

Estimating Dynamic Affine Term Structure Models

Zachry Pitsillis

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.

February 12, 2015

*M. Phil. in Mathematical Finance,
University of Cape Town*



The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

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

Signed by candidate

signature removed

Zachry Pitsillis

February 12, 2015

Abstract

Duffee and Stanton (2012) demonstrated some pointed problems in estimating affine term structure models when the price of risk is dynamic, that is, risk factor dependent. The risk neutral parameters are estimated with precision, while the price of risk parameters are not. For the Gaussian models they investigated, these problems are replicated and are shown to stem from a lack of curvature in the log-likelihood function. This geometric issue for identifying the maximum of an essentially horizontal log-likelihood has statistical meaning. The Fisher information for the price of risk parameters is multiple orders of magnitude smaller than that of the risk neutral parameters. Prompted by the recent results of Christoffersen *et al.* (2014) a remedy to the lack of curvature is attempted. An unscented Kalman filter is used to estimate models where the observations are portfolios of FRAs, Swaps and Zero Coupon Bond Options. While the unscented Kalman filter performs admirably in identifying the unobserved risk factor processes, there is little improvement in the Fisher information.

Acknowledgements

I would like to express a deep thank you to my supervisors; doctor Peter Ouwehand and associate professor Thomas McWalter. Their guidance and insight has been invaluable. I am also deeply grateful for the warm and generous support from my parents. This dissertation could not have been completed without the love, friendship and direction which they provided.

Contents

1	Introduction	1
2	Affine Term Structure Models	3
2.1	Overview of the Affine Term Structure Literature	3
2.2	Mathematical Framework for Term Structure Modelling	6
2.3	Affine Term Structure Models	8
2.4	Empirical Limitations of CATSMs	10
2.5	Gaussian Models and a Dynamic Price of Risk	11
3	Estimating Dynamic Affine Term Structure Models	14
3.1	Kalman Filtering and Gaussian DATSMs	14
3.2	The Unscented Kalman Filter	16
3.3	Maximum Likelihood Parameter Estimates	21
4	Simulation Based Study of Parameter Estimates	23
4.1	Yields and the Kalman Filter	23
4.2	FRAs, Swaps and Bond Options and the Unscented Kalman Filter	29
5	Conclusion	33
	Appendix A	34
	Appendix B	35
	References	36

List of Figures

1	Log-likelihoods at various parameter values using the Kalman filter and zero coupon bond yields.	27
2	Simulated and filtered states for the two factor model, using yields and estimated parameters.	28
3	Simulated and filtered short rate using a portfolio of zero coupon bond options and the unscented Kalman filter.	31
4	Log-likelihoods at various parameter values using the unscented Kalman filter and zero coupon bond options.	32

List of Tables

1	One factor Gaussian model parameter estimation results.	25
2	Two factor Gaussian model parameter estimation results.	25
3	Fisher information for the one factor Gaussian model estimated via Kalman filter.	25
4	RMSE for the one and two factor models using true and estimated parameters.	26
5	The portfolio of zero coupon bond options used to investigate the performance of the unscented Kalman filter.	29
6	RMSE for the one factor models, using the true parameters with the unscented Kalman filter and portfolios of non-linear instruments . .	29
7	Fisher information for the one factor Gaussian model estimated via the Kalman filter and unscented Kalman filter.	30

1 Introduction

Interest rate yield curves are inherently forward looking; they describe the costs of transferring cash-flows between parties through time. The term structure of interest rates thus has critical implications for future economic states. Piazzesi (2009) highlights the term structure's importance for forecasting of interest rates, government debt policy, monetary policy and derivative and security pricing.

The modern economic treatment of term structures has centred on continuous time no arbitrage analysis of interest rates. Dai and Singleton (2003) view continuous time yield curve models in terms of three ingredients; the specification of risk neutral dynamics of risk factors, the dependence of interest rates on these risk factors and the associated risk factor premia. These three ingredients define a theoretical term structure model and are the key pieces to consider in connecting theoretical models to empirical observations.

Much of the econometric term structure modelling literature uses the affine class of models characterised by Duffie and Kan (1996). These models, termed completely affine, use a risk factor premium that is proportional to the volatility of the risk factors and fixed in sign (typically negative). Empirical observations have forced affine models of the term structure to incorporate a more dynamic specification of their risk factor premia. In order to explain the joint behaviour of excess bond returns and bond yield volatility, the price of risk must be dynamic (state dependent). Sufficient theoretical flexibility, however, comes with challenges in estimating dynamic model parameters. These challenges were addressed by Duffee and Stanton (2012) for the recently developed dynamic term structure models. They investigated the capabilities of various methodologies for estimating models through a simulation based study.

The first aim of this dissertation is to follow in the footsteps of Duffee and Stanton (2012) in assessing methodologies for estimating dynamic affine term structure models. In particular, the efficacy of the Kalman filter when used to implement maximum likelihood parameter estimation will be investigated. Duffee and Stanton (2012) found that identification of parameters which characterise the dynamic price of risk is particularly problematic for Kalman filtering based estimation. These failings are expounded and analysed through simulation based studies.

It is found that the Fisher information for the parameters associated with the price of risk is orders of magnitude less than that of the other parameters. This has a natural geometric interpretation; the log-likelihood is effectively flat. Consequently, there is little precision in maximising the log-likelihood (estimating the parameters).

The second aim of this dissertation is to investigate a potential innovation to the estimation of dynamic affine term structure models. It has recently been shown

by Christoffersen *et al.* (2014) that non-linear securities such as swaps and bond options can be used in an unscented Kalman Filter to effectively filter and forecast interest rates. Their investigation was conducted in a completely affine framework. They suggest that filters which incorporate enriched term structure information offer potential estimation improvements. This dissertation assesses these potential improvements for the estimation of dynamic affine term structure models. Despite the unscented Kalman filter providing quite remarkable accuracy for the identification of the state process, there is little improvement in the Fisher information.

In Chapter 2, the development of completely affine and then dynamic affine term structure models is reviewed. This is accompanied with an overview of the observations which force a dynamic price of risk and the subsequent estimation difficulties. Chapter 3 provides an exposition of the estimation procedures being investigated, while Chapter 4 contains the investigation methodology and results.

2 Affine Term Structure Models

2.1 Overview of the Affine Term Structure Literature

The empirical and theoretical investigation of interest rate term structures goes back to the work of Macaulay (1938), and there has since been a proliferation of literature on yield curves. Continuous time arbitrage free models of the yield curve were developed by Vasicek (1977) and Dothan (1978). Both authors relied on the specification of an underlying short rate diffusion. Combining the model of the instantaneous short rate with no arbitrage arguments, bond prices and hence a term structure can be derived. This approach was augmented with more unobserved state variables and generalised diffusion specifications, some notable examples of which are the work of Cox *et al.* (1985), Longstaff and Schwartz (1992) and Chen and Scott (1992). Many of the models were generalised and brought into a single consistent class of models by Duffie and Kan (1996). The class is referred to as affine term structure models, and has received much attention in the literature.

In the affine class of models bond yields are linear functions of the state variables. Duffee (2002) outlines that, depending on the form of the market price of interest rate risk, the linearity occurs in both the equivalent martingale measure and the physical measure. This relies on the market price of risk being proportional to the volatility of the risk factors, in which case the model is termed a completely affine term structure model (CATSM). This relationship is monotone; the market price of risk is only increasing or only decreasing with respect to the risk factor volatilities.

The strict proportional characterisation of the market price of risk is, however, counter-factual. The observed means and volatilities of excess returns on longer dated bonds, driven by the market price of risk, is incompatible with the behaviour of yields in a completely affine framework. In particular, Duffee (2002) identified that CATSMs cannot simultaneously match the observed distribution of yields and capture the predictability of excess bond returns (from the slope of the yield curve). Consequently, Duffee (2002) finds that the forecast accuracy of CATSMs is poor due to these empirical mismatchings. Similar empirical failings of CATSMs were identified by Dai and Singleton (2002) and Duarte (2004).

Dai and Singleton (2003) emphasises that despite the particular failings the affine term structure models have provided much insight into the expectations hypothesis and term structures as a whole. Thus, it was not appropriate to discard the affine models, but rather, their limitations prompted the development of a second generation of models.

The second generation of affine term structure models include the classes of essentially affine models introduced by Duffee (2002) and the semi-affine class intro-

duced by Duarte (2004). Here, they are jointly termed dynamic affine term structure models (DATSMs). These models allow for a more general specification of the price of risk while remaining as tractable as the original affine models. Duffee (2002) modifies the price of risk in his essentially affine model so that strict dependence on the volatility of the factors is relaxed. The price of risk is allowed to vary with the level of the risk factors and it is also possible for it to change sign. These extensions facilitate term structures to be modelled in a more empirically consistent manner. Duarte (2004) used the same specification as Duffee (2002) for the market price of interest rate risk with an additional term, in essence allowing for even more volatility independent variation in the price of risk.

The additional flexibility in the specification of the price of risk in these dynamic affine models comes at the cost of parameter estimation. Duffee and Stanton (2012) show that the empirical literature assessing the performance of the second generation affine models is sparse. In particular the finite sample behaviour of various estimation procedures requires attention. The asymptotic performance of maximum likelihood estimators, efficient method of moment estimators and Kalman Filter quasi-maximum likelihood is understood in general. However, the finite sample performance of these techniques when applied to second generation models requires further study.

Initial analysis of the use of Kalman filter in the context of first generation term structure models was undertaken by Duan and Simonato (1999) as well as by Lund (1995) and Lund (1997). The topic is still open with contemporary investigation of the use of the Kalman filter in completely affine term structure models undertaken by Christoffersen *et al.* (2014). Duffee and Stanton (2012) performed a Monte-Carlo study of the finite sample performance of various parameter estimation techniques for dynamic affine term structure models. Under various model specifications, with given parameters the state processes were simulated via a Milstein scheme and then different estimation techniques were applied. The input to the calibration techniques are bond yields, with or without added “observation” noise.

Duffee and Stanton (2012) find that for DATSMs some maximum likelihood parameters are estimated with little precision and are strongly biased. In addition they find that when the likelihood is intractable the simulation based efficient method of moments does not yield usable parameter estimates while the Kalman filter was found to give parameter estimates which had similar finite sample biases to maximum likelihood. Note that the Kalman filter can be used to estimate both maximum likelihood models and models where exact maximum likelihood is not possible. When maximum likelihood is infeasible and the Kalman filter is used, the estimates obtained are quasi-maximum likelihood estimates.

The high levels of parameter estimate bias and standard deviation are endemic to the parameters which characterise the dynamic price of risk. In all of the DATSMs which Duffee and Stanton (2012) investigated the bias and standard deviation for these parameters far exceeded the bias and standard deviation for the risk neutral parameters. Recall that the combination of risk neutral parameters with the price of risk parameters determines the dynamics of the risk factor processes under the real world probability measure.

The bias and standard error associated with the risk neutral drift and volatility terms of the risk factor processes is small because these parameters are accessible through the entire term structure of bond yields at each observation. That is, risk neutral parameters are identified through the use of panel data and longitudinal data. Because of the dynamic specification of the price of risk only longitudinal information supports the estimation of the price of risk parameters. Duffee and Stanton (2012) state that, in a completely affine model, the risk neutral drift parameters “pin down” the drift of the risk factors in the real world, in essence, directing the price of risk parameter estimates. In the DATSMs the real world drift is decoupled from the risk neutral drift and hence the cross sectional information in term structure at each observation does not provide information for the price of risk parameters, only the time series evolution of the yields does so.

The lack of cross sectional information about the price of risk parameters in DATSMs motivates the need to enrich the longitudinal information available. In this regard Christoffersen *et al.* (2014) showed that non-linear security prices¹ (e.g. swaps and caps) can be used to effectively filter for the risk factors in completely affine term structure models (with known parameters). They investigate three types of filters incorporating non-linear dependence of the observed security prices on the risk factors; the extended Kalman Filter, the unscented Kalman filter and the particle filter. Each of these filters receives extensive exposition in the book by Haykin, *et al.* (2001). Motivation to use non-linear instruments as observations is given by the results of Almeida *et al.* (2011), who find that using non-linear instruments greatly improves the forecasting of excess returns for swaps.

Christoffersen *et al.* (2014) find that the unscented Kalman filter provides accurate filtering of non-linear securities without the exorbitant computational burden of the particle filter. The extended Kalman filter was found to be inaccurate when the securities were highly non-linear functions of the state process. The accuracy which the unscented Kalman filter provides was found to facilitate effective parameter estimation in the completely affine context. However, this was established via asymptotic argument rather than an entire Monte Carlo experimentation.

¹Non-linear in the risk factors

The success of Almeida *et al.* (2011) in using instruments other than zero coupon bonds for the estimation of CATSMs naturally prompts the same inquiry into DATSMs. The further successes of Christoffersen *et al.* (2014) in the application of the unscented Kalman filter with non-linear instruments also encourages investigation. These results prompted the second step of this research to assess the viability of the unscented Kalman Filter for estimation of DATSMs. The driving question is whether the non-linearities in the risk factors for instruments such as FRAs, swaps and zero coupon bond options would resolve the lack of precision in the parameter estimates.

2.2 Mathematical Framework for Term Structure Modelling

This section is a brief exposition of term structure models based on the presentation given by Piazzesi (2009). There are a number of important assumptions which underlie the construction of a term structure model. In particular, the assumptions about risk preferences in the market as well as the absence of arbitrage and completeness of a market dictate the framework in which bond yields are modelled. The fundamental building blocks of a term structure model is the risk free zero coupon bond $P_t^{(\tau)}$ with tenor $\tau = T - t$ and the instantaneous risk free short rate r_t . The yield to maturity, or yield, of a bond is the log return associated with holding the bond to maturity denoted by

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

The short rate corresponds to the yield on an instantaneously maturing bond, that is $r_t = \lim_{\tau \downarrow 0} y_t^{(\tau)}$. Bond prices and the short rate can be related in any number of ways based on the assumptions made about the market in which they are determined.

A particularly general framework is that of martingale pricing theory. The details are suppressed here, suffice to say that if the market is arbitrage free and complete there exists a unique martingale probability measure \mathbb{Q} , equivalent to the real-world (data generating or physical) probability measure \mathbb{P} , under which bond prices are given by

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

where \mathcal{F}_t is the filtration generated by the relevant stochastic processes. In this context we have used a money market account earning the short rate as a numéraire which motivates calling \mathbb{Q} the risk neutral measure. The details surrounding equivalent martingale pricing theory can be found in Björk (2004).

It is worth emphasising at this point that the observed yield curve at each date t is determined by the equivalent martingale measure \mathbb{Q} . However, the evolution

through time for each tenor, τ , is determined by the real world measure \mathbb{P} . So the cross sectional term structure is \mathbb{Q} dependent and the longitudinal evolution is \mathbb{P} dependent.

The sources of random variation which drive the evolution of bond prices (and hence the term structure of interest rates) are modelled as n risk factors, $X_t \in \mathbb{R}^n$. These risk factors determine the short rate process $r_t = R(X_t)$ and are modelled as Itô processes with a stochastic differential equation

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

under \mathbb{Q} . Here \widetilde{W}_t is n -dimensional \mathbb{Q} Brownian motion, $\mu(t, X_t)$ is an $n \times 1$ vector and $\sigma(t, X_t)$ is an $n \times n$ matrix. We write the corresponding \mathbb{P} stochastic differential equation for X_t as

$$dX_t = \nu(t, X_t)dt + \sigma(t, X_t)dW_t$$

where W_t is n -dimensional \mathbb{P} Brownian motion, $\mu(t, X_t)$ is an $n \times 1$ vector. The technical restrictions on μ, ν and σ are suppressed for now.

The connection between \mathbb{P} and \mathbb{Q} is established in the specification of the one dimensional state price deflater process. Duffee (2002) gives that the state price deflater has the SDE

$$d\pi_t = -r_t\pi_t dt - \pi_t\Lambda(t, X_t)'dW_t$$

where $\Lambda(t, X_t)$ is an $n \times 1$ vector whose i th element represents the market price of risk associated with the Brownian motion $W_t^{(i)}$. That is, the measure \mathbb{Q} is the equivalent martingale measure given by Girsanov's theorem such that \widetilde{W}_t with SDE

$$d\widetilde{W}_t = dW_t + \Lambda(t, X_t)dt$$

is standard Brownian motion under \mathbb{Q} . Consequently, the \mathbb{P} dynamics of the risk factors can be written as

$$dX_t = [\mu(t, X_t) + \sigma(t, X_t)\Lambda(t, X_t)] dt + \sigma(t, X_t)dW_t$$

The advantage of continuous time arbitrage free asset pricing is, as Piazzesi (2009) points out, that Itô's lemma facilitates the identification of arbitrage free prices from the specification of the \mathbb{Q} dynamics of the risk factors. The assumption that the risk factors are Itô processes and hence Markovian implies we can rewrite the bond pricing equation as a direct function of the risk factors X_t , that is

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

Now if the function F is sufficiently smooth, Itô's lemma can be used to establish a stochastic differential equation for $P_t^{(\tau)}$. Combining the resultant stochastic differential equation with the known risk neutral drift of $P_t^{(\tau)}$ allows for a partial differential equation for the bond price to be established, as follows.

The bond price has a \mathbb{Q} SDE given by

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

From Itô's lemma we can identify

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

where $F_t(t, \tau, X_t) = \frac{\partial F(t, \tau, X_t)}{\partial t}$, $F_X(t, \tau, X_t)$ is an $1 \times n$ gradient vector and $F_{XX}(t, \tau, X_t)$ is an $n \times n$ Hessian matrix. In addition, we have that

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

Now, the risk neutral dynamics of $F(t, \tau, X_t)$ are known to follow

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

by the definition of the measure \mathbb{Q} . Thus the bond price must satisfy $r_t F(t, \tau, X_t) = \mu_F(t, \tau, X_t)$, which gives the bond price partial differential equation

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

2.3 Affine Term Structure Models

The affine class of models characterised by Duffie and Kan (1996) give a set of conditions on the risk factors, the short rate and the bond price function which are numerically tractable and have received much attention in the literature. The primary assumption of the affine class of models is that the bond price has an exponential-affine form

$$F(t, \tau, X_t) = \exp [A(\tau) - B(\tau)'X_t],$$

where A and B are at least twice differentiable functions of τ . Duffie and Kan (1996) show that, under quite general technical conditions, the assumption of exponentially affine bond prices implies that $\mu(t, X_t)$, $\sigma(t, X_t)\sigma(t, X_t)'$ and r_t are also affine functions of X_t .

Following in the notation of Dai and Singleton (2000) we can write the parametrisation for affine term structure models as follows. The short rate is given by $r_t = \delta_0 + \delta_X'X_t$ where δ_0 is a real scalar and δ_X is an $n \times 1$ real vector. The SDE of X_t is taken to be

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

where $\tilde{\kappa}$ and Σ are $n \times n$ matrices, $\tilde{\theta}$ is an $n \times 1$ vector and S_t is a diagonal matrix with elements

$$[S_t]_{ii} = \alpha_i + \beta'_i X_t.$$

Under this parametrisation the bond prices can be identified through the solution of the coupled Ricatti equations

$$\frac{dA(\tau)}{d\tau} = -\tilde{\theta}'\tilde{\kappa}'B(\tau) + \frac{1}{2} \sum_{i=1}^n [\Sigma' B(\tau)]_i^2 \alpha_i - \delta_0$$

and

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

Here the notation $[\Sigma' B(\tau)]_i$ indicates the i th element of the vector $\Sigma' B(\tau)$, and the boundary conditions are $B(0) = 0_{n \times 1}$ and $A(0) = 0$.

The completely affine term structure models are those in which the \mathbb{P} dynamics of the risk factors are also required to be affine. That is the drift and the square of the diffusion are required to be affine functions of the risk factors. This imposes the assumption that the market price of risk is

$$\Lambda_t = \sqrt{S_t} \lambda,$$

where λ is a $n \times 1$ real vector. The market price of risk relates only to the risk factors and does not depend on the maturity of any of the bonds in the market. This is an essential feature of all continuous time arbitrage free term structure models. In completely affine term structure models the market price of risk is strictly proportional to $\sqrt{S_t}$ which drives the volatility of the risk factors. Note also that the price of risk associated with each Brownian motion is either positive or negative.

With this specification of the price of risk the \mathbb{P} SDE of X_t is

$$\begin{aligned} dX_t &= \left[\tilde{\kappa}(\tilde{\theta} - X_t) + \Sigma S_t \lambda \right] dt + \Sigma \sqrt{S_t} dW_t \\ dX_t &= \kappa(\Theta - X_t) dt + \Sigma \sqrt{S_t} dW_t, \end{aligned}$$

where $\kappa = \tilde{\kappa} - \Sigma \Phi$ and $\Theta = \kappa^{-1}(\tilde{\kappa} \tilde{\theta} + \Sigma \psi)$, with the i th row of Φ given by $\lambda_i \beta'_i$ and ψ is an $n \times 1$ vector with elements $\lambda_i \alpha_i$. From now on a completely affine term structure model of this form will be referred to as a CATSM.

Naturally, there are a number of restrictions on the parameters which are required to ensure that the SDEs at hand have (strong) solutions. These restrictions, along with a classification scheme were developed by Dai and Singleton (2000). They call a model *admissible* if its parameters ensure that $[S(t)]_{ii}$ is strictly positive for all i .

The primary characteristic of a CATSM is the number of factors, n . Then for n -factor CATSMs Dai and Singleton (2000) show that there are $n + 1$ non-nested (distinct) subfamilies of admissible models each denoted by $\mathbb{A}_m(n)$ with $m = 0, 1, \dots, n$. The m here refers to the number of linearly independent combinations of risk factors which appear in the diffusion coefficient $\Sigma\sqrt{S_t}$. More specifically if we set $\mathcal{B} = [\beta_1, \dots, \beta_n]$ then $m = \text{rank}(\mathcal{B})$. Essentially, each sub-family is characterised by the total number of risk factors, n , and the number of these factors which affect the volatilities of the system, m .

Dai and Singleton (2000) define a *canonical representation of the subfamily* $\mathbb{A}_m(n)$ as the least restrictive n -factor CATSM parametrisation which is admissible and has $m = \text{rank}(\mathcal{B})$. There are technicalities about the uniqueness of the canonical representation given in the appendices of Dai and Singleton (2000). The subfamily $\mathbb{A}_m(n)$ is precisely defined as the set of all CATSMs which are nested special cases of the canonical representation for $\mathbb{A}_m(n)$. One salient point to note is that the process X_t is stationary if the eigenvalues of κ are strictly positive. In addition, Σ is restricted to be the identity matrix, essentially requiring each risk factor to represent a principle component of the term structure model. The specific parameter restrictions given by Dai and Singleton (2000) are listed in appendix A.

2.4 Empirical Limitations of CATSMs

There are a number of limitations associated with the completely affine interest rate models. Duffee (2002) reviews these limitations and summarises the mismatch between the empirical term structures and the models'. He highlights that affine models have poor forecast accuracy of expected excess returns on long dated bonds. This forecast inaccuracy stems from misspecification of the market price of interest rate risk. Problematically, the forecast errors in affine models are most pronounced when the yield curve is steeply upwards sloping. Non-zero correlations amongst the factors in combination with the price of risk specification implies unrealistic excess returns in long dated bonds. Furthermore, there is non-linear transformation in the term structure that the models fail to capture.

Duffee (2002) illustrates the impact of completely affine specification of the price of risk by considering the bond price SDE given by

$$\frac{dP_t^{(\tau)}}{P_t^{(\tau)}} = (r_t + e_t^{(\tau)})dt + v_t^{(\tau)}dW_t.$$

In a completely affine context the excess return and volatility for a bond are

$$e_t^{(\tau)} = -B(\tau)' \Sigma \sqrt{S_t} \Lambda_t = -B(\tau)' \Sigma S_t \lambda$$

and

$$v_t^{(\tau)} = -B(\tau)' \Sigma \sqrt{S_t}$$

respectively. Note the strict dependence on the risk factor volatility in both, and the consequent tight link between bond excess returns and bond return volatility.

Pronounced empirical mismatches relate to two of the stylised facts of term structures. Duarte (2004) summarises these stylised facts as i) “the excess return of treasury bonds has a high time variability” and ii) “the volatility of interest rates is time varying”. Affine term structure models in their initial form struggled to accommodate both these stylised facts due, in many cases, to the rigid specification of the market price of interest rate risk. In essence this is because there is a dynamic relationship between excess returns on treasury bonds and interest rate volatility. Dai and Singleton (2003) conclude that the empirical failings of affine term structure models are due to a “tension” within affine models between balancing the conditional expected returns and conditional volatilities with the fulcrum of the market price of interest rate risk. Specifically, the completely affine specification of the price of interest rate risk enforces a counter-factual proportional dependence of excess expected bond returns on risk factor volatilities.

2.5 Gaussian Models and a Dynamic Price of Risk

A second generation of affine term structure models was developed by Duarte (2004) and Duffee (2002), amongst others. Their innovation was to decouple the price of risk from the risk factor volatility. They both allow the price of risk to vary with the level of the risk factors. It was parameter estimates for the models developed by these authors which Duffee and Stanton (2012) investigated.

In order to keep the scope of this research manageable only the Gaussian models investigated by Duffee and Stanton (2012) are directly assessed here. The restriction of the investigation to only Gaussian models means that no approximations need to be made in both the simulation and estimation of the models. The Gaussian specification allows for affine yields and closed form bond option prices, in addition, the exact transition density for r_t can be identified. These standard results are presented in appendix B.

The focus on Gaussian models restricts the sources of error to the estimation procedures being investigated (as opposed to any approximations). In addition, it allows for computationally tractable extensions from the Kalman filter to the unscented filter. The existence of closed form pricing formulae for non-linear instruments means that the unscented Kalman filter can be implemented without the need to invoke finite difference methods.

In the one factor Gaussian model the risk neutral dynamics of the short rate are

$$dr_t = (\kappa\theta - \kappa r_t)dt + \sigma d\widetilde{W}_t.$$

The two factor Gaussian model is given in Duffee and Stanton (2012) as

$$\begin{aligned} r_t &= \delta_0 + X_{1t} + X_{2t} \\ dX_{it} &= -\kappa_i X_{it}dt + \sigma_i d\widetilde{W}_{it} \end{aligned}$$

for $i = 1, 2$.

There are many alternative specifications of the price of risk in an affine modelling framework. The price of risk specification for Gaussian models which Duffee and Stanton (2012) investigate is given by

$$\Lambda_t = \frac{\lambda_1 + \lambda_2 r_t}{\sigma}$$

in the one factor case and by

$$\Lambda_{it} = \frac{\lambda_{i1} + \lambda_{i2} X_{it}}{\sigma_i}$$

in the two factor case. These correspond to a special version of the price of risk supplied in Duffee (2002). Models with these specifications for the price of risk are termed here dynamic affine term structure models (DATSMs).

The key characteristic of these specifications is that the sign of the price of risk can change. This gives sufficient flexibility in the model to accommodate a dynamic relationship between volatility of bond returns (as well as risk factors) and the excess expected return. Duffee (2002) emphasises this flexibility. The interpretation is that DATSMs allow investors to seek and avoid different types of interest rate risk, this in turn facilitates (though does not guarantee) better empirical fit.

The added flexibility can be seen in the real world dynamics of the risk processes. In the one factor model the real world dynamics are

$$dr_t = ((\kappa\theta + \lambda_1) - (\kappa - \lambda_2)r_t)dt + \sigma dW_t.$$

and in the two factor model

$$dX_{it} = (\lambda_{i1} - (\kappa_i - \lambda_{i2})X_{it})dt + \sigma_i d\widetilde{W}_{it}.$$

The added flexibility in the price of risk decouples the risk neutral and real world drifts of the DATSM risk processes. This decoupling makes the identification of the price of risk parameters difficult, only longitudinal information can be used to identify the price of risk parameters. Cross-sectional information in the term structure only identifies the risk neutral parameters.

Through simulations Duffee and Stanton (2012) identify that the parameter estimates associated with the price of risk (λ_1 , λ_2 and the corresponding 2 factor parameters) suffer from pronounced biases. In addition the standard deviations associated with these parameters were much larger than those associated with the risk neutral parameters. The characteristics of the parameters estimates were shared across the different estimation methods which Duffee and Stanton (2012) used. Both direct maximum likelihood and Kalman filtering (quasi or exact maximum likelihood) produced biases and standard errors of the same magnitude. The efficient method of moments was found to produce much more severe biases and standard deviations.

3 Estimating Dynamic Affine Term Structure Models

3.1 Kalman Filtering and Gaussian DATSMs

The Kalman filter is a natural tool for the estimation of Gaussian affine term structure models, enabled by the linear relationship between yields and risk factors. This presentation follows closely that of Duffee and Stanton (2012) with a little discussion.

Consider the case of a known set of model parameters ρ . We observe a panel of d zero-coupon bond yields as linear functions of the m underlying risk factors with added observational noise

$$y_t = H_0(\rho) + H_1(\rho)x_t + \epsilon_t.$$

It is typical to assume that observation noise is a d dimensional serially independent Gaussian white noise with variance covariance matrix $R(\rho)$. The observations noise is typically taken to represent market micro-structure noise, liquidity effects and other sources of variation not captured by the model.

For a chosen set of yield tenors $\{\tau_1, \tau_2, \dots, \tau_d\}$ the observation matrices are given by

$$H_0(\rho) = \begin{bmatrix} \delta_0 - \sum_{i=1}^m \frac{A_i(\tau_1)}{\tau_1} \\ \vdots \\ \delta_0 - \sum_{i=1}^m \frac{A_i(\tau_d)}{\tau_d} \end{bmatrix}, \quad H_1(\rho) = \begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} & \dots & \frac{B_m(\tau_1)}{\tau_1} \\ \vdots & & \vdots \\ \frac{B_1(\tau_d)}{\tau_d} & \dots & \frac{B_m(\tau_d)}{\tau_d} \end{bmatrix}.$$

Using Gaussian risk factors means that the unobserved state process evolves over discrete intervals in a linear fashion. In vector form

$$x_{t+1} = F_0(\rho) + F_1(\rho)x_t + v_{t+1}$$

where v_{t+1} is the innovations or news process driving the real world evolution of the risk factors. The innovations have variance covariance matrix $Q(\rho)$. For the 2 factor case the vector F_0 and matrices F_1 and Q have values

$$\begin{aligned} F_0(\rho) &= \left[\frac{\kappa\theta_i + \lambda_{i1}}{\kappa_i - \lambda_{i2}} \left(1 - e^{-(\kappa_i - \lambda_{i2})\Delta t} \right) \right]_{i=1,2} \\ F_1(\rho) &= \text{diag} \left[1 - e^{-(\kappa_i - \lambda_{i2})\Delta t} \right]_{i=1,2} \\ Q(\rho) &= \text{diag} \left[\frac{\sigma_i^2}{2(\kappa_i - \lambda_{i2})} \left(1 - e^{-2(\kappa_i - \lambda_{i2})\Delta t} \right) \right]_{i=1,2}. \end{aligned}$$

In the one factor case, the above all become 1 dimensional and the i subscripts are dropped.

Note that H_0 and H_1 depend on the \mathbb{Q} parameters and the cross-sectional tenors, while the F_0 , F_1 and Q depend on the \mathbb{P} parameters and the observation time step Δt .

For known parameters Haykin (2001)² shows that the Kalman filter facilitates the minimum mean square error identification of the state variable x_t . The estimate of the state is denoted $x_{t|t}$ and its covariance matrix $P_{t|t}$. The Kalman filter requires values for the unconditional mean and covariance of the state to start the process. These values can be calculated in the Gaussian case as

$$x_{0|0} = E^{\mathbb{P}}[x_t] = \left[\frac{\kappa\theta_i + \lambda_{i1}}{\kappa_i - \lambda_{i2}} \right]_{i=1\dots m}$$

$$P_{0|0} = Var^{\mathbb{P}}[x_t] = diag \left[\frac{\sigma_i^2}{2(\kappa_i - \lambda_{i2})} \right]_{i=1\dots m}.$$

At each time step with given estimates the step ahead forecasts of the state $x_{t+1|t}$ and its covariance $P_{t+1|t}$ are calculated from the state evolution equation by

$$x_{t+1|t} = E^{\mathbb{P}}[x_{t+1}|x_t = x_{t|t}] = F_0(\rho) + F_1(\rho)x_{t|t}$$

and

$$P_{t+1|t} = Var^{\mathbb{P}}[x_{t+1|t}|x_t = x_{t|t}] = F_1(\rho)'P_{t|t}F_1(\rho) + Q(\rho).$$

This step ahead forecast facilitates a prediction of the observations

$$y_{t+1|t} = E^{\mathbb{P}}[y_{t+1}|x_t = x_{t|t}] = H_0(\rho) + H_1(\rho)x_{t+1|t}$$

and the observation covariance

$$V_{t+1|t} = Var^{\mathbb{P}}[y_{t+1}|x_t = x_{t|t}] = H_1(\rho)'P_{t+1|t}H_1(\rho) + R(\rho).$$

The prediction can be compared to the actual observation at $t + 1$, which can then be used to improve the original state and covariance. The prediction error is given by

$$e_{t+1} = y_{t+1} - y_{t+1|t},$$

which gives an updated estimator of the state and its covariance as

$$x_{t+1|t+1} = x_{t+1|t} + P_{t+1|t}H_1(\rho)V_{t+1|t}^{-1}e_{t+1}$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}H_1(\rho)V_{t+1|t}^{-1}H_1(\rho)'P_{t+1|t}.$$

The problem of finding the unknown true parameters ρ° requires an appropriate objective function. This is supplied by the log-likelihood of the observations, which when maximised jointly minimises the observation prediction error and supplies maximum likelihood parameter estimates.

²The book *Kalman Filters and Neural Networks* is edited by Haykin and chapter 1 on Kalman filtering is also written by Haykin.

With the assumption of Gaussian measurement errors, the conditional density of the observations $f(y_t|y_{t-1})$ is Gaussian with mean $y_{t|t-1}$ and covariance matrix $V_{t|t-1}$. The likelihood for a set of observations $y = \{y_1, \dots, y_T\}$ and for any parameter vector ρ corresponds to a product of conditional densities

$$\mathcal{L}(y_t; \rho) = \prod_{t=1}^T f(y_t|y_{t-1}; \rho).$$

The parameters for a Gaussian DATSM model can thus be estimated by identifying the parameters which yield the maximum likelihood. This is achieved by finding the parameters which maximise the log-likelihood

$$\ell(\rho; y) = -\frac{1}{2} \sum_{t=1}^T \left[d \log(2\pi) + \log |V_{t|t-1}| + e_t' V_{t|t-1}^{-1} e_t \right],$$

so that the parameter estimates are

$$\hat{\rho}(y) = \max_{\rho} \{\ell(\rho; y)\} = \max_{\rho} \{\mathcal{L}(\rho; y)\}.$$

The resulting optimisation problem of finding maximum likelihood parameter estimates can be implemented with standard computation software available in packages such as MATLAB. However, a number of challenges exist when doing so. Gupta and Mehra (1974) describe the difficulties associated with finding the true maximum likelihood. It is necessary to ensure that the starting point of the optimisation procedure are appropriate. In addition, it is necessary to conduct the optimisation multiple times using different starting points and algorithms in order to avoid accepting a local maximum. It is also important to enforce the parameter restrictions implicit in the model.

The computational burden associated with the optimisation can become onerous. Analytical expressions for Jacobian (score) and Hessian matrices are possible, but difficult to establish with the models at hand. Consequently, computationally intensive derivative free optimisation needs to be employed. Even with analytical expressions of the derivatives numerical issues persist, remedies have been attempted such as square root filters implemented by Bierman *et al.* (1990) and later by Kulikova (2009). These approaches may provide further numerical stability to the DATSM parameter estimation, but are not investigate here.

3.2 The Unscented Kalman Filter

The unscented Kalman filter was developed by Julier and Uhlmann (1997) for non-linear systems estimation and relies on the unscented transformation. This allows for

more accurate state estimation than if an extended Kalman filter³ were used. The key idea, as described by Wan and van der Merwe (2001), is to accurately capture the effect of non-linear transformations on means and variances with deterministic sampling points. This explanation follows that given by Wan and van der Merwe (2001) with simplifications to the application for the DATSMs at hand.

Firstly, it is necessary to understand the unscented transformation also developed by Julier and Uhlmann (1996). The unscented transformation aims to closely approximate the mean and variance of non-linear transformation, G , of a vector random variable z with mean μ_z and covariance Σ_z . If the dimension of z is m then $2m + 1$ vectors, called sigma vectors, are formed from points at scaled standard deviations away from the mean. The sigma vectors \mathcal{Z}_i are formed by

$$\begin{aligned}\mathcal{Z}_0 &= \mu_z \\ \mathcal{Z}_i &= \mu_z + (\sqrt{(m + \xi)\Sigma_z})_i \quad \text{for } i = 1 \dots m \\ \mathcal{Z}_i &= \mu_z - (\sqrt{(m + \xi)\Sigma_z})_i \quad \text{for } i = m + 1 \dots 2m.\end{aligned}$$

where $(\sqrt{(m + \xi)\Sigma_z})_i$ is the i th column of the Cholesky decomposition of $(m + \xi)\Sigma_z$ and ξ is a scaling parameter. In the Gaussian case the scaling parameter is determined by $\xi = 3\alpha^2 - m$, where α is set to a small positive number (typically less than 10^{-3}). Each sigma vector is passed through the non-linear transformation to yield points

$$\mathcal{G}_i = G(\mathcal{Z}_i), \quad i = 0 \dots 2m,$$

which are used in a weighted average to estimate the mean and covariance

$$\begin{aligned}E[G(z)] &\approx \sum_{i=0}^{2m} w_i^{(1)} \mathcal{G}_i = \hat{\mu}_g \\ \text{Var}[G(z)] &\approx \sum_{i=0}^{2m} w_i^{(2)} (\mathcal{G}_i - \hat{\mu}_g)(\mathcal{G}_i - \hat{\mu}_g)' = \hat{\Sigma}_g.\end{aligned}$$

The mean weights are given by

$$\begin{aligned}w_0^{(1)} &= \frac{\xi}{m + \xi} \\ w_i^{(1)} &= \frac{1}{2(m + \xi)} \quad i = 0 \dots 2m\end{aligned}$$

³The extended Kalman filter is much the same as the Kalman filter discussed in section 3.1, with the addition of first order Taylor series terms to the transition and observation equations to account for non-linearities. Most applications in which the state is non-Gaussian, i.e. a CIR process, use the extended Kalman Filter

and covariance weights

$$w_0^{(2)} = \frac{\xi}{m + \xi} + 3 - \alpha^2$$

$$w_i^{(2)} = \frac{1}{2(m + \xi)} \quad i = 0 \dots 2m.$$

The unscented Kalman filter follows the same steps as the Kalman filter with the use of the unscented transformation. Because of the use of Gaussian processes in the DATSM being investigated here a slightly simplified form of the unscented Kalman filter can be implemented. This corresponds to the unscented Kalman filter with additive noise as presented by Wan and van der Merwe (2001). We have that the process evolves linearly as before

$$x_{t+1} = F_0(\rho) + F_1(\rho)x_t + v_{t+1}$$

where v_{t+1} is the innovations with variance covariance matrix $Q(\rho)$. However, the observation are a non-linear function of the states

$$y_t = H(x_t; \rho)$$

that consists of a panel of security prices or rates. The panel of prices can include any number of securities from zero coupon bond yields to swap rates to bond option rates⁴

Again, consider the case of a known set of parameters ρ , we can initialise the state process as before at the unconditional mean and variance of the process

$$x_{0|0} = E^{\mathbb{P}}[x_t]$$

$$P_{0|0} = Var^{\mathbb{P}}[x_t]$$

Then for each time step of the observation $t \in \{1 \dots T\}$ evolve the state and its covariance forward using the unscented transformation. This involves creating a set of sigma points⁵

$$\mathcal{X}_{t-1} = \left[x_{t-1|t-1}, \quad x_{t-1|t-1} + \gamma\sqrt{P_{t-1|t-1}}, \quad x_{t-1|t-1} - \gamma\sqrt{P_{t-1|t-1}} \right]$$

which are evolved through the state transition

$$\mathcal{X}_t^* = F(\mathcal{X}_{t-1}; \rho)$$

⁴Note one major advantage of this approach is that direct market observations can be used. For instance in the South African context coupon bearing bond yields (via the BESA formula) or FRA and swap rates including day count conventions

⁵Here $\gamma = \sqrt{m + \xi}$ and $x_{t-1|t-1} + \gamma\sqrt{P_{t-1|t-1}}$ means the matrix formed by the addition of each column of $\gamma\sqrt{P_{t-1|t-1}}$ to $x_{t-1|t-1}$.

where $F(x) = F_0(\rho) + F_1(\rho)x$. These evolved sigma points facilitate estimation of the state mean and covariance

$$x_{t|t-1} = \sum_{i=0}^{(2m)} w_i^{(1)} \mathcal{X}_{i,t}^*$$

$$P_{t|t-1} = \sum_{i=0}^{(2m)} w_i^{(2)} (\mathcal{X}_{i,t}^* - x_{t|t-1})(\mathcal{X}_{i,t}^* - x_{t|t-1})' + Q(\rho).$$

To forecast the observation it is first necessary to resample from the state process incorporating the forecast state covariance

$$\mathcal{X}_{t|t-1} = \left[\mathcal{X}_{0,t}^*, \quad \mathcal{X}_{0,t}^* + \gamma\sqrt{P_{t|t-1}}, \quad \mathcal{X}_{0,t}^* - \gamma\sqrt{P_{t|t-1}} \right].$$

These form estimates of the conditional observation mean and covariance

$$\mathcal{Y}_{t|t-1} = H(\mathcal{X}_{t|t-1})$$

$$y_{t|t-1} = \sum_{i=0}^{(2m)} w_i^{(1)} \mathcal{Y}_{i,t}$$

$$V_{t|t-1} = \sum_{i=0}^{(2m)} w_i^{(2)} (\mathcal{Y}_{i,t} - y_{t|t-1})(\mathcal{Y}_{i,t} - y_{t|t-1})' + R(\rho)$$

as well as the covariance between the observations and the state process

$$J_{t|t-1} = \sum_{i=0}^{(2m)} w_i^{(2)} (\mathcal{X}_{i,t} - x_{t|t-1})(\mathcal{Y}_{i,t} - y_{t|t-1})'.$$

In the same manner as the Kalman filter the observation forecast error ($y_t - y_{t|t-1}$) is used to update the estimations of the state and its covariance

$$K_t = J_{t|t-1} V_{t|t-1}^{-1}$$

$$x_{t|t} = x_{t|t-1} + K_t (y_t - y_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - K_t V_{t|t-1} K_t'.$$

In the case of unknown parameters it is, in general, not possible to identify the exact likelihood function. Because of the non-linear transformation the (in this application) Gaussian state process does not give a Gaussian observations. However, the unscented filter can be used to implement quasi-maximum likelihood, as the conditional mean and variance of the observations are accurately captured. So as with the Kalman filter an objective function can be formed from the (quasi) Gaussian log-likelihood. This assumption is typical in the literature, both Duffee and Stanton

(2012) and Christoffersen *et al.* (2014) use a Gaussian log-likelihood as an objective function.

The use of a quasi log-likelihood has been well established in the econometric literature. A deep presentation of this is give by White (1996). The core result is that in many cases the use of an appropriate quasi likelihood maintains consistency and lack of bias in the parameter estimates. With careful design the efficiency of the parameter estimates is only marginally comprised. Of course none of these favourable characteristic are always guaranteed in the limit, as in the case of the maximum likelihood parameters. It depends critically on the choice of the quasi likelihoods in relation to the true likelihood. And of course, the use of finite samples further obscures the behaviour of quasi maximum likelihood parameter estimates.

Christoffersen *et al.* (2014) use an approximate asymptotic argument to suggest the efficacy of the unscented Kalman filter with Gaussian quasi likelihood for parameter estimation. They construct an approximation to the parameter estimation error by first estimating the score statistics at the true parameters

$$\hat{U}(\rho_i^\circ) = \frac{\ell(y; \rho_i^\circ + \varepsilon) - \ell(y; \rho_i^\circ - \varepsilon)}{2\varepsilon} \quad i = 1 \dots L$$

where ε is set to 10^{-6} , L is the number of parameters and ρ° is the vector of parameters. Then, using the likelihood Hessian matrix $H(\rho^\circ)$ found using perfect filtering and second differencing, and they form an approximation to the parameter estimate error

$$\hat{\rho}_i - \rho_i^\circ \approx \hat{U}(\rho_i^\circ) H^{-1}(\rho^\circ)_{ii}.$$

This is not strictly a measure of parameter error as Christoffersen *et al.* (2014) describe it. Most notably the scale of the values they report (in their Table 8) is vast compared to the actual parameters. Rather the quantity above, is a Hessian scaled score, and indicates the extent to which the log-likelihood has a maximum at the true parameter value, and thus the strength of the possibility that a maximum likelihood parameter estimate will be close to the true value. Their interpretation of the distribution of this error estimate is, however, appropriate. That the median Hessian scaled score is zero and that the unscented Kalman filter has a much smaller dispersion are good (though not conclusive) indicators of the efficacy of parameter estimation via the unscented Kalman filter. Zero valued scores are necessary for the parameter estimates to be close to the true values but not sufficient. The same complications discussed with maximising the log-likelihood supplied by the Kalman filter also apply to unscented Kalman filters.

3.3 Maximum Likelihood Parameter Estimates

A brief review of the theory of maximum likelihood parameter estimation is useful to illuminate some of the issues driving this investigation. Maximum likelihood estimators (MLEs) are central to the theory and practice of statistics. Their use is in many cases motivated by their limiting properties. An exposition of the theory of maximum likelihood and statistical inference are available in the book by Millar (2011).

The log-likelihood function is maximised by setting the scores

$$U(\rho) = \frac{\partial}{\partial \rho} \ell(Y; \rho)$$

to zero. At the maximum likelihood parameters the scores are random variables with

$$E^{\mathbb{P}}[U(\rho^{\circ})] = 0$$

The maximum likelihood parameters $\hat{\rho}(y)$ given in the case of the Kalman filter are random variables, dependent on the observations y . Provided the distribution of the sample satisfies some regularity conditions, MLEs have very favourable characteristics as the sample size tends to infinity. Firstly they are consistent, which is to say $\hat{\rho}(y)$ converges in probability to the true parameters ρ_0 . In addition, the limiting distribution is Gaussian,

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

where $\mathcal{I}(\rho_0)$ is the Fisher information matrix associated with the parameters. This introduces the efficiency of MLEs, in the limit they achieve the Cramer-Rao lower bound on their covariance $\mathcal{I}(\rho_0)^{-1}$.

The Fisher information matrix is defined as

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

which, under certain regularity conditions, becomes

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

The interpretation of the above quantity as information is seen through the geometric notion of curvature. Barring some pathologies, the greater the (negative) curvature of the log density function around the true parameter the more information available for the parameter estimate. So when there is greater curvature in

the log-likelihood the the maximum of the likelihood can be identified with more precision. This feeds naturally into the statistical notion of variance, the greater the information the less variance there is, in the limit, for the MLE parameter estimate.

The favourable performance of a maximum likelihood estimator for a finite sample is not guaranteed from these characteristics. In fact it is the problems of finite sample MLEs that Duffee and Stanton (2012) investigate. Evaluation and analysis of the performance of MLEs is possible through comparison of the sample mean and variability of the estimators with their limiting distributional characteristics. Naturally, an assessment of the scores and Fisher information associated with each parameter then becomes useful.

4 Simulation Based Study of Parameter Estimates

The first step of this investigation is to replicate the results of Duffee and Stanton (2012) for one and two factor Gaussian models. The replication is largely successful, with very similar results being obtained. Then the significant parameter biases which Duffee and Stanton (2012) identified are explained through an analysis of the Fisher information for the parameters. It is found that the log-likelihood is essentially flat for the price of risk parameters. Consequently, the Fisher information for the price of risk parameters is comparatively tiny and hence they are estimated with little precision.

In the second step an unscented Kalman filter is used to assess the viability of non-linear instruments for parameter estimation. Analysis of the quasi-log likelihood function is used to show that with moderately non-linear instruments (FRAs and Swaps) there is no improvement in the Fisher information. The use of zero coupon bond options provides a marginal improvement to the Fisher information. However, given the scale of the deficit of Fisher information, this improvement is not significant.

4.1 Yields and the Kalman Filter

The one and two factor Gaussian models described in section 2.5 were simulated using time steps of one week ($\Delta t = \frac{1}{52}$) and $T = 1000$ observations. There were $n = 500$ simulations conducted. The parameters used were the same as Duffee and Stanton (2012) (their Table 4) and are given here in Tables 1 and 2. From each simulation of the risk factors a set of yields were then calculated on zero coupon bonds with tenors of 3 months, 6 months, 1 years, 5 years and 10 years. Then Gaussian noise, with standard deviation \sqrt{V} was added to the bond yields.

Each set of bond yields with noise y_k , $k \in \{1, \dots, n\}$, was passed into an optimisation routine which searches the parameter space for optimal parameters. At each parameter iteration the Kalman filter is called and the log-likelihood is identified, the optimisation iterates until a maximum for the log-likelihood is found. The optimisation was conducted at least twice for each simulation, the optimisations are started with random parameters (in the space of possible parameters). In the event that the optimisation failed⁶ another set of starting parameters are generated. The constrained optimisation routines used were the MATLAB routines `active-set` and `sqp`⁷.

⁶Optimisation could fail for reasons including too many iterations, parameters on their boundary, failure of the Kalman filter for the given parameters and others

⁷For details on these routines see <http://www.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html>

To assess the performance of the parameter estimates Duffee and Stanton (2012) estimate the mean of the parameter estimates and their standard deviation. These quantities are compared to the true parameter and the estimated asymptotic standard error of the parameter estimate. This asymptotic standard error is calculated here using the inverse of the estimated Fisher information matrix. This is calculated by approximating the scores for each parameter and sample as a finite difference

$$\hat{U}(\rho_i; y_k) = \frac{\ell(y_k; \rho_i + \varepsilon_i) - \ell(y_k; \rho_i - \varepsilon_i)}{2\varepsilon_i} \quad i = 1 \dots L, \quad k = 1, \dots, n$$

where $\varepsilon_i = \frac{\rho_i}{20000}$, the Fisher information is estimated by their covariance

$$\hat{\mathcal{I}}(\rho)_{ij} = \text{cov}(\hat{U}(\rho_i; y), \hat{U}(\rho_j; y)).$$

The standard error is then

$$s.e.(\rho_i) = \sqrt{[\hat{\mathcal{I}}(\rho)^{-1}]_{ii}}.$$

These estimates are shown in Tables 1 and 2. The estimates for the mean and standard deviation found in this study are close to those found by Duffee and Stanton (2012) (their Table 4). The standard error estimates found in this study are usually a bit lower than that of Duffee and Stanton (2012). The risk neutral parameters are estimated accurately, well within a standard deviation of their true values. The price of risk parameters, conversely, are estimated with little precision. There is substantial bias in the estimates and the standard deviation is an order of magnitude or two larger than that of the risk neutral parameters. The standard deviations of the risk neutral parameters are typically close to their lower bound⁸. For the price of risk parameters, the standard deviations were much larger than their asymptotic lower bounds. The impact of the bias in the price of risk parameters is econometrically significant. It leads to a misidentification of the expected excess bond return as well as an over estimation of the half life of economic shocks. These issues receive a full presentation in Duffee and Stanton (2012). The contribution of this research is rather to identify the cause of the parameter estimate inexactitude.

The issue can be clearly seen in Figure 1 which shows the log likelihood function at various parameter values for 20 samples of the one factor model. In each subplot the parameter in the x -axis is varied while the other parameters are kept at their true values. For the two risk neutral parameters plotted (κ and σ) the log likelihoods are concave and peaked close to the true parameter value. In such an environment the maximum is easy to identify (at least in one dimension). For

⁸The standard deviations for σ and δ_0 are actually below their lower bound estimate, this is probably due to the use of a finite difference calculation for the scores.

Table 1: One factor Gaussian model parameter estimation results.

	True Value	Mean Estimate	Std. Dev.	Std. Err.
$\kappa \times 10^3$	65	65.26	3.39	2.34
$\kappa\theta \times 10^3$	8.4	8.43	0.28	0.20
$\sigma \times 10^3$	17.5	17.54	0.74	0.92
$\lambda_1 \times 10^3$	5	23.2	19.34	10.87
$\lambda_2 \times 10^3$	-140	-398.52	254.98	146.96
$\sqrt{V} \times 10^3$	6	6.00	0.06	0.06

Table 2: Two factor Gaussian model parameter estimation results.

	True Value	Mean Estimate	Std. Dev.	Std. Err.
$\kappa_1 \times 10^3$	700	700.44	21.70	17.55
$\kappa_2 \times 10^3$	20	20.36	3.94	2.40
$\sigma_1 \times 10^3$	20	20.02	0.79	0.79
$\sigma_2 \times 10^3$	13	13.04	0.59	0.55
$\lambda_{11} \times 10^3$	-10	-14.82	8.95	7.47
$\lambda_{12} \times 10^3$	-4	-12.02	9.38	4.79
$\lambda_{21} \times 10^3$	100	-133.41	292.61	122.11
$\lambda_{22} \times 10^3$	-120	-412.42	222.84	183.34
$\delta_0 \times 10^3$	100	100.20	5.39	5.61
$\sqrt{V} \times 10^3$	2	2.00	0.02	0.02

Table 3: Fisher information for the one factor Gaussian model estimated via Kalman filter.

	Fisher Information Estimate $\hat{\mathcal{I}}(\rho^\circ)_{ii}$
κ	2385286
$\kappa\theta$	400917347
σ	4670803
λ_1	64627
λ_2	355
\sqrt{V}	250647858

the price of risk parameters, however, the likelihood is essentially horizontal. It is not surprising therefore that the parameters are not identified with a great deal of precision. The initial visual evidence in Figure 1 is corroborated by the values in Table 3. The average curvature (Fisher information) of the likelihood in the price of risk parameters is substantially lower than that of the other parameters. Similar results were found in the two factor model.

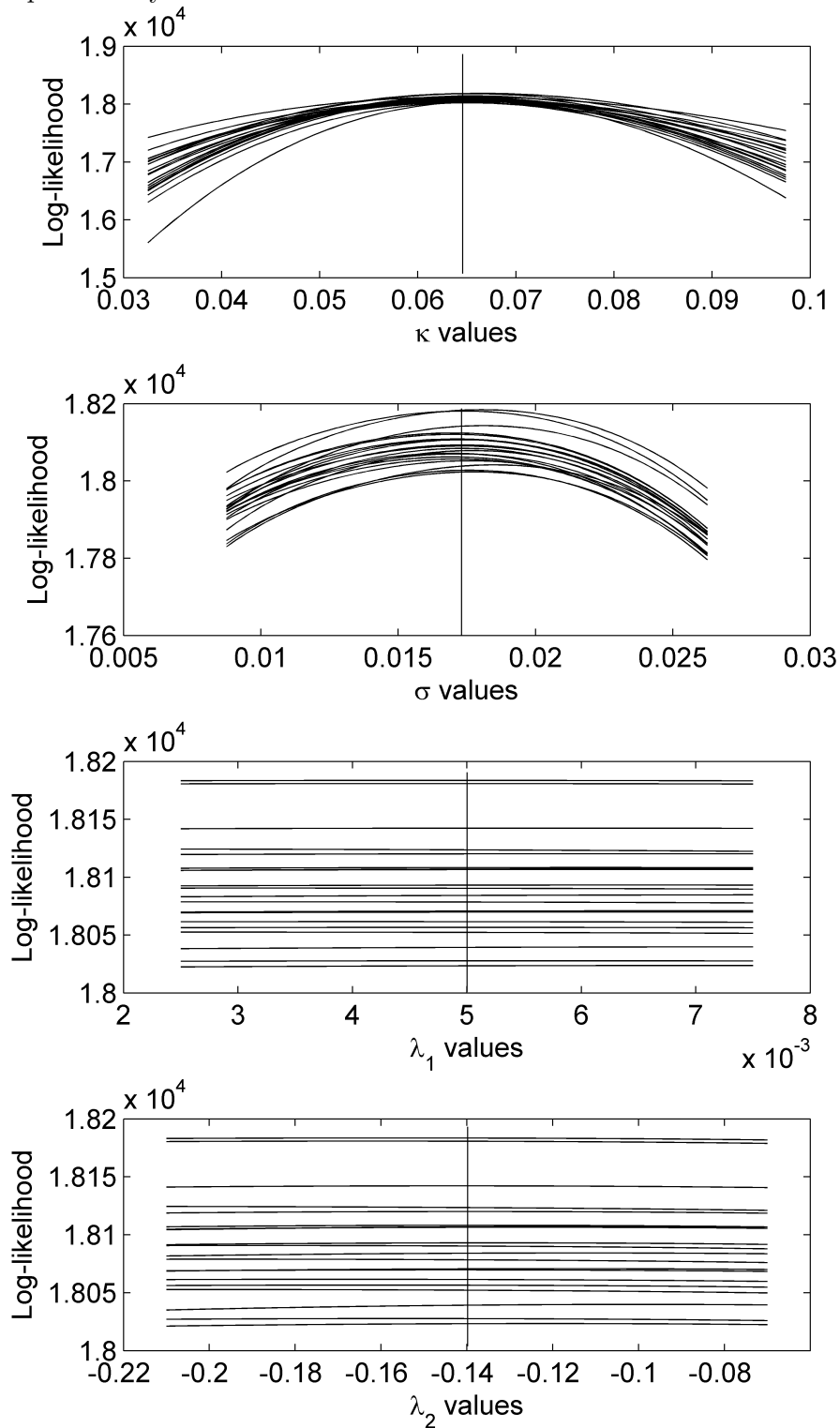
One final point to note about the Kalman filter implementation is its efficacy in identifying the states. Table 6 shows the Root Mean Square Error (RMSE) of the filtered state space when both the true parameters and estimated parameters are used. The scale of the risk factors in the simulation is in the order of 10^{-2} , while the error is of the order 10^{-3} . So the Kalman filter is accurately identifying the states when the true parameters are used. When the parameters are estimated the results are also good, with the two factor model providing slightly less accuracy. Figure 2 demonstrates these results for the simulated and filtered states for the two factor model, using estimated parameters.

Table 4: RMSE for the one and two factor models using true and estimated parameters.

	True Parameters	Estimated Parameters
One factor model	2.183×10^{-3}	2.184×10^{-3}
Two factor model	1.833×10^{-3}	4.402×10^{-3}

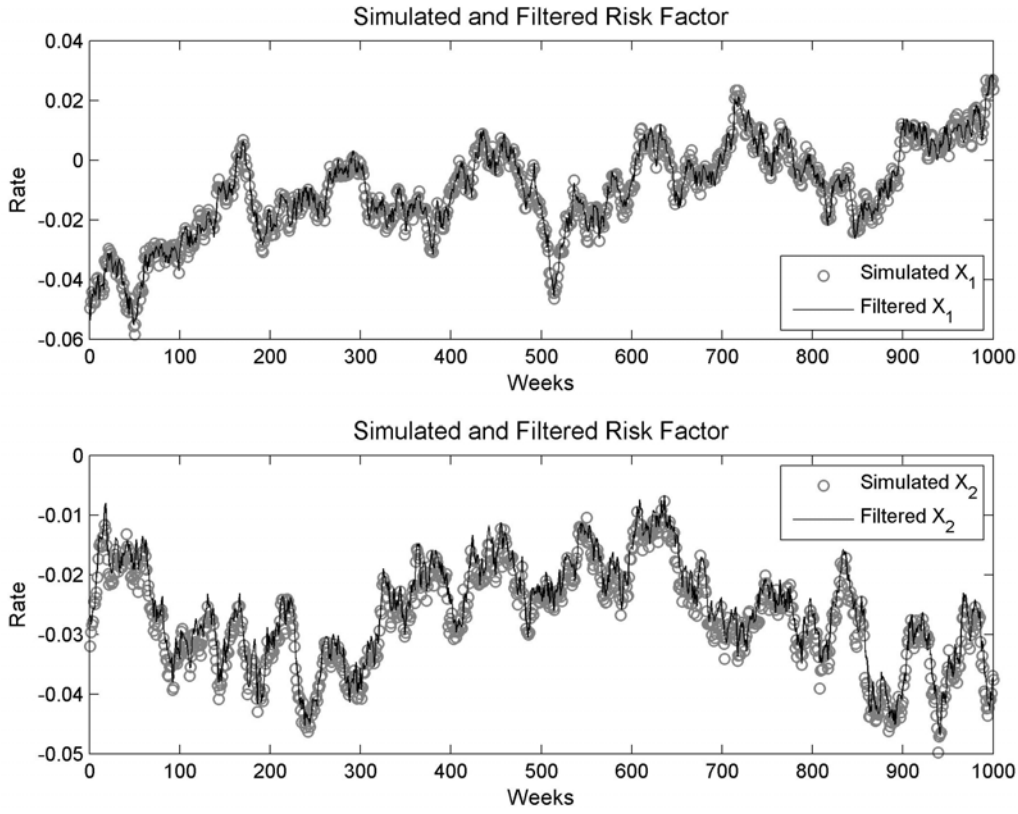
Note, for each path the parameter estimates found for that path are used.

Figure 1: Log-likelihoods at various parameter values using the Kalman filter and zero coupon bond yields.



Each x axis covers 50% to 150% of the true parameter value, the true parameter is indicated with a vertical line.

Figure 2: Simulated and filtered states for the two factor model, using yields and estimated parameters.



4.2 FRAs, Swaps and Bond Options and the Unscented Kalman Filter

A one factor model is investigated using simulation and asymptotic arguments. The unscented filter is found to accurately filter for the state space process. However, there is only marginal improvement in the Fisher information for the price of risk parameters. It is found, however, that the use of options as observations improves the Fisher information for the volatility parameter.

Two investigations were undertaken. Both simulated the short rate from the one factor model, using the same true parameters as in table 1. The first investigation then calculated the rates for a 3m×6m FRA, 6m×1yr FRA, 2 year, 5 year and 10 year swap. The payment frequencies for the swaps are quarterly for the 2 year swap and half yearly for the 5 and 10 year swaps. The second investigation used a portfolio of zero coupon bond options, whose characteristics are given in Table 5. For the set of rates in the first investigation an observation error of $\sqrt{V} = 0.006$ was maintained, while for the options the observation error was set to $\sqrt{V} = 0.6$

Table 5: The portfolio of zero coupon bond options used to investigate the performance of the unscented Kalman filter.

Option	Type	Underlying Bond Maturity	Option Maturity	Strike (% of Underlying)
1	Call	2 Year	1 Year	98%
2	Put	2 Year	$\frac{1}{2}$ Year	110%
3	Call	5 Year	2 Year	110%
4	Put	5 Year	2 Year	130%
5	Put	10 Year	3 Year	130%

All options had a nominal of 100.

Table 6: RMSE for the one factor models, using the true parameters with the unscented Kalman filter and portfolios of non-linear instruments

	RMSE with true parameters
FRAs and Swaps	2.751×10^{-3}
Bond Options	1.217×10^{-3}

The unscented Kalman filter was found to be remarkably accurate at identifying the state process. The RMSE for both portfolios shown in Table 6. When zero coupon bond options are used the RMSE was just above 10 basis points, which is lower than when yields are used in the Kalman filter. This filtering efficacy is demonstrated in Figure 3. The RMSE for portfolio of FRAs and swaps was not as good, but still performed well at just under 30 basis points.

In order to assess the unscented Kalman filter’s capabilities for estimation, the

(quasi) Fisher information is calculated. A necessary, though not sufficient, condition for the unscented Kalman filter to provide a solution to the parameter estimate problem is for there to be a substantial improvement in the curvature of the quasi likelihood. Unfortunately, this is not provided. The Fisher information estimate for the original Kalman filter is given in Table 7 alongside the Fisher information obtained using the unscented Kalman filter for the two portfolios. There is no substantial improvement in the Fisher information when a portfolio of non-linear instruments is used. It is encouraging, however, to see the substantial improvement in the Fisher information for the volatility parameter when zero coupon bond options are used. This is to be expected given the sensitivity of options to volatility parameters. Figure 4 shows the quasi log-likelihood for 20 samples using zero coupon bond options, demonstrating the improved curvature in σ and lack of curvature in λ_1 and λ_2 .

Table 7: Fisher information for the one factor Gaussian model estimated via the Kalman filter and unscented Kalman filter.

Parameter	Fisher information for various filtering techniques $\hat{\mathcal{I}}(\rho^\circ)_{ii}$		
	KF with yields	UKF with FRAs and Swaps	UKF with ZCB options
κ	2385286	153325	308923
$\kappa\theta$	400917347	24266013	406251179
σ	4670803	1125481	275263064
λ_1	64627	66055	71486
λ_2	355	344	363
\sqrt{V}	250647858	239824899	90648

The one factor Gaussian DATSM is the simplest possible model to assess the efficacy of parameter estimates. In the case of using yields as observations the Kalman filter provides exact maximum likelihood. The lack of curvature in the (quasi) Gaussian log-likelihood when various instruments are used is deeply problematic for the estimation of DATSMs. Any additional complexity such as the addition of factors or the use of square root processes requires further analytic and computational approximation. While these more complex models could be investigated, any meaningful improvement on the failures of the most basic case are extremely unlikely.

Figure 3: Simulated and filtered short rate using a portfolio of zero coupon bond options and the unscented Kalman filter.

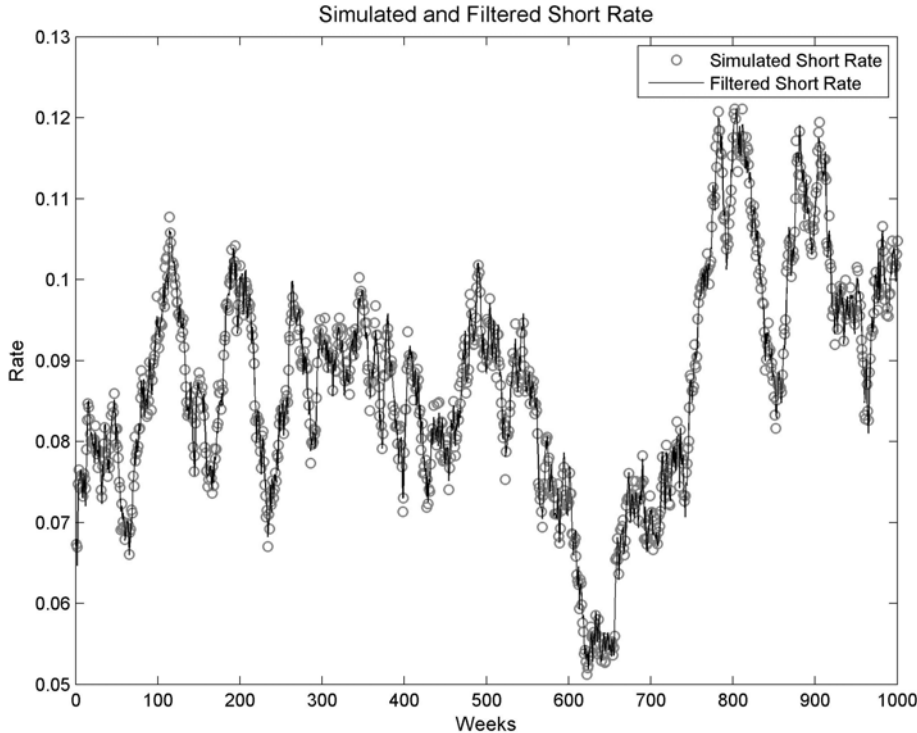
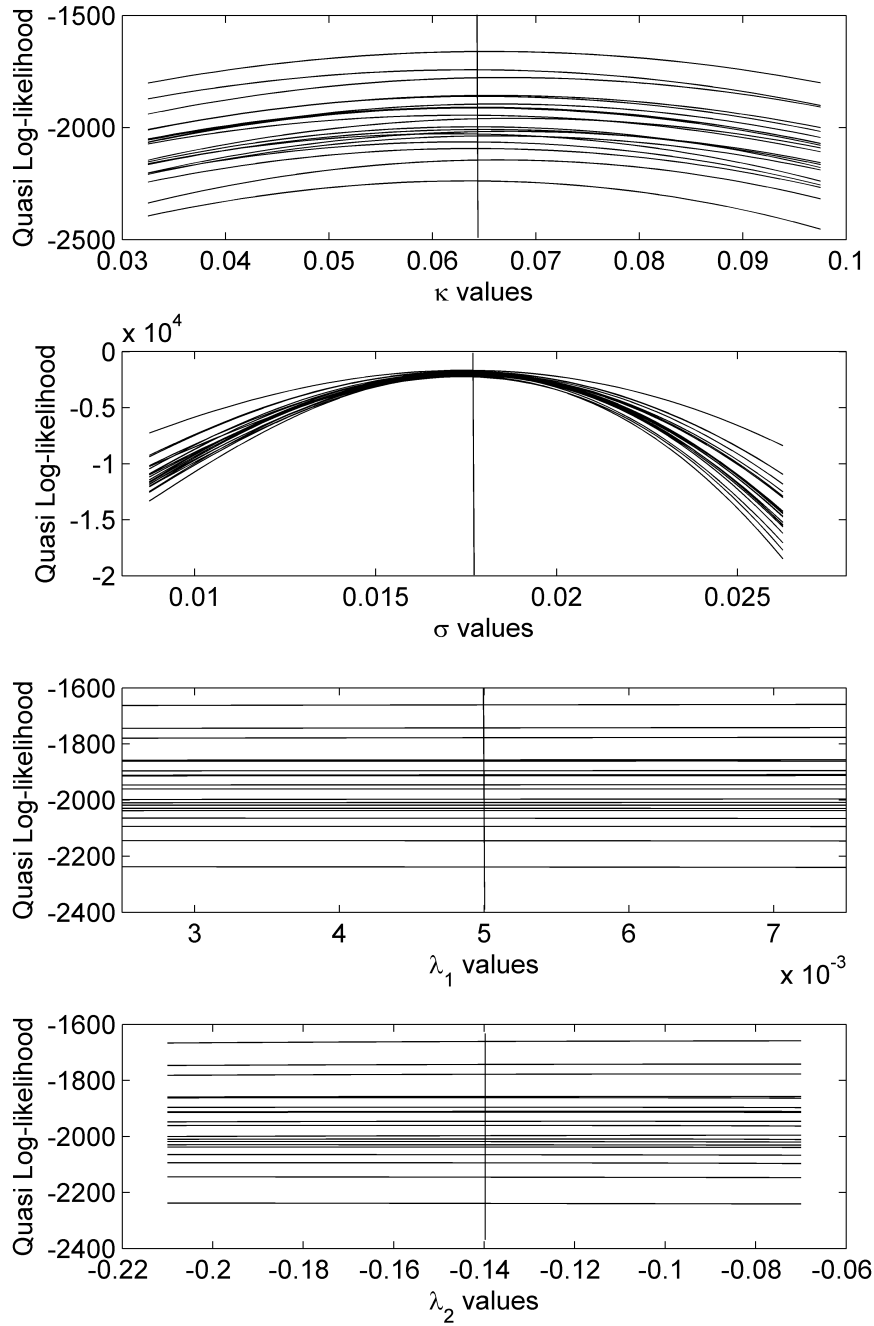


Figure 4: Log-likelihoods at various parameter values using the unscented Kalman filter and zero coupon bond options.



Each x axis covers 50% to 150% of the true parameter value, the true parameter is indicated with a vertical line.

5 Conclusion

Term structure modellers must sympathise with Yossarian⁹; observations have necessitated the use of dynamic affine models, however observations are of little use for estimating the dynamic price of risk parameters. In the most basic Gaussian one factor model the price of risk parameters cannot be estimated with great precision with a reasonably sized sample. This effect is even more pronounced for the parameter which is introduced to make the price of risk state dependent. The lack of Fisher information for the price of risk parameters makes the log-likelihood effectively flat in these parameters. Consequently, any attempt to find a maximum is fated to be imprecise.

The root cause of the imprecision is identified by Duffee and Stanton (2012). When the price of risk is allowed vary, the drift of the real world risk factors is “decoupled” from the risk neutral parameters. More generally, a state dependent Radon-Nikodym derivative estranges the equivalent real world and risk neutral measures. This constrains the statistical power and precision of parameter estimates. The rich information in the cross sectional observations, defined by the risk neutral measure, provides little guidance for the longitudinal processes, defined by the real world measure.

The difficulties induced by DATSMs are not, however, without value. They illustrate a key point for financial modelling. While many model classes may be able to recover stylised facts, the ability of the model to provide parameter estimates which are precise is, in many ways, more important. A sufficiently flexible model is of little use if the parameter estimates are too loose. In this regard, a simulation based analysis of the Fisher information for the model parameters before the use of data is vitally important.

⁹The lead character in Heller’s *Catch-22*

Appendix A

The following definition of a *canonical representation of the subfamily* $\mathbb{A}_m(n)$ is taken directly from Dai and Singleton (2000).

For each m , we partition X_t as $X = (X^B, Y^D)'$, where X^B is $m \times 1$ and X^D is $(n - m) \times 1$, and define the canonical representation of $\mathbb{A}_m(n)$ as the special case of

$$dX_t = \kappa(\Theta - X_t)dt + \Sigma\sqrt{S_t}dW_t,$$

with the following restrictions. For $m > 0$ we have

$$\kappa = \begin{bmatrix} \kappa_{m \times m}^{BB} & 0_{m \times (N-m)} \\ \kappa_{(n-m) \times m}^{DB} & \kappa_{(n-m) \times (n-m)}^{DD} \end{bmatrix}$$

and if $m = 0$, κ must be upper or lower triangular. The other restrictions are

$$\Theta = \begin{bmatrix} \Theta_{m \times 1}^B \\ \Theta_{(n-m) \times 1}^D \end{bmatrix},$$

$$\Sigma = I_{n \times n},$$

$$\alpha = \begin{bmatrix} 0_{m \times 1} \\ 1_{(n-m) \times 1} \end{bmatrix},$$

$$\mathcal{B} = \begin{bmatrix} I_{m \times m} & \mathcal{B}_{m \times (N-m)} \\ 0_{(n-m) \times m}^{DB} & 0_{(n-m) \times (n-m)}^{DD} \end{bmatrix}.$$

The corresponding restrictions on the parameters of the CATSM are

$$\begin{aligned} \delta_{X_i} &\geq 0, \quad m + 1 \leq i \leq n, \\ \kappa_i \Theta &\equiv \sum_{j=1}^m \kappa_{ij} \Theta_j \geq 0, \quad 1 \leq i \leq m, \\ \kappa_{ij} &\leq 0 \quad 1 \leq j \leq m, \quad j \neq i \\ \Theta_i &\geq 0, \quad 1 \leq i \leq m, \\ \mathcal{B}_{m \times (N-m)} &\geq 0. \end{aligned}$$

See Dai and Singleton (2000) for a discussion on why these restrictions ensure the admissibility of the model, and on the uniqueness of the representation.

Appendix B

These formulae are all based on the presentation given by Bolder (2001). The restrictions implicit in the DATSM model are that both σ and $(\kappa - \lambda_2)$ must be strictly positive. For ease of estimation $\kappa\theta$ is treated as a single parameter.

In the one factor Gaussian model the risk neutral dynamics of the short rate are

$$dr_t = (\kappa\theta - \kappa r_t)dt + \sigma d\widetilde{W}_t$$

and with the price of risk specified as $\Lambda_{it} = \frac{\lambda_1 + \lambda_2 r_t}{\sigma}$, the real world dynamics are

$$dr_t = ((\kappa\theta + \lambda_1) - (\kappa - \lambda_2)r_t)dt + \sigma dW_t.$$

Then the affine yields for tenor τ are given by the standard Vasicek formulae

$$Y(\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau}r_t$$

where

$$B(\tau) = \frac{1 - e^{-\kappa\tau}}{\tau}$$

and

$$A(\tau) = \frac{\kappa^2\theta - \frac{1}{2}\sigma^2}{\kappa^2} (B(\tau) - \tau) - \frac{\sigma^2}{4\kappa} B^2(\tau).$$

Bond options with maturity T and strike K on a bond with tenor τ , P_t^τ , are given by

$$\begin{aligned} c_t &= P_t^\tau \mathcal{N}(d_1) - P_t^T K \mathcal{N}(d_2) \\ p_t &= P_t^T K \mathcal{N}(-d_2) - P_t^\tau \mathcal{N}(-d_1) \end{aligned}$$

where

$$\begin{aligned} d_1 &= \frac{\log\left(\frac{P_t^\tau}{K P_t^T}\right) + \frac{1}{2}\sigma_B^2}{\sigma_B} \\ d_2 &= d_1 - \frac{1}{2}\sigma_B \end{aligned}$$

and

$$\sigma_B^2 = \sigma^2 \frac{1 - e^{-\kappa(\tau-T)}}{\kappa^2} \frac{(1 - e^{-2\kappa T})}{2\kappa}.$$

Explicit formulae for the transition density are easily available with standard Ornstein-Uhlenbeck results. For given times $t_2 > t_1$ the distribution of r_{t_2} given r_{t_1} is Gaussian with

$$E^\mathbb{P}[r_{t_2}|r_{t_1}] = \frac{\kappa\theta + \lambda_1}{\kappa - \lambda_2} \left(1 - e^{-(\kappa - \lambda_2)(t_2 - t_1)}\right) + e^{-(\kappa - \lambda_2)(t_2 - t_1)} r_{t_1}$$

and

$$Var^\mathbb{P}[r_{t_2}|r_{t_1}] = \frac{\sigma^2}{2(\kappa - \lambda_2)} \left(1 - e^{-2(\kappa - \lambda_2)(t_2 - t_1)}\right).$$

References

- Almeida, C., Graveline, J. J. and Joslin, S. (2011). Do interest rate options contain information about excess returns?, *Journal of Econometrics* **164**(1): 35 – 44. Annals Issue on Forecasting.
URL: <http://www.sciencedirect.com/science/article/pii/S0304407611000340>
- Bierman, G., Belzer, M., Vandergraft, J. and Porter, D. (1990). Maximum likelihood estimation using square root information filters, *Automatic Control, IEEE Transactions on* **35**(12): 1293–1298.
- Björk, T. (2004). *Arbitrage theory in continuous time*, Oxford university press.
- Bolder, D. (2001). Affine term-structure models: Theory and implementation, *Working Paper 01-15*, Bank of Canada.
- Chen, R.-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.
- Christoffersen, P., Dorion, C., Jacobs, K. and Karoui, L. (2014). Nonlinear Kalman filtering in affine term structure models, *Working Paper 14-04*, Centre interuniversitaire sur le risque, les politiques économiques et l’emploi.
- Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica: Journal of the Econometric Society* pp. 385–407.
- Dai, Q. and Singleton, K. (2003). Term structure dynamics in theory and reality, *Review of Financial Studies* **16**(3): 631–678.
- Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models, *The Journal of Finance* **55**(5): 1943–1978.
- Dai, Q. and Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure, *Journal of financial Economics* **63**(3): 415–441.
- Dothan, L. U. (1978). On the term structure of interest rates, *Journal of Financial Economics* **6**(1): 59–69.
- 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**(2): 111–135.

- Duarte, J. (2004). Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* **17**(2): 379–404.
- 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**(02).
- Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates, *Mathematical finance* **6**(4): 379–406.
- Gupta, N. and Mehra, R. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations, *Automatic Control, IEEE Transactions on* **19**(6): 774–783.
- Haykin, et al. (2001). *Kalman Filtering and Neural Networks*, John Wiley & Sons.
- Haykin, S. (2001). *Kalman Filtering and Neural Networks*, 1 edn, John Wiley & Sons, chapter 1 Kalman Filters, pp. 1–20.
- Julier, S. J. and Uhlmann, J. K. (1996). A general method for approximating nonlinear transformations of probability distributions, *Technical report*, Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford.
- Julier, S. J. and Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems, *Int. symp. aerospace/defense sensing, simul. and controls*, Vol. 3, Orlando, FL, pp. 3–2.
- Kulikova, M. (2009). Likelihood gradient evaluation using square-root covariance filters, *Automatic Control, IEEE Transactions on* **54**(3): 646–651.
- 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. (1938). Some theoretical problems suggested by the movements of interest rates, bond yields and stock prices in the united states since 1856, *NBER Books* .
- Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, Vol. 111, John Wiley & Sons.
- Piazzesi, M. (2009). *Handbook of Financial Econometrics*, Vol. 1, Elsevier, chapter Affine Term Structure Models, pp. 691 – 766.
- Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**(2): 177–188.
- Wan and van der Merwe (2001). *Kalman Filtering and Neural Networks*, 1 edn, John Wiley & Sons, chapter 7 The Unscented Kalman filter, pp. 221–280.
- White, H. (1996). *Estimation, inference and specification analysis*, number 22, Cambridge university press.