclusions

In conclusion, the *raison d'être* of this dissertation was to employ numerical experiments to explain the existence disparate parameter estimates in multi-factor affine term structure models and to establish whether these disparities matter as far as option pricing is concerned. To guide our conclusions we answer the questions outlined at the start of the previous chapter.

The disparate parameter sets essentially lead to the same short rate process. This was demonstrated in the assessment of the three-factor Vasicek model because it resulted in different parameter recoveries. The AFNS model recovers its parameters adequately except for the long term drift of the curvature which is different in every run. It was shown that the cause of the optimisation difficulties in the Vasicek model is that the model is specified with the same functional form for each variable so that it is not known in advance what the unobservable factors are. The specification of the AFNS on the other hand, is such that every factor has a distinct functional form so there is no ambiguity about what each factor represents. Thus the AFNS precludes possibilities for different parameter sets.

The likelihood function has many essentially equivalent maxima and possibly a global maximum value corresponding to several parameter sets. This follows logically from the fact that the parameter sets are representations of the same short rate process. Since they model the same short rate process, which in turn produces the yields, the parameters should have the same likelihood.

Finally, the models generally lead to the same option prices, however it is possible to have different prices as a result of the precision of the optimisation. Therefore one should use a precise optimiser.

Bibliography

Babbs, S. H. and Nowman, K. B. (1999). Kalman Filtering of Generalized Vasicek Term Structure models, *Journal of Financial and Quantitative Analysis* pp. 115–130.

Björk, T. (2009). *Arbitrage theory in continuous time*, Oxford university press.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum Likelihood Estimation of Dynamic Models with Time-varying Covariances, *Econometric Reviews* (11): 143–172.

Chatterjee, S. (2005). Application of the Kalman Filter for Estimating Continuous Time Models: The Case of UK and Germany. Department of Economics University of Glasgow January, 2005.

Chen, R. and Scott, L. (1992). Pricing Interest Rate Options in a Two-factor Cox-Ingersoll-Ross Model of the Term structure, *Review of Financial Studies* **5**(4): 613–636.

Christensen, H. E. J., Diebold, X. F. and Rudebusch, G. D. (2011). The Affine Arbitrage-free Class of Nelson-Siegel Term Structure Models, *Journal of Econometrics* .

Cox, J. C., , Ingersoll, J. E. and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates, *Econometrica: Journal of the Econometric Society* p. 385407.

Dai, Q. and Singleton, K. (2000). Specification Analysis of Affine Term Structure Models, *The Journal of Finance* **55**(5): 1943–1978.

Dai, Q. and Singleton, K. (2003). Term Structure Dynamics in Theory and Reality, *Review of Financial Studies* **16**(3): 631–678.

De Jong, F. (2000). Time-series and Cross-section Information in Affine Term Structure Models, *Journal of Business and Economic Statistics* (18): 300–314.

Diebold, X. F. and Li, C. (2005). Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics* .

Duan, J. C. and Simonato, J. G. (1999). Estimating and Testing Exponential-affine Term Structure Models by Kalman Filter, *Review of Quantitative Finance and Accounting* **13**(3): 111–135.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models, *The Journal of Finance* **57**(1): 405–443.

Duffee, G. R. and Stanton, R. H. (2012). Estimation of Dynamic Term Structure Models, *The Quarterly Journal of Finance* **2**(2).

Duffie, D. and Kan, R. (1996). A Yield-factor Model of Interest Rates, *Mathematical Finance* **6**(4): 379–406.

Gasha, G., He, Y., Medeiros, C., Rodriguez, M., Salvati, J. and Yi, J. (2010). On the Estimation of Term Structure Models and An Application to the United States. IMF Working Paper Monetary and Capital Markets Department.

Geyer, A. L. J. and Pichler, S. (1999). A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates, *Journal of Financial Research* **22**(1): 107–130.

Gupta, H. N. and Mehra, K. R. (1974). Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations, *IEEE Transactions on Automatic Control* **19**(6): 774–783.

Kim, D. H. and Orphanides, A. (2005). Term Structure Estimation with Survey Data on Interest Rate Forecasts, *Board of Governors of the Federal Reserve System* .

Langetieg, T. C. (1980). A Multivariate Model of the Term Structure, *Journal of Finance* (35): 71–97.

Longstaff, F. A. and Schwartz, E. S. (1992). Interest Rate Volatility and the Term Structure: A Two-factor General Equilibrium Model, *The Journal of Finance* **47**(4): 1259–1282.

Lund, J. (1995). Econometric Analysis of Continuous-time: Arbitrage-free Models of the Term Structure of Interest Rates, *Technical Report* . URL: http://www.jesperlund.com/papers/kflin.pdf.

Lund, J. (1997). Non-linear Kalman Filtering Techniques for Term-structure Models, *Technical Report* . URL: http://www.jesperlund.com/papers/kfnlin.pdf.

Macaulay, F. R. *et al. (1938). Some theoretical problems suggested by the movements of interest rates, bond yields and stock prices in the united states since 1856,* NBER Books .

Maes, K. (2000). Estimating Affine Models by Kalman Filter.
URL: http://www.econ.kuleuven.ac.be/ew/academic/intecon/staff/stijn/KonstantijnMaes.htm

Millar, R. B. (2011). Maximum likelihood estimation and inference: with examples in R, SAS and ADMB, *Vol. 111, John Wiley & Sons.*

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves, Journal of business *pp. 473–489.*

Piazzesi, M. (2009). Affine Term Structure Models, Handbook of Financial Econometrics **1***: 691–766.*

Pitsillis, Z. (2015). Estimating Dynamic Affine Term Structure Models, *Master's thesis, University of Cape Town.*

Stewart, J. (2015). Calculus, *Cengage Learning.*

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure, Journal of Financial Economics **5***(2): 177–188.*

Veerhuis, P. (2011). Arbitrage-free Nelson Siegel Yield Curve Mode: An Application to Assess Unconventional Monetary Policy, *Master's thesis, University of New South Wales.*

Wilson, T. (1994). "Debunking the Myths", Risk *pp. 67–72.*

Appendix A

Proof of Proposition 2.1

A.1 Derivation of B(t,T) Term

As explained by Christensen *et al.* (2011), we begin the proof of this proposition by limiting the volatility to be constant. The system of ODEs for $B(t,T)$ then becomes

$$\frac{dB(t,T)}{dt} = \delta'_X + \tilde{\kappa}' B(t,T), \ B(T,T) = 0.$$

Since

$$\frac{d}{dt}[e^{\tilde{\kappa}'(T-t)} B(t,T)] = e^{\tilde{\kappa}'(T-t)}\frac{dB(t,T)}{dt} - \tilde{\kappa}' e^{\tilde{\kappa}'(T-t)} B(t,T)$$

we can conclude that

$$\int_t^T \frac{d}{dt}[e^{\tilde{\kappa}'(T-s)} B(s,T)]ds = \int_t^T e^{\tilde{\kappa}'(T-s)} \delta_X ds$$

moreover, when we invoke the boundary conditions it is easy to see that

$$B(t,T) = -e^{\tilde{\kappa}'(T-t)} \int_t^T e^{\tilde{\kappa}'(T-s)} \delta_X ds.$$

Now define $\tilde{\kappa}'$ and δ_X as

$$\tilde{\kappa}' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix}, \ \delta_X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

We can then show that

$$e^{\tilde{\kappa}'(T-t)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & -\lambda(T-t)e^{\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix}$$

and

$$e^{-\tilde{\kappa}'(T-t)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix}.$$

Substituting into the ODE

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

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

and because

$$\int_T^t ds = T - t,$$

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

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

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

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

the solution becomes

$$B(t,T) = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda(T-t)} & 0 \\ 0 & \lambda(T-t)e^{\lambda(T-t)} & e^{-\lambda(T-t)} \end{pmatrix} \begin{pmatrix} T-t \\ -\frac{1 - e^{\lambda(T-t)}}{\lambda} \\ -(T-t)e^{\lambda(T-t)} - \frac{1 - e^{\lambda(T-t)}}{\lambda} \end{pmatrix}$$

$$= \begin{pmatrix} -(T-t) \\ -\frac{1 - e^{-\lambda(T-t)}}{\lambda} \\ (T-t)e^{-\lambda(T-t)} - \frac{1 - e^{-\lambda(T-t)}}{\lambda} \end{pmatrix}.$$

which concludes our proof for $B(t,T)$.

A.2   Derivation of Yield Adjustment Term

$A(t,T)$ as expressed in proposition 2.1 is just a simple result of matrix multiplication for the fist term. The closed form of the yield adjustment term is long and deriving it is an arduous undertaking, we therefore present it in Appendix A.

The last term of $A(t,T)$ may be derived by equating θ to zero. In that case the yield adjustment term has this general form

$$\frac{A(t,T)}{T-t} = \frac{1}{2} \frac{1}{T-t} \int_t^T \sum_{j=1}^3 (\Sigma' B(s,T) B(s,T)' \Sigma)_{j,j} ds.$$

Given the volatility matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

the analytical solution may be evaluated as follows

$$\frac{A(t,T)}{T-t} = \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^3 \left[\begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} \\ \sigma_{12} & \sigma_{22} & \sigma_{32} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \right.$$

$$\left. \left(B^1(t,T) \quad B^2(t,T) \quad B^3(t,T) \right) \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \right]_{j,j}$$

$$= \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2 ds + \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2 ds$$

$$+ \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2 ds$$

$$+ \bar{D}\frac{1}{T-t}\int_t^T B^1(s,T)B^2(s,T) ds$$

$$+ \bar{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T) ds$$

$$+ \bar{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T) ds$$

where

$$\bar{A} = \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2,$$
$$\bar{B} = \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{33}^2,$$
$$\bar{C} = \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2,$$
$$\bar{D} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22} + \sigma_{13}\sigma_{23},$$
$$\bar{E} = \sigma_{11}\sigma_{31} + \sigma_{12}\sigma_{32} + \sigma_{13}\sigma_{33},$$
$$\bar{F} = \sigma_{21}\sigma_{31} + \sigma_{22}\sigma_{32} + \sigma_{23}\sigma_{33}.$$

The six integrals above can further be evaluated as

$$I_1 = \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T B^1(s,T)^2 ds$$

$$= \frac{\bar{A}}{2}\frac{1}{T-t}\int_t^T (T-s)^2 ds = \frac{\bar{A}}{6}(T-t)^2$$

$$I_2 = \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T B^2(s,T)^2 ds$$

$$= \frac{\bar{B}}{2}\frac{1}{T-t}\int_t^T \left[-\frac{1-e^{-\lambda(T-s)}}{\lambda}\right]^2 ds$$

$$= \bar{B}\left[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right]$$

$$I_3 = \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T B^3(s,T)^2 ds$$

$$= \frac{\bar{C}}{2}\frac{1}{T-t}\int_t^T \left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right] ds$$

$$= \bar{C}\left[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{2\lambda(T-t)} - \frac{3}{4\lambda^2}(T-t)e^{2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right.$$

$$\left. - \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right]$$

$$I_4 = \frac{\bar{D}}{T-t}\int_t^T B^1(s,T)B^2(s,T) ds$$

$$= \frac{\bar{D}}{T-t}\int_t^T (T-s)\frac{1-e^{-\lambda(T-s)}}{\lambda} ds$$

$$\bar{D}\left[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right]$$

$$I_5 = \bar{E}\frac{1}{T-t}\int_t^T B^1(s,T)B^3(s,T) ds$$

$$= \bar{E}\frac{1}{T-t}\int_t^T -(T-s)\left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right] ds$$

$$= \bar{E}\left[\frac{3}{\lambda^2}e^{-\lambda(T-1)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(t-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right]$$

$$I_6 = \bar{F}\frac{1}{T-t}\int_t^T B^2(s,T)B^3(s,T) ds$$

$$= \bar{F}\frac{1}{T-t}\int_t^T \left[-\frac{1-e^{-\lambda(T-t)}}{T-t}\right]\left[(T-s)e^{-\lambda(T-s)} - \frac{1-e^{-\lambda(T-s)}}{\lambda}\right] ds$$

$$= \bar{F}\left[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{2}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right].$$