BANK OF ENGLAND

Working Paper No. 481

Likelihood inference in non-linear term structure models: the importance of the lower bound

Martin Andreasen and Andrew Meldrum

December 2013

BANK OF ENGLAND

Working Paper No. 481

Likelihood inference in non-linear term structure models: the importance of the lower bound

Martin Andreasen[(1)] and Andrew Meldrum[(2)]

Abstract

This paper shows how to use adaptive particle filtering and Markov chain Monte Carlo methods to estimate quadratic term structure models (QTSMs) by likelihood inference. The procedure is applied to a quadratic model for the United States during the recent financial crisis. We find that this model provides a better statistical description of the data than a Gaussian affine term structure model. In addition, QTSMs account perfectly for the lower bound whereas Gaussian affine models frequently imply forecast distributions with negative interest rates. Such predictions appear during the recent financial crisis but also prior to the crisis.

Key words: Adaptive particle filtering, Bayesian inference, higher-order moments, PMCMC, quadratic term structure models.

JEL classification: C1, C58, G12.

---

(1) Aarhus University. Email: mandreasen@econ.au.dk
(2) Bank of England. Email: andrew.meldrum@bankofengland.co.uk

The views expressed are those of the authors and do not necessarily reflect those of the Bank of England or members of the Monetary Policy Committee or Financial Policy Committee. We thank Francis Breedon, Sean Holly, Oreste Tristani, Donald Robertson, participants at the fourth CSDA International Conference on Computational and Financial Econometrics 2010 and three anonymous referees for useful comments and discussions. Martin Andreasen acknowledges access to computer facilities provided by the Danish Center for Scientific Computation (DCSC) and financial support to Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. This paper was finalised on 8 October 2013.

The Bank of England's working paper series is externally refereed.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

Summary

The yields on government bonds are of interest to monetary policy makers partly because they reflect financial market participants' expectations of future policy rates. As with any asset price, however, they also reflect the additional return - or ‘risk premia’ - that investors require to compensate them for the uncertainty surrounding future returns on the asset. And yields also play an independent and important role in the transmission mechanism of monetary policy. Central banks therefore make widespread use of models to both forecast yields and to decompose them into expectations of future policy rates and risk premia.

Perhaps the most popular type of model among central bankers, academics and financial market practitioners is the ‘affine term structure model’ (ATSM), where yields are a linear function of some underlying variables. This makes for tractability. These statistical models of bond yields are consistent with the standard assumption that investors cannot make risk-free arbitrage profits (ie, investors cannot make profits by buying and selling different categories of bonds in such a way that the expected return from holding that portfolio is positive). But ATSMs do not impose the restriction that nominal interest rates are subject to a lower bound. This feature of the model is likely to have become more important in recent years given the historically low level of nominal bond yields.

Quadratic term structure models (QTSMs), in contrast, are more general and can be specified to be consistent with a lower bound. They are, however, substantially harder to estimate than ATSMs. This paper demonstrates for the first time that it is possible to use a numerical technique known as ‘Particle Markov chain Monte Carlo’ to estimate these models. This technique involves the random generation of many different candidate values for the model parameters. Each candidate draw of parameter values depends on the previous draws. Whether the candidate is accepted or rejected depends in part on how well it matches the observed data. This in turn is established using a different simulation technique known as a ‘particle filter’, which involves simulating many possible scenarios from the model and establishing how likely each scenario is given the observed data. Once we have considered a sufficiently large number of draws, the distribution of possible parameters will cease to change, known as convergence. This way of estimating these models has some desirable features relative to the methods that have been used previously. In particular, the statistical properties of the estimated model parameters can be more accurately established.

We apply the technique to estimate a QTSM using US nominal bond yields for the period 1962-2012. We find that the presence of the zero lower bound on nominal interest rates has important implications when using term structure models to forecast bond yields and short-term policy interest rates. Standard ATSMs imply around a 5-15% probability of negative policy rates in ten years' time throughout the estimation period. During the recent financial crisis the ATSM implies probabilities of negative policy rates of more than 40% at shorter horizons. The QTSM rules this out by construction. The difference between policy rate forecasts from the two models becomes more important as bond yields approach the lower bound.

1 Introduction

The lower bound on nominal interest rates is often ignored in dynamic term structure models, partly because bond yields have been far from this bound during most of the post-war period in many countries. However, the recent financial crisis has driven bond yields to historically low levels, with short rates at or close to zero in several countries. These low interest rates have highlighted a key shortcoming of many affine term structure models (ATSMs), namely their inability to account for the zero lower bound.

The class of quadratic term structure models (QTSMs) can address this shortcoming, as emphasized by Ahn et al. (2002) and Lieppold and Wu (2002). These models are, however, quite challenging to estimate because their likelihood function does not have a closed-form expression when accounting for measurement errors in bond yields. It is therefore only possible to approximate the likelihood function for a QTSM by Monte Carlo simulation. This procedure has been considered numerically infeasible for many years, and researchers have therefore used alternative estimation methods when taking QTSMs to the data. One strand of the literature has used Quasi Maximum Likelihood (QML) based on local linearisations and the Kalman filter, although the asymptotic distribution of this estimator is unknown (see Taulbjerg (2002), Li and Zhao (2006), Kim (2007), Kim and Singleton (2011), among others).[1] Another approach is to use Generalised Method of Moments (GMM) when moments have closed-form solutions as in Lieppold and Wu (2003). The well-known shortcomings of this alternative is that GMM may be less efficient than Maximum Likelihood (ML) and GMM is often sensitive to the selected moments. The obvious alternative to GMM is therefore to follow Gallant and Tauchen (1996) and use Efficient Methods of Moments (EMM), where moments are determined from a flexible auxiliary model that allows EMM to attain the same asymptotic efficiency as ML. An application of EMM to QTSMs is provided in Ahn et al. (2002). However, recent Monte Carlo evidence by Duffee and Stanton (2012) indicates that EMM may have poor finite sample properties when data display the same high degree of persistence as bond yields. Moreover, GMM and EMM do not estimate the latent factors, which prevent derivation of model-implied time series for term premia, predictions of future interest rates, etc. Given these shortcomings of QML, GMM, and EMM, likelihood inference still remains a

[1] Lieppold and Wu (2007) use the unscented Kalman filter for QML estimation of QTSMs, but the asymptotic properties of this estimator are also unknown.

very attractive estimation approach with no clear substitutes.

In this paper we show how to make likelihood inference feasible for QTSMs. We focus on the case where bond yields are measured with errors and a particle filter is used to get an unbiased estimate of the likelihood function. In this setting, model parameters can be estimated by Bayesian inference using Markov chain Monte Carlo (MCMC) as shown in Andrieu et al. (2010). To make the estimation process feasible, it is essential to get a fast and reliable estimate of the likelihood function. This is challenging for dynamic term structure models due to the presence of multiple latent factors and typically small measurement errors in bond yields. We overcome these difficulties by using the adaptive particle filter by Andreasen (2010) to improve the efficiency of particle filtering and hence make likelihood inference feasible for QTSMs. Although this paper focuses exclusively on QTSMs, our procedure applies more generally to any non-linear and potentially non-Gaussian dynamic term structure model.

The suggested procedure is illustrated by estimating the most flexible two-factor QTSM on US data from 1962 to 2012. We highlight the following results. First, the adaptive particle filter by Andreasen (2010) is shown to provide a fast and reliable estimate of the log-likelihood function for QTSMs, unlike the standard particle filter by Gordon et al. (1993). Second, adopting a Bayesian estimation approach when studying the zero lower bound for bond yields may be useful, because this constraint may easily imply that parameters are at the boundary of their domain, thus rendering standard classical inference problematic. Third, the quadratic model gives a much better statistical description of the data than a Gaussian ATSM. Bayes factors and posterior Bayes factors are clearly in favour of the quadratic model. Fourth, the estimated quadratic model displays strong non-linearities and ensures that all forecast distributions imply positive interest rates. This property is in contrast to the Gaussian ATSM where a large proportion of the forecast distributions imply negative interest rates during the recent financial crisis but also prior to the crisis. Based on these findings, we therefore conclude that the zero lower bound is an important constraint to account for in US data.

The rest of this paper is organised as follows. Section 2 presents the canonical QTSM and its state space representation. Likelihood estimation of this model is discussed in Section 3. The QTSM is estimated in Section 4, where we also analyse the importance of the zero lower bound. Concluding comments are provided in Section 5.

2 Quadratic term structure models

This section introduces the canonical QTSM and its state space representation. For numerical convenience, we follow Realdon (2006) and consider a discrete-time version of the model because it gives a closed-form solution to zero-coupon bond prices even with multiple correlated factors. This timing assumption also simplifies the subsequent estimation as we can ignore issues related to discretisation of continuous-time diffusions.[2]

2.1 Model assumptions and identification

We assume that the one-period risk-free interest rate r_t is a quadratic function of an $m \times 1$ vector of latent factors $\mathbf{x}_t$, i.e.

$$r_t = \alpha + \boldsymbol{\beta}' \mathbf{x}_t + \mathbf{x}_t' \boldsymbol{\Psi} \mathbf{x}_t. \tag{1}$$

Here α is a scalar, $\boldsymbol{\beta}$ is an $m \times 1$ vector, and Ψ is an $m \times m$ matrix. The factors evolve according to a VAR(1) process under the risk-neutral measure $\mathbb{Q}$

$$\mathbf{x}_{t+1} = (\mathbf{I} - \boldsymbol{\Phi})\, \mathbf{x}_t + \boldsymbol{\Phi}\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}}, \tag{2}$$

where $\boldsymbol{\varepsilon}_{t+1}^{\mathbb{Q}} \sim \mathcal{NID}\left(\mathbf{0}, \mathbf{I}\right)$. The mean level of the factors is given by $\boldsymbol{\mu}$ with dimensions $m \times 1$, while Φ and Σ are $m \times m$ matrices. Absence of arbitrage implies an equivalent risk-neutral probability measure, and the price at time t of an i-period zero-coupon bond $P_{t,i}$ is therefore

$$P_{t,i} = E_t^{\mathbb{Q}} \left[\exp\left\{ -r_t \right\} P_{t+1,i-1} \right]. \tag{3}$$

Given these assumptions, bond prices are exponential-quadratic in the factors, i.e.

$$P_{t,i} = \exp\left\{ A_i + \mathbf{B}_i' \mathbf{x}_t + \mathbf{x}_t' \mathbf{C}_i \mathbf{x}_t \right\} \tag{4}$$

for $i = 1, 2, ..., k$ with the recursive formulae for A_i, $\mathbf{B}_i$, and $\mathbf{C}_i$ derived in Realdon (2006).

We consider a standard affine specification for the market price of risk, i.e.

$$\mathbf{f}\left(\mathbf{x}_t\right) = -\mathbf{f}_0 - \mathbf{f}_1 \mathbf{x}_t, \tag{5}$$

[2] Accounting for discretisation of continuous-time diffusions in the particle filter is straightforward (see for instance Pitt (2002)).

where $\mathbf{f}_0$ and $\mathbf{f}_1$ having dimensions $m \times 1$ and $m \times m$, respectively.

Within a QTSM, the policy rate, and hence all bond yields, are non-negative if $\alpha \geq \frac{1}{4}\boldsymbol{\beta}'\boldsymbol{\Psi}^{-1}\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ is positive semi-definite (Realdon (2006)). Restricting nominal yields to be non-negative is desirable from an economic perspective because investors always have the option of holding banknotes, paying zero interest. In practice, holding cash is not costless and there are a few cases where nominal interest rates have been slightly negative. For example, the central bank of Denmark has remunerated excess reserves at interest rates as low as -0.2% since July 2012. Another example is the US, where secondary market rates on four-week Treasury bills have occasionally been negative. But these violations of the zero lower bound in the US have been small (no lower than -0.01% in the daily data published by the Federal Reserve Board) and short-lived. On the other hand, all interest rates used in this paper for the US are positive.[3]

The most flexible normalization is adopted for identification (see Ahn et al. (2002) and Realdon (2006)). That is, $\boldsymbol{\Psi}$ is symmetric and all diagonal elements equal one, $\boldsymbol{\mu} \geq 0$, $\boldsymbol{\Sigma}$ is a diagonal matrix, and $\boldsymbol{\Phi}$ is lower triangular. To impose the zero lower bound, we require $\alpha \geq 0$, $\boldsymbol{\beta} = \mathbf{0}$, and $\boldsymbol{\Psi}$ to be positive semi-definite.

2.2 The state space representation

Our specification for the market price of risk implies that the law of motion for the factors under the physical measure is given by

$$\mathbf{x}_{t+1} = (\mathbf{I} - \boldsymbol{\Phi} + \boldsymbol{\Sigma}\mathbf{f}_1)\,\mathbf{x}_t + \boldsymbol{\Phi}\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{f}_0 + \boldsymbol{\Sigma}\boldsymbol{\varepsilon}_{t+1}, \tag{6}$$

where $\boldsymbol{\varepsilon}_{t+1} \sim \mathcal{NID}\,(\mathbf{0}, \mathbf{I})$. We denote the implied conditional distribution for the factors by $p\,(\mathbf{x}_{t+1}|\mathbf{x}_t)$. The measurement equations are constructed from continuously compounded bond yields $y_{t,i} = -\frac{1}{i}\ln P_{t,i}$ for a selected number of maturities, $\mathcal{I} = \{i_1, i_2, ..., i_k\}$. As is standard practice when taking dynamic term structure models to the data, we account for measurement errors in these yields. Such errors may arise when extracting bond yields from a panel of coupon-bonds due to bid-ask spreads, non-synchronized trading, etc. These errors are denoted by $\mathbf{w}_t$ and we let $\mathbf{w}_t \sim \mathcal{NID}\left(\mathbf{0}, s_w^2\mathbf{I}\right)$. The measurement equations for our QTSM are then

[3] It would be straightforward to allow for a lower bound below zero within the QTSM, for instance by letting $\alpha < 0$.

given by

$$\begin{bmatrix} y_{t,i_1} \\ y_{t,i_2} \\ \dots \\ y_{t,i_k} \end{bmatrix} = \begin{bmatrix} -\frac{1}{i_1}\left(A_{i_1} + \mathbf{B}'_{i_1}\mathbf{x}_t + \mathbf{x}'_t\mathbf{C}_{i_1}\mathbf{x}_t\right) \\ -\frac{1}{i_2}\left(A_{i_2} + \mathbf{B}'_{i_2}\mathbf{x}_t + \mathbf{x}'_t\mathbf{C}_{i_2}\mathbf{x}_t\right) \\ \dots \\ -\frac{1}{i_k}\left(A_{i_k} + \mathbf{B}'_{i_k}\mathbf{x}_t + \mathbf{x}'_t\mathbf{C}_{i_k}\mathbf{x}_t\right) \end{bmatrix} + \begin{bmatrix} w_{i_1,t} \\ w_{i_2,t} \\ \dots \\ w_{i_k,t} \end{bmatrix}. \tag{7}$$

We define $\mathbf{y}_t \equiv \left[y_{t,i_1}\, y_{t,i_2} \dots y_{t,i_k}\right]'$ and let $p(\mathbf{y}_t|\mathbf{x}_t)$ denote the conditional distribution of $\mathbf{y}_t$ given the factors. Equations 6 and 7 define a non-linear state space system with k observed bond yields and m latent factors. The log-likelihood function for this system is denoted by $L(\boldsymbol{\theta}|\mathbf{y}_{1:T})$ where $\mathbf{y}_{1:T} \equiv \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T\}$ and $\boldsymbol{\theta}$ contains the model parameters.

3 Likelihood-based inference

This section discusses likelihood inference for the considered QTSM. We start in Section 3.1 by describing how to approximate the value of the log-likelihood function and how model parameters may be estimated from a classical perspective. Section 3.2 is devoted to Bayesian inference, and Section 3.3 addresses the challenging task of getting a fast and reliable estimate of the log-likelihood function.

3.1 Approximating the likelihood function and classical inference

Evaluation of the likelihood function for a QTSM requires computing a multidimensional integral over the latent factors $\mathbf{x}_{1:T} \equiv \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$. This is challenging because the integral does not have a closed-form expression, unlike in linear and Gaussian models where the solution is given by the Kalman filter. One way to proceed is to use a particle filter and estimate the log-likelihood function by repeated use of importance sampling and resampling (Doucet et al. (2001)). Particle filters only require distributions for $p(\mathbf{y}_t|\,\mathbf{x}_t)$ and $p(\mathbf{x}_{t+1}|\,\mathbf{x}_t)$, and they may therefore be applied to a wide class of non-linear and potentially non-Gaussian dynamic term structure models.

The parameters in the model can then be estimated by maximising the likelihood function implied by the particle filter, i.e. by Maximum Simulated Likelihood (MSL). However, the approximated likelihood function is not smooth in $\boldsymbol{\theta}$ due to the resampling step. It is therefore very challenging to optimise this objective function and compute asymptotic standard errors.[4] Reliable inference is further com-

[4]The particle filters by Pitt (2002) and Flury and Shephard (2009) are important exceptions where the estimated likelihood function is smooth in $\boldsymbol{\theta}$. The survey by Kantas et al. (2009) discusses other

plicated by the demanding requirement that the number of particles must tend to infinity for consistency of the MSL estimator (see for instance Hajivassiliou and Ruud (1994)).

These practical and theoretical problems related with MSL have to some extent reduced the use of MSL. Much focus from a classical perspective has therefore been devoted to simulation-based moment estimators (i.e. Simulated Method of Moments, Indirect Inference, and EMM) which are easier to implement compared to MSL.[5]

3.2 Bayesian inference by PMCMC

Andrieu et al. (2010) introduce particle Markov chain Monte Carlo (PMCMC) methods to estimate $\boldsymbol{\theta}$ from a Bayesian perspective when only an approximated value of the log-likelihood function is available using a particle filter. One of the suggested methods is the particle marginal Metropolis-Hastings (PMMH) algorithm to draw from the posterior distribution of $\boldsymbol{\theta}$. This sampler is similar to the well-known Metropolis-Hastings algorithm except the estimated likelihood function is used instead of the true likelihood function. They show that the equilibrium distribution for this Markov chain is unaffected by the Monte Carlo variation in the likelihood function. In other words, using an approximated likelihood function does not induce any bias in the estimates of $\boldsymbol{\theta}$, and this constitutes a clear advantage of PMCMC in comparison to MSL.[6]

However, the Monte Carlo variation in the likelihood function does affect the mixing properties of the PMMH algorithm as illustrated by Flury and Shephard (2011), Andrieu et al. (2010), among others. This is because a noisy estimate of the likelihood function makes it difficult for the algorithm to explore the parameter space as the chain may easily get stuck at certain points for a long time. Hence, a high Monte Carlo variation in the likelihood function increases the probability that the Markov chain does not converge or that it converges around a local mode.

3.3 Particle filters and QTSMs

As discussed by Flury and Shephard (2011), there is a computational trade-off in PMCMC estimation. If the estimation of the likelihood is more accurate, fewer

methods to address the mentioned problems for MSL.

[5]See Carrasco and Florens (2002) for an introduction to simulation-based moment estimators.

[6]In addition to the assumptions in Andrieu et al. (2010), this result relies on different random numbers being used in the particle filter throughout the Markov chain, as emphasised by Flury and Shephard (2011).

draws will be required before the Markov chain converges to the posterior distribution for the model parameters, but the longer it takes to evaluate the likelihood at each step in the chain. A reliable estimate of the likelihood function is unfortunately difficult to obtain for dynamic term structure models. This is due to the presence of multiple latent factors in these models and the typical finding that bond yields are observed with small measurement errors. The latter implies that the conditional distribution of yields given the factors $p\left(\mathbf{y}_t | \mathbf{x}_t\right)$ is very peaked and may therefore be difficult to approximate.

We follow Andrieu et al. (2010) and start by exploring the properties of the Standard Particle Filter (SPF) by Gordon et al. (1993) where the proposal distribution for the importance sampling step is given by the state transition distribution, i.e. $p\left(\mathbf{x}_{t+1} | \mathbf{x}_t\right)$. At the posterior mean parameters reported below, it was not feasible to use the SPF to evaluate the likelihood function of the QTSM without an extremely large number of particles. With $100,000$ particles, the filter diverged - i.e. the conditional density of all particles was too small to be represented with reasonable numerical accuracy - in 77 out of 100 attempts to evaluate it using different random number seeds. If we raise the standard deviation of the measurement error in bond yields from 22 to 40 basis points, the filter does not diverge if $100,000$ particles are used. Given this configuration, the time taken for each evaluation of the log-likelihood function is more than 15 seconds and the Monte Carlo variability in the estimated log-likelihood function is much too high to make estimation feasible (5.09 which is well above the range suggested by Flury and Shephard (2011)).[7]

The poor performance of the SPF mainly relates to the fact that bond yields in the next period, i.e. $\mathbf{y}_{t+1}$, are not used in the proposal distribution for the importance sampling step. Filters with this adaptive property were first discussed in Pitt and Shephard (1999), where they suggest the use of auxiliary variables.[8] The adaptive filter applied in this paper is the optimised Central Difference Particle Filter (CDPF) by Andreasen (2010), where the proposal distribution is generated as follows. Based on the work of Norgaard et al. (2000), the Central Difference Kalman Filter (CDKF) is first used to compute preliminary estimates of first and second moments of $p\left(\mathbf{x}_{1:t+1} | \mathbf{y}_{1:t+1}\right)$. These moments are then used to construct a Gaussian proposal distribution which therefore contains information from bond

[7]The computation times reported in this section were based on code written in Fortran 90 run on a PC with 3GHz Intel Core 2 Duo processors and 3.5GB RAM.

[8]Other well-known alternatives include local linearisation as in Doucet et al. (2000) or an MCMC step inside the particle filter (Gilks and Berzuini (2001)).

yields in the next period, i.e. $\mathbf{y}_{t+1}$. That is, the proposal distribution is given by

$$\mathbf{x}_{t+1}^{(i)} = \widehat{\mathbf{x}}_{t+1}^{CDKF} + \gamma_{t+1}\widehat{\mathbf{S}}_{\mathbf{x},t+1}^{CDKF}\boldsymbol{\epsilon}_{t+1}^{(i)} \quad \text{for } i = 1, 2, ..., N, \tag{8}$$

where $\boldsymbol{\epsilon}_{t+1}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Here, $\widehat{\mathbf{x}}_{t+1}^{CDKF}$ denotes the posterior mean in the CDKF, and $\widehat{\mathbf{S}}_{\mathbf{x},t+1}^{CDKF}$ is the Cholesky factor of the covariance matrix for this estimate. The role of the free parameter γ_{t+1} is to increase the variance of the proposal distribution to account for fat tails and other deviations from normality in $p(\mathbf{x}_{1:t+1}|\,\mathbf{y}_{1:t+1})$. Following the work by Richard and Zhang (2007) on efficient importance sampling, Andreasen (2010) determines the value of γ_{t+1} by maximising the effective sample size in every period - which is equivalent to minimising the distance to the optimal proposal distribution. This optimisation step is normally done using a subset of the particles $N_{opt} << N$, and only in the event where the effective sample size is lower than a given threshold N_{eff}^{low} are all particles used in the optimisation.[9]

At the posterior mean for parameters reported below, using only 5,000 particles and optimising γ_{t+1} if the effective number of particles is below 250, we find that it only takes 0.9 seconds to evaluate the likelihood function and the standard deviation of the log likelihood function across 100 evaluations of the filter is just 0.28. Based on these simulation results and the convergence of the Markov chains presented below, we conclude that it is possible to get a sufficiently fast and reliable estimate of the log-likelihood function for a two-factor QTSM to make estimation by likelihood inference feasible.

4 An estimated QTSM for the US

This section uses the optimised CDPF to estimate a QTSM on US data. We proceed as follows. Section 4.1 presents the data and Section 4.2 describes restrictions on the model and settings for the optimised CDPF. Estimation results for the QTSM are presented in Section 4.3, and we compare the performance of the QTSM to the ATSM in Section 4.4. The importance of the zero lower bound for bond yields and forecast distributions are studied in Section 4.5 with special focus devoted to the recent financial crisis.

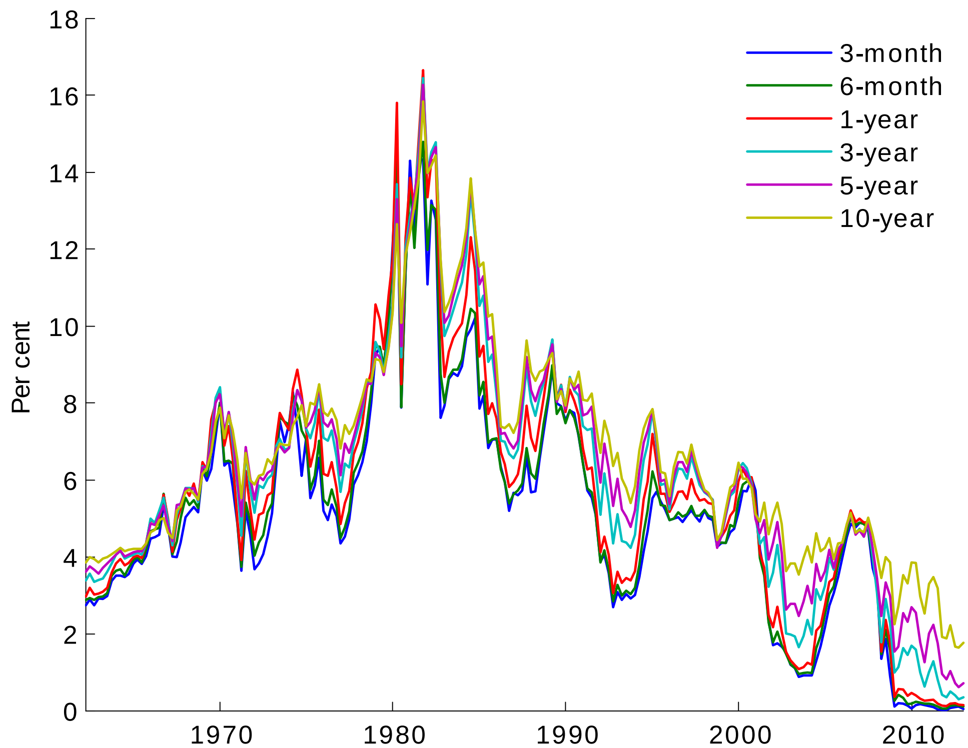
4.1 The data

We use end-of-quarter nominal zero-coupon bond yields for the US from 1962Q1 to 2012Q4. Six yields with maturities of 1, 2, 4, 12, 20 and 40 quarters are selected

[9] We refer to Andreasen (2010) for further details on the implementation of the optimised CDPF.

for the estimation. The data are from the Federal Reserve Economic Database provided by the Federal Reserve Bank of St Louis. Figure 1 displays the data and summary statistics are provided in Table 1. The mean term structure is upward sloping with average spreads between the 10-year and 1-quarter bond yields of about 1.5 percentage points. Broadly speaking, the volatility of quarter-on-quarter changes in yields falls gradually with maturity. We also note that the distributions of changes in bond yields are asymmetric and fat-tailed relative to a Gaussian distribution as evident from values of skewness and kurtosis in Table 1; skewness is negative at all maturities and kurtosis is decreasing with maturity.

Figure 1: US bond yields, 1962Q1-2012Q4



4.2 Priors

Following Chib and Ergashev (2009), we impose the prior that the unconditional yield curve is upward sloping, i.e. $y_{t,i_1} < y_{t,i_2} < ... < y_{t,i_k}$. We also impose the loose prior that the unconditional averages of all yields are between 0% and 10%.

Table 1: Summary statistics for US bond yields, 1962Q1-2012Q4

Maturity (quarters)	1	2	4	8	20	40
Mean (Percentage points)	5.10	5.25	5.68	6.09	6.33	6.63
Std. dev. (Percentage points)	0.97	0.95	1.04	0.86	0.79	0.66
Skewness	-1.95	-2.05	-1.64	-0.72	-0.62	-0.57
Kurtosis	13.46	15.57	14.36	4.86	4.03	2.72

The mean is for the level of bond yields. All other moments are computed for quarter-on-quarter changes in bond yields.

In addition, we impose all eigenvalues of $\mathbf{I} - \mathbf{\Phi}$ and $\mathbf{I} - \mathbf{\Phi} + \mathbf{\Sigma f}_1$ to be less than one in absolute value, meaning that the factor dynamics are stable under both the risk-neutral and physical probability measures. Apart from these restrictions, no further prior information is imposed for the estimation, i.e. we use flat priors on all coefficients.

We sample from the posterior distribution for the parameters using the Delayed Rejection Adaptive Metropolis algorithm of Haario et al. (2006), details of which are provided in Appendix 1. We run a Markov chain with a length of $100,000$. Convergence of this chain following a burn-in of $b_\theta = 50,000$ draws is verified in two ways. First, cumulated averages and trace plots are plotted to identify any obvious problems of convergence. Second, we use the formal test of Geweke (1992) to check that the posterior means from the first half of the chain (after discarding the burn-in) are not significantly different from the means in the second half of the chain.

4.3 Estimation results

Estimation results are reported in Table 4 in Appendix 2. Most parameters are estimated quite accurately as seen from the 5^{th} and 95^{th} percentiles of the posterior distributions. Given the normalization of the QTSM, the lower bound of the short interest rate r_t is given by α. Our mean posterior estimate of α is 5.2993×10^{-5}, and we therefore have a lower bound of just over 0.02% for the annualized short rate, which seems quite realistic. This is contrary to the results in Ahn et al. (2002) where the lower bound for the short rate is estimated to have an unrealistically high level (their sample finishes in 1991, well before the current period of low interest rates). Plots of the posterior distributions in Figure 11 in Appendix 2 show that the restriction $\alpha \geq 0$ is binding and the distribution for α is skewed to the right. We also observe a skewed distribution for ϕ_{11} and μ_2, as constraints on these parameters are binding. Hence, classical inference for this model may be challenging because several parameters are close to their boundaries, including those that enforce positive interest rates.[10] We illustrate this further by estimating the model by QML, using a quasi-likelihood function based on the CDKF. The QML estimate of α is 1.8722×10^{-13}, which is extremely close to the zero lower bound. A large part of the QML asymptotic density covers the invalid negative region (here the percentiles are based on 100,000 simulations from the asymptotic distribution).

4.4 Comparing QTSMs to ATSMs

This section compares some of the properties implied by the quadratic model to a Gaussian ATSM, i.e. a model similar to the one presented in Section 2 except $\mathbf{\Psi} = \mathbf{0}$. The ATSM is estimated from a Bayesian perspective using the Kalman filter to evaluate the likelihood function and the same restrictions as for the quadratic model described in Section 4.2. We start by reporting log-marginal data densities $\log p\left(y_{1:T}\right)$ for the quadratic and affine models in Table 2. We have $\log p\left(y_{1:T}\right) \approx -485$ in the quadratic model and $\log p\left(y_{1:T}\right) \approx -566$ in the affine model. However, the presence of flat priors in both models means that the two marginal data densities are not directly comparable because the resulting Bayes factor contains an undefined constant. To get an idea for the size of this constant, consider the case where all parameters in the two models are uniformly distributed on $[-C, C]$ where $C > 0$. The quadratic model has four additional parameters compared to the affine model (α, ψ_{21}, μ_1, μ_2), and the Bayes factor is therefore $e^{81-4\log(2C)}$. A very conservative value of C is 20, which gives a Bayes factor of e^{66} in favour of the quadratic model. Following the work of Aitkin (1991), another procedure for model comparison is to compute the posterior Bayes factor which is the ratio of the mean posterior likelihood in the two models. Table 2 implies a posterior Bayes factor of e^{77} in favour of the quadratic model. Thus, the stochastic structure implied by the quadratic model is more in line with the data than the structure implied by the affine model. Ahn et al. (2002) reach a similar conclusion, as the QTSM outperforms an affine model when measured in terms of the ability to match moments from the auxiliary model.

The difference between the quadratic and affine models is the non-linear terms in bond yields. Our finding that the log-marginal data density is substantially higher for the quadratic models must therefore imply that these non-linear terms

[10] Points estimates on the boundary of their domain are also evident in the QML estimates by Kim and Singleton (2011) on Japanese bond yields after 1995.

Table 2: Marginal data densities and posterior means of the likehood function

	Truncation parameter	QTSM	ATSM
$\log p(\mathbf{y}_{1:T})$	0.1	-484.77	-565.09
$\log p(\mathbf{y}_{1:T})$	0.3	-485.95	-565.70
$\log p(\mathbf{y}_{1:T})$	0.5	-486.22	-566.00
$\log p(\mathbf{y}_{1:T})$	0.7	-486.45	-566.19
$\log(E[p(\mathbf{y}_{1:T})])$	-	-492.63	-570.43

All marginal data densities are computed by the harmonic mean estimator using a truncated multivariate normal distribution.

play an important role. Another way to evaluate the impact of these terms is to compute unconditional higher-order moments such as skewness and kurtosis. This is done in Table 3 using the mean posterior parameter estimates. The model-implied moments are computed based on a simulated time series with $1,000,000$ observations. For the QTSM we see that quarter-on-quarter changes in bond yields have almost no skewness, and kurtosis ranging from 4.88 at the one-quarter maturity to 3.96 at the ten-year maturity. The magnitude of these deviations from the Gaussian distribution are not sufficient to match the unconditional moments in the data. This is somewhat similar to the results in Ahn et al. (2002) as they conclude that QTSMs with Gaussian factor dynamics cannot fully match the conditional volatility in US bond yields.

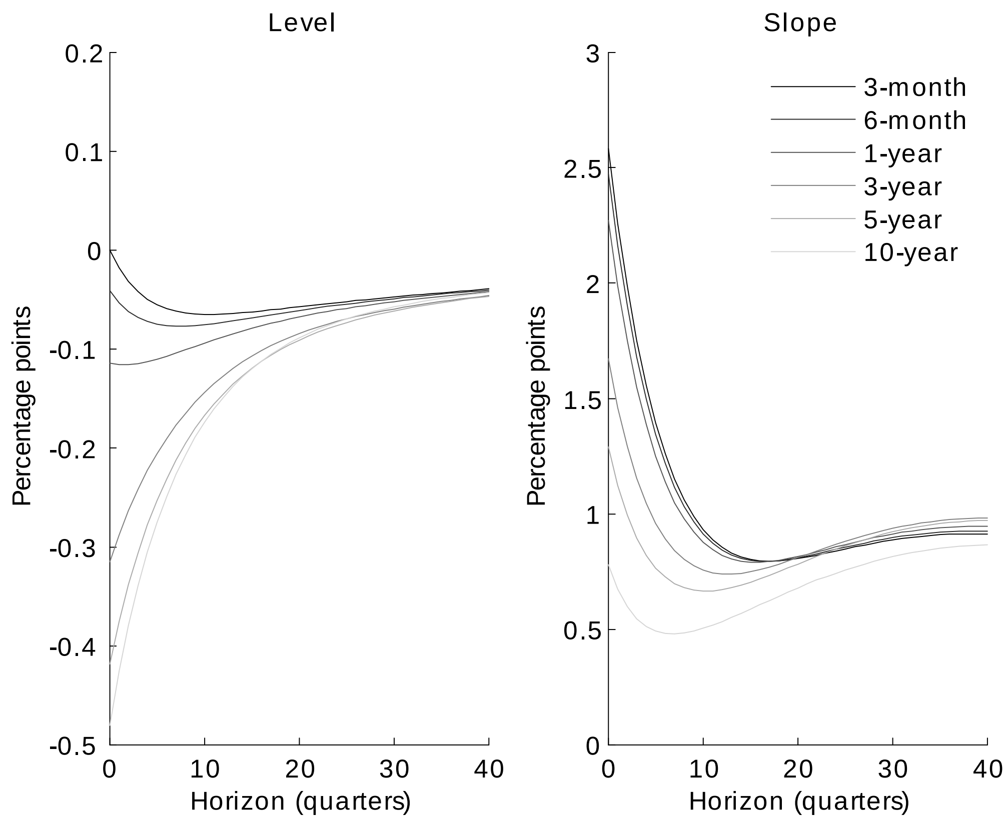
Table 3: Skewness and kurtosis in quarter-on-quarter changes in bond yields

Maturity (quarters)	1	2	4	8	20	40
Skewness:						
QTSM	0.01	0.01	0.00	-0.00	-0.00	-0.00
Data	-1.95	-2.05	-1.64	-0.72	-0.62	-0.57
Kurtosis:						
QTSM	4.88	4.75	4.57	4.22	4.07	3.96
Data	13.46	15.57	14.36	4.86	4.03	2.72

Moments in the QTSM are computed based on a time series with 1 million observations. All moments in the QTSMs are computed at the mean of the posterior distributions.

A third way to illustrate the impact of the non-linear terms in the QTSM is to compare their impulse response functions to those from the affine model. As a benchmark, Figure 2 reports the impulse response functions in the ATSM following a positive shock. (All impulse response functions reported in this paper are estimated as the average across 10,000 draws from the joint posterior distribution of the parameters.) Shocks to the first factor primarily affect long-term interest rates, while shocks to the second factor primarily affect the short-term interest rate. As is common in the term structure literature, we therefore refer to these factors as driving the 'level' and 'slope' of the term structure, respectively.

Figure 2: Impulse response functions from the ATSM
Impulse response functions for a one standard deviation shock. The responses are reported in percentage points from the steady state.



The corresponding responses in the QTSMs are more complicated due to the non-linear terms, and these functions therefore depend on i) the sign of the shock, ii) the size of the shock, and iii) the initial factor values. To illustrate some of these dependencies we compute the generalised impulse response functions (GIRFs) proposed by Koop et al. (1996) for positive and negative shocks and at two different dates: i) 1992Q4 where the initial factors imply bond yields between about 3% and 7% and ii) 2012Q4 where the initial factors give low bond yields close to the zero lower bound at short maturities.

Figures 3 and 4 show GIRFs following shocks to the first and second factor, respectively. In each case, the top row of sub-plots shows impulse responses for 1992Q4 and the bottom row for 2012Q4. The first column shows impulse responses for a positive shock and the second column for a negative shock. The broad shapes of the impulse responses are similar to those from the ATSM, in that the first factor (the 'level') has a greater impact on long-term yields and the second factor (the 'slope') has a greater impact on short-term yields. For both the level and slope, the responses are larger in 1992Q4 (interest rates are far from the zero-lower bound) compared with the responses in 2012Q4 (interest rates are close to the bound). For a given date, responses to a level shock are broadly symmetric. But the response of short rates to a positive slope shock is larger than the response to a negative slope in 2012Q4. This is because short rates are bounded below by the zero lower bound. These differences indicate that the estimated quadratic model displays strong non-linearities.

Figure 3: Impulse response functions to the level factor in the QTSM
Impulse response functions for a one standard deviation shock to the slope factor. The responses are reported in percentage points (pp) from the steady state.

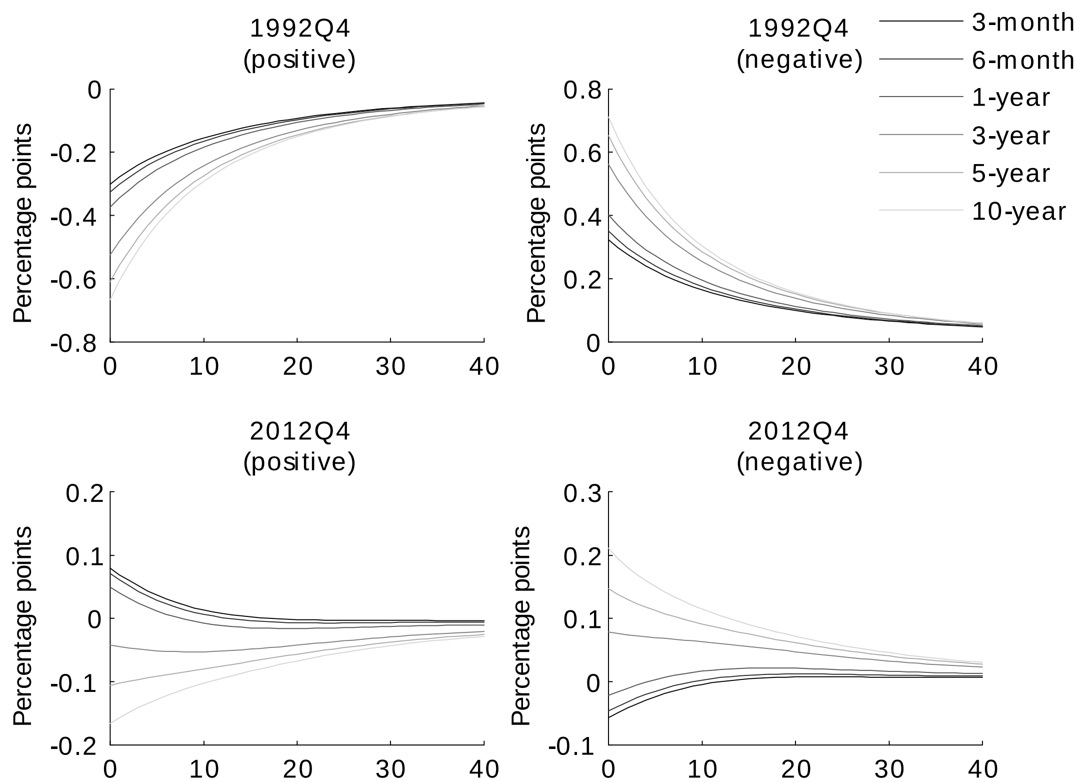
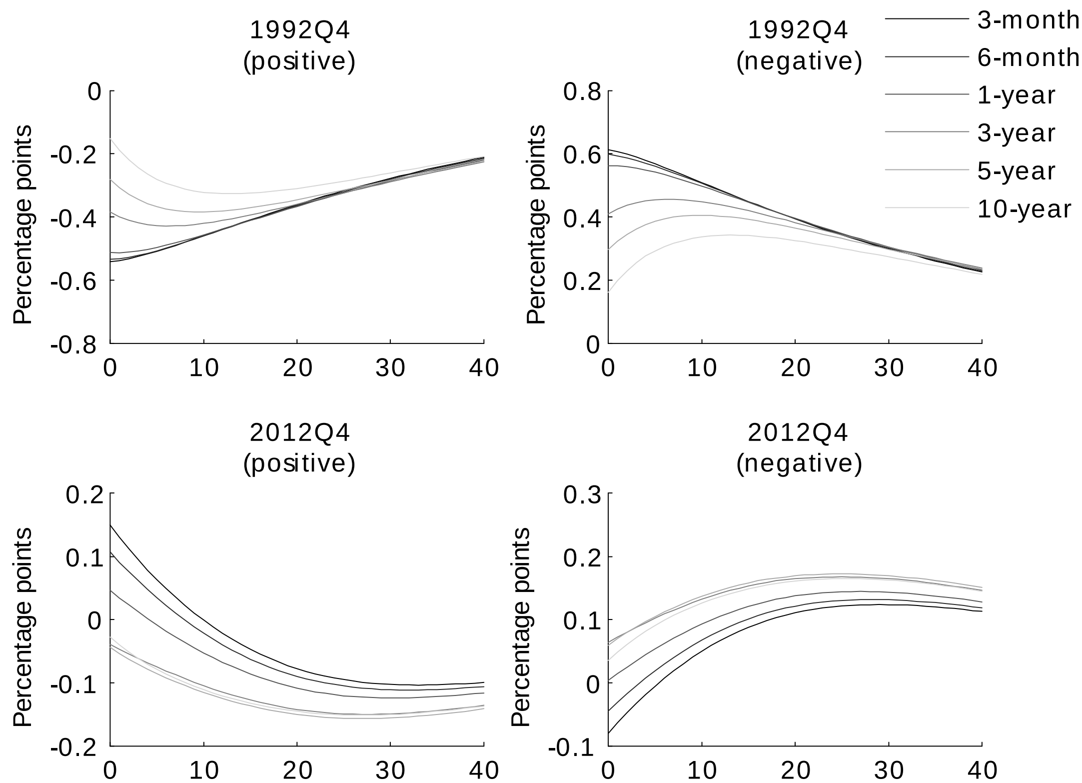


Figure 4: Impulse response functions to the slope factor in the QTSM
Impulse response functions for a one standard deviation shock to the slope factor. The responses are reported in percentage points (pp) from the steady state.



Dynamic term structure models are often applied to extract measures of term premia. We therefore end this section by comparing the derived measures of term premia from the considered models in Figures 5 and 6. Following Dai and Singleton (2002), term premia are here defined as the difference between the long rate and average expected future short rates, i.e.

$$TP_{t,\tau} = y_{t,\tau} - \frac{1}{\tau} \textstyle\sum_{i=0}^{\tau-1} E_t \left[r_{t+i} \right] \tag{9}$$

in period t for maturity τ. The non-linear terms in the QTSM appear in the long rate $y_{t,\tau}$ and in the expected future short rates $\frac{1}{\tau} \sum_{i=0}^{\tau-1} E_t \left[r_{t+i} \right]$, making it difficult a priori to assess their impact on term premia. But it is striking that estimates of term premia are broadly similar across the two models (all estimates of term premia are computed as averages across $10,000$ draws from the posterior densities for the parameters). Some more noticeable differences do emerge towards the end of the sample, when yields are closer to the lower bound. This suggests that estimates of term premia from the ATSM may be less robust when interest rates are low.

Figure 5: Model-implied term premia from the QTSM

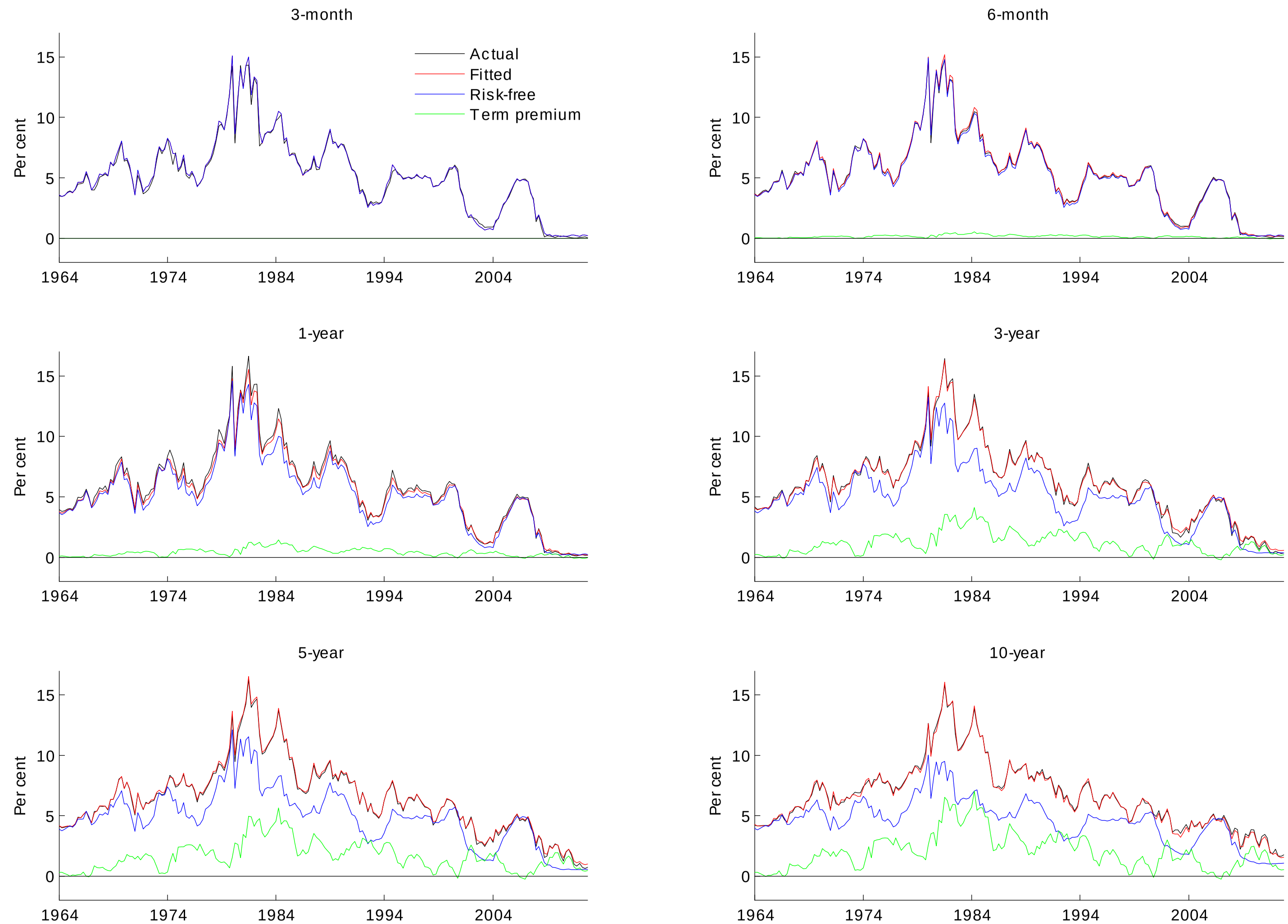
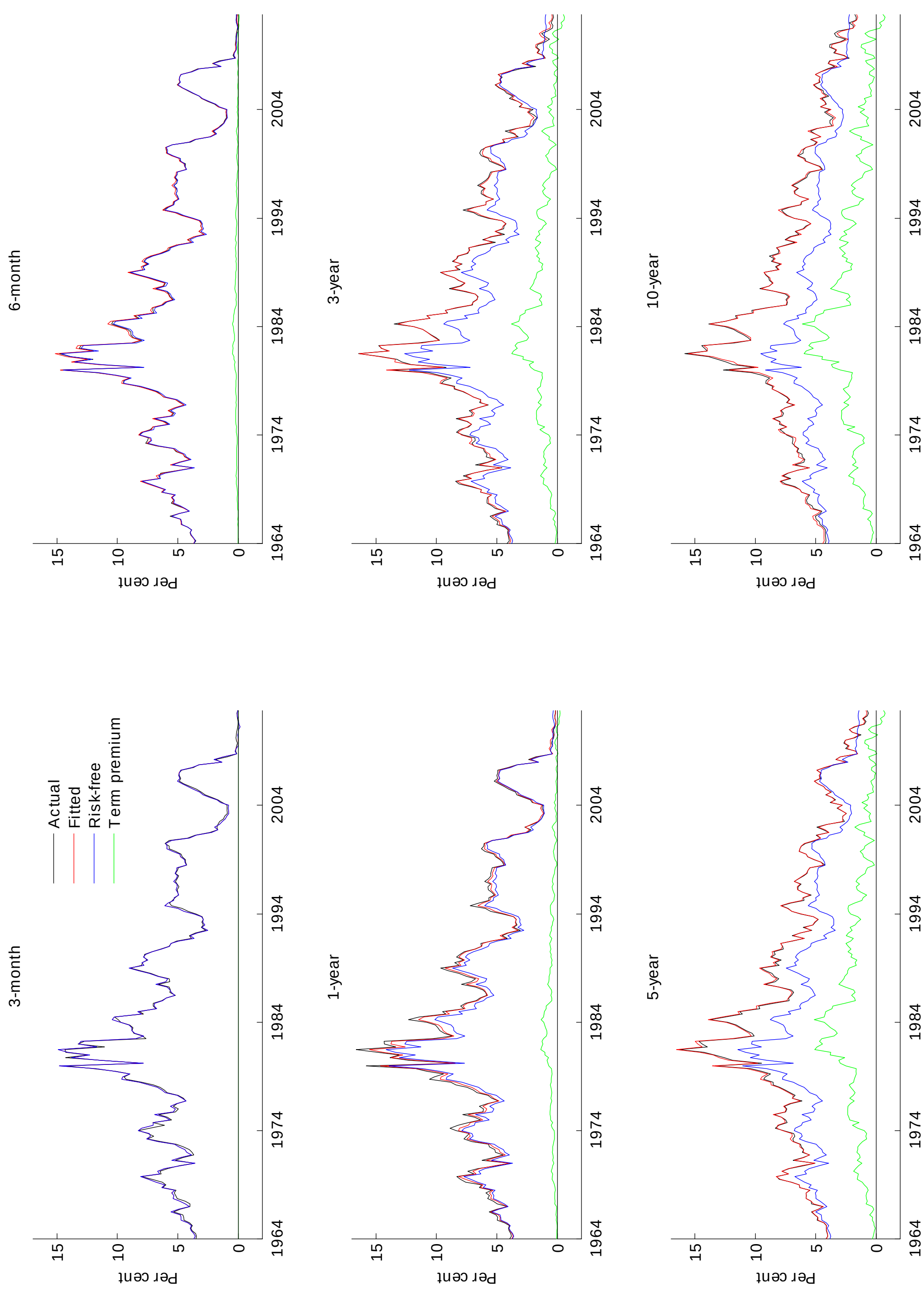


Figure 6: Model-implied term premia from the ATSM



To summarise, we find that the estimated quadratic model provides a better statistical description of the data than a Gaussian ATSM. This difference relates to the presence of sizeable non-linearities in the quadratic model, which leads to non-trivial higher-order moments in bond yields and factor-dependent impulse response functions.

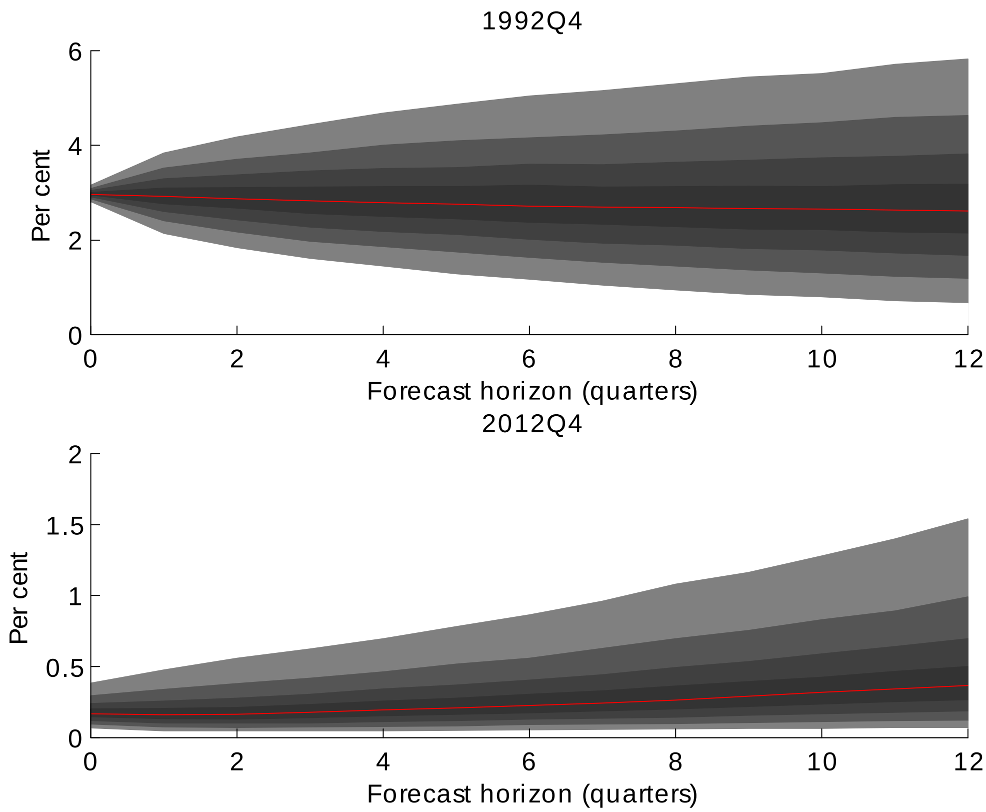
4.5 The importance of the zero lower bound

This section studies the importance of the zero lower bound in the considered affine and quadratic term structure models. Fitted one-quarter rates from the ATSM are not always positive, falling as low as -0.14% in 2011Q2. In contrast, the QTSM ensures that fitted short-term interest rates remain positive by construction. Moreover, forecast distributions implied by the ATSM also assign probability to negative interest rates even when the model implied short rate is positive. We illustrate this point in Figures 8 and 9 which show forecast distributions for the one-quarter interest rate at different forecast horizons for two dates: 1992Q4 and 2012Q4. Here, the central red line shows the median forecast path, and each progressively lighter shading regions cover ten percentiles of the distributions. In total, the shaded areas cover 80% of the distributions. For the affine model, we see that the median projection in 2012Q4 remains above the zero lower bound but there is nevertheless a high probability of negative interest rates at all horizons. This shortcoming of the model is due to the fact that these forecast distributions are Gaussian and hence symmetric. The quadratic model does not have the same problem and perfectly accounts for the zero lower bound at all forecast horizons. This is possible because bond yields in this model are a mixture of Gaussian and Chi-squared distributed variables, with the latter dominating at the zero lower bound. As a result, forecast distributions display a high degree of positive skewness when interest rates are close to the bound. When interest rates are further from the zero lower bound, as in 1992Q4, the forecast distribution is more symmetric.

To evaluate the historical importance of the zero lower bound, Figure 9 shows the proportion of the ATSM forecast distributions for the one-quarter rate which imply negative rates at different horizons (these are computed from 10,000 draws from the posterior parameter distributions). At short forecast horizons - for example, one and four quarters ahead - these proportions are generally below 5% until the mid-2000s, when they increase to about 20%. Towards the end of the sample these probabilities spike to over 30%, as short rates approach the lower bound. At

Figure 7: Model implied probability distributions for the US one-quarter interest rate from the QTSM

The probability distributions are computed using 10,000 simulations of the model parameters and factor dynamics. The red lines show the median of the distributions. Progressively lighter shading covers ten-percentile regions of the distributions with the total shaded area covering percentiles 10-90.

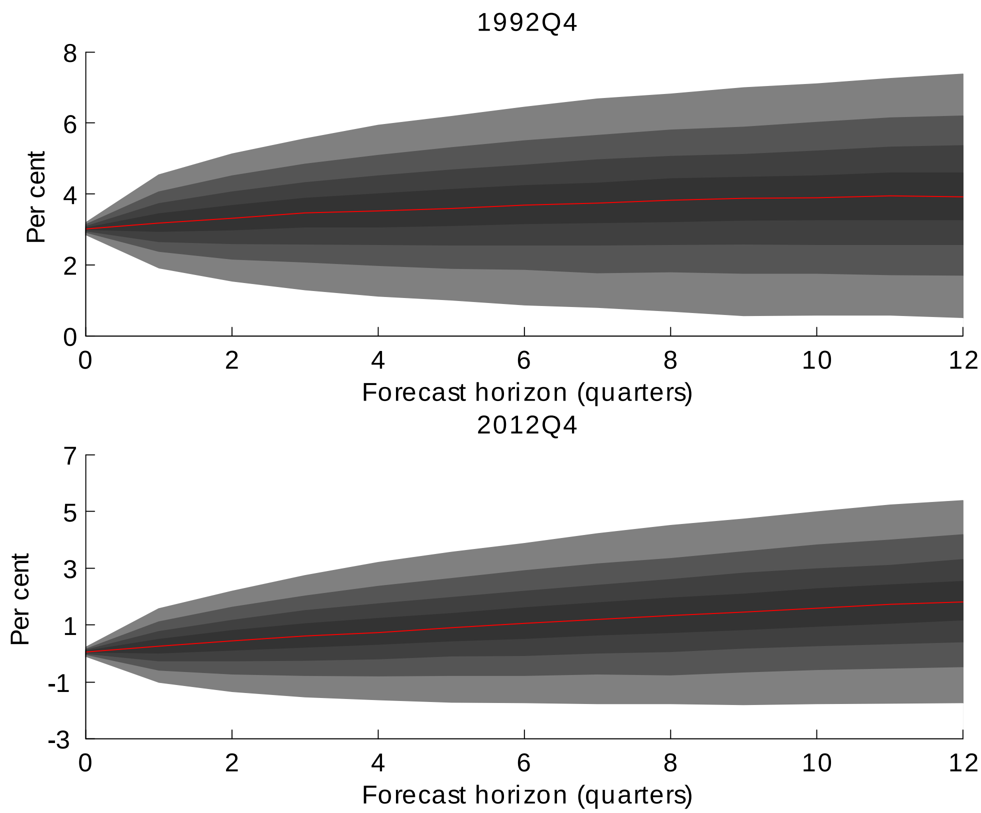


a forty-quarter forecast horizon, the probability of negative rates is around 10% throughout the sample, although this also drifts up towards the end of the sample. Hence, the constraint implied by the zero lower bound is not just binding in relation to the recent financial crisis but has actually been historically relevant.

As illustrated above, the quadratic model is able to account for the zero lower bound due to the non-linear terms and in effect truncates forecast distributions at α to ensure positive interest rates. This implies that forecast distributions in this model are skewed to the right. Another way to evaluate the importance of the non-linear terms in accounting for the lower bound is therefore to study the skewness in forecast distributions as done in Figure 10 (these are computed from

Figure 8: Model implied probability distributions for the US one-quarter interest rate from the ATSM

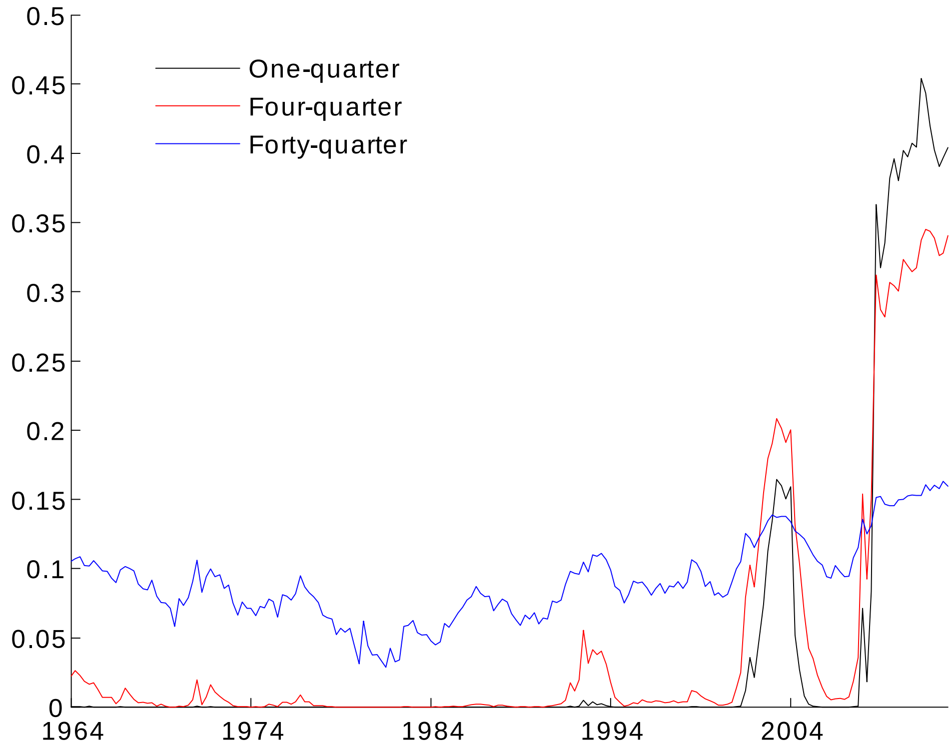
The probability distributions are computed using 10,000 simulations of the model parameters and factor dynamics. The red lines show the median of the distributions. Progressively lighter shading covers ten-percentile regions of the distributions with the total shaded area covering percentiles 10-90.



10,000 draws from the posterior parameter distributions). We see that values of skewness increase as the one-quarter rate gets close to the zero lower bound. This appears around 1993, after 2001, and in relation to the recent financial crisis.

To conclude, we find that the zero lower bound for bond yields is an important constraint to account for and that an ATSM may easily forecast negative interest rates. Such predictions appear during the recent financial crisis in the US but also prior to this crisis. On the other hand, a quadratic model perfectly accounts for the zero lower bound and therefore ensures sensible forecast distributions for bond yields.

Figure 9: Probability of forecasting negative 1-quarter interest rates in the ATSM
Each line relates to a different forecast horizon for the 1-quarter interest rate.

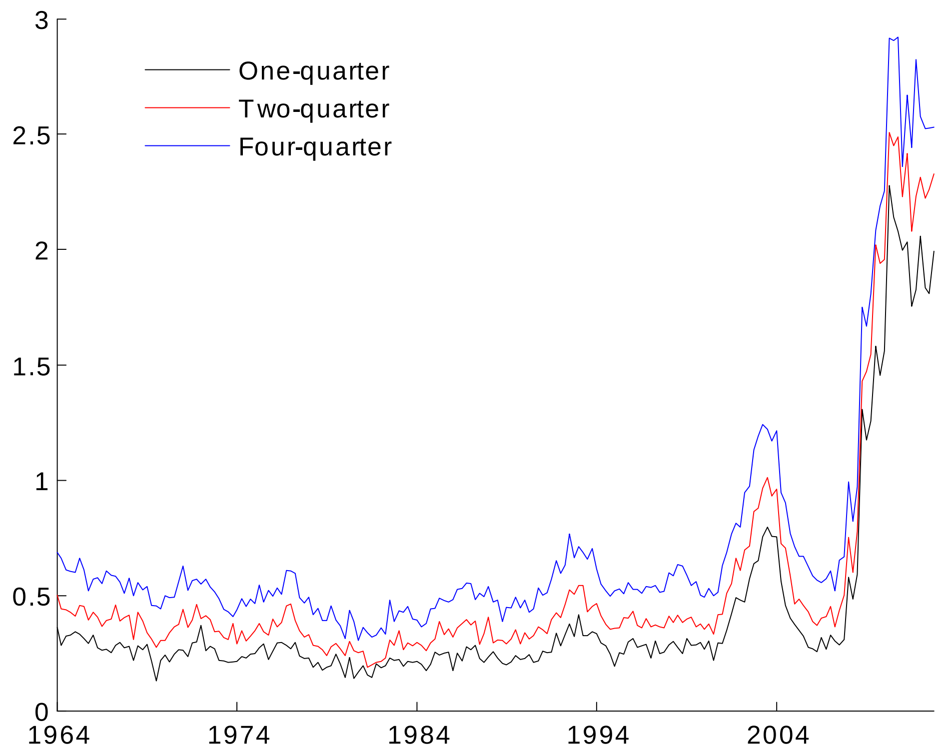


5 Conclusion

This paper shows how to use particle filtering and MCMC to estimate quadratic term structure models by likelihood inference. The procedure is made numerical feasible by using the optimised CDPF, which relies on the CDKF and a small optimisation step. We show that this adaptive particle filter provides a fast and reliable estimate of the log-likelihood function for QTSMs, in contrast to the standard particle filter. The suggested procedure is illustrated by estimating the most flexible QTSM on quarterly data for the US. We find that the QTSM provides a much better statistical description of the data than a Gaussian ATSM. Bayes factors and posterior Bayes factors are clearly in favour of the quadratic model. We also find that a Bayesian perspective is useful when accounting for the zero lower bound because this constraint may easily imply parameters at the boundary of their domain, thus rendering standard classical inference problematic. The quadratic model displays

Figure 10: Skewness of forecast distributions for the 1-quarter interest rates in the QTSM

Each line relates to a different forecast horizon for the 1-quarter interest rate. Skewness is computed based on 10,000 draws from the posterior parameter distributions, filtered states and state disturbances.



strong non-linearities and ensures that all forecast distributions imply positive interest rates. This property is in contrast to the Gaussian ATSM where a large proportion of the forecast distributions implies negative interest rates. Accordingly, the lower bound is an important constraint to account for in the US, and this constraint has been active before the recent financial crisis.

Although this paper has focused on QTSMs, the suggested likelihood procedure applies more generally to any non-linear and potentially non-Gaussian dynamic term structure model. Other applications include a quadratic model with non-linearities in the market price of risk to further improve the performance of the model. In relation to the lower bound, another interesting application would be to estimate a model with a shadow interest rate as suggested by Black (1995).

We leave these and other applications for future work.

Appendix 1: Details on the implementation of the PMMH algorithm

Delayed Rejection Adaptive Metropolis algorithm

We use the Delayed Rejection Adaptive Metropolis (DRAM) algorithm of Haario et al. (2006) to sample from the posterior distribution for the model parameters. At the i^{th} step in the chain we first draw a proposed parameter vector $\boldsymbol{\theta}'$ according to:

$$\boldsymbol{\theta}' = \boldsymbol{\theta}^{(i-1)} + \nu \mathbf{S}_{\theta}^{(i)} \boldsymbol{\xi}' \tag{10}$$

where $\boldsymbol{\theta}^{(i-1)}$ is the $i-1^{th}$ member of the chain, $\boldsymbol{\xi}' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\nu > 0$ is a constant and $\mathbf{S}_{\theta}^{(i)}$ is the Cholesky factor of the proposal covariance matrix (see below for further details). The proposal is accepted with probability $\min\{a_1, 1\}$, where (with a flat prior) a_1 is the ratio of the likelihood functions, i.e.

$$a_1 = \frac{p\left(\mathbf{y}_{1:T}|\boldsymbol{\theta}'\right)}{p\left(\mathbf{y}_{1:T}|\boldsymbol{\theta}^{(i-1)}\right)} \tag{11}$$

If the point is accepted, then $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}'$. If it is rejected, we draw a second proposed parameter vector $\boldsymbol{\theta}''$ from a scaled-down proposal distribution:

$$\boldsymbol{\theta}'' = \boldsymbol{\theta}^{(i-1)} + \delta \mathbf{S}_{\theta}^{(i)} \boldsymbol{\xi}''$$

where $\boldsymbol{\xi}'' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $0 < \delta < \nu$ is a constant. This proposal is accepted with probability $\min\{a_2, 1\}$, where a_2 is modified to account for the rejection at the first step, as in Haario et al. (2006).

For the first $b_s = 10,000$ observations in the chain $\mathbf{S}_{\theta}^{(i)}$ is fixed. After b_s observations, we follow Haario et al. (2001) and update the value of $\mathbf{S}_{\theta}^{(i)}$ based on all previous draws in the chain. That is we compute the mean of all values in the chain so far:

$$\overline{\boldsymbol{\theta}}^{(1:i)} = \begin{cases} \widehat{\boldsymbol{\theta}}^{CDKF} & \text{for } i = 0 \\ \frac{1}{i+1}\left[i\overline{\boldsymbol{\theta}}^{(1:i-1)} + \boldsymbol{\theta}^{(i)}\right] & \text{for } i \geq 1 \end{cases}, \tag{12}$$

and replace $\mathbf{S}_\theta^{(0)}$ in 10 by $\mathbf{S}_\theta^{(i-1)}$, where $\mathbf{R}_\theta^{(i-1)} = \mathbf{S}_\theta^{(i-1)}\mathbf{S}_\theta^{(i-1)\prime}$ and:

$$\mathbf{R}^{(i+1)} = \left\{ \begin{array}{lr} \frac{1}{b_s+1}\left[\mathbf{R}_\theta^{CDKF} + \sum_{i=1}^{b_s}\left(\boldsymbol{\theta}^{(i)} - \overline{\boldsymbol{\theta}}^{(1:b_\theta)}\right)\left(\boldsymbol{\theta}^{(i)} - \overline{\boldsymbol{\theta}}^{(1:b_\theta)}\right)'\right] & \text{if } i = b_s \\ \frac{i-1}{i}\mathbf{R}^{(i)} + \frac{s_d}{i}\left[i\overline{\boldsymbol{\theta}}^{(1:i-1)}\overline{\boldsymbol{\theta}}^{(1:i-1)\prime} - (i+1)\,\overline{\boldsymbol{\theta}}^{(1:i)}\overline{\boldsymbol{\theta}}^{(1:i)\prime} + \boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(i)\prime} + \epsilon\mathbf{I}\right] & \text{if } i > b_s \end{array} \right\}. \tag{13}$$

Here, $s_d = 2.4^2/N_\theta$ and $\epsilon > 0$. After constructing M chains of length N_θ, we then discard 'burn-in' periods of length $b_\theta < N_\theta$ from each chain and combine the remaining portions of the chains.

The sampling from the posterior distribution is initialised as follows. We first estimate the parameters ($\widehat{\boldsymbol{\theta}}^{CDKF}$) that maximise a quasi-likelihood function constructed using the Central Difference Kalman Filter (CDKF) of Norgaard et al. (2000). Here we apply the evolutionary optimisation routine CMA-ES which can deal with arbitrary non-linear constraints and multimodal objective functions (see Hansen et al. (2003) and Hansen and Kern (2004)). A Gaussian approximation around this point is constructed using a Hessian-based covariance estimator ($\mathbf{R}_\theta^{CDKF}$). A chain initialized by $\boldsymbol{\theta}^{(0)} = \widehat{\boldsymbol{\theta}}^{CDKF}$ and $\mathbf{R}_\theta^{CDKF} = \mathbf{S}_\theta^{(0)}\mathbf{S}_\theta^{(0)\prime}$ converges quite slowly to the posterior distribution. In a preliminary step, we therefore construct a chain of $N_\theta = 1,000,000$ replacing the likelihood function with the quasi-likelihood constructed using the CDKF - which is much quicker to evaluate that the CDPF - and discard the first $b_\theta = 500,000$ observations as burn-in. We then initialise the final chain using the mean and covariance of parameters from this preliminary quasi-MCMC chain.

The optimised Central Difference Particle Filter

We use the notation $[\mathbf{x}_t, \mathbf{S}_\mathbf{x}(t)] = CDKF(\mathbf{S}_\mathbf{x}(t-1), \mathbf{x}_{t-1}, \mathbf{y}_t)$ to denote one iteration in the CDKF from time point $t-1$ to time point t based on $\mathbf{S}_\mathbf{x}(t-1)$, $\mathbf{x}_{t-1}$, and $\mathbf{y}_t$.

- Initialisation: $t = 0$

  For $i = 1, ..., N$ draw particles $\widehat{\mathbf{x}}_0^{(i)}$ from $p(\mathbf{x}_0)$ and let $w_0^{(i)} = \frac{1}{N}$ for all i. The posterior state estimate: $\widehat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{x}}_t^{(i)}$

- For $t > 0$

  1. Importance sampling step

     – $\left[\widehat{\mathbf{x}}_t^{CDKF}, \widehat{\mathbf{S}}_\mathbf{x}^{CDKF}(t)\right] = CDKF\left(\widehat{\mathbf{S}}_\mathbf{x}^{CDKF}(t-1), \widehat{\mathbf{x}}_{t-1}, \mathbf{y}_t\right)$

– Determine the value of γ_t by numerical optimisation

– Draw particles $\mathbf{x}_t^{(i)}$ from $\mathcal{N}\left(\mathbf{x}_t^{(i)}\middle|\widehat{\mathbf{x}}_t^{CDKF},\gamma_t^2\widehat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)\left(\widehat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)\right)'\right)$ for $i=1,...,N$

– Evaluate the importance weights:
$w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t\middle|\mathbf{x}_t^{(i)};\boldsymbol{\theta}\right)p\left(\mathbf{x}_t^{(i)}\middle|\widehat{\mathbf{x}}_{t-1}^{(i)};\boldsymbol{\theta}\right)}{\mathcal{N}\left(\mathbf{x}_t^{(i)}\middle|\widehat{\mathbf{x}}_t^{CDKF},\gamma_t^2\widehat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)\left(\widehat{\mathbf{S}}_{\mathbf{x}}^{CDKF}(t)\right)'\right)}$ for $i=1,...,N$

– The contribution to the log-likelihood function: $L_t = L_{t-1} + \log(\sum_{i=1}^N w_t^{(i)})$

– For $i=1,...,N$ normalise the importance weights $\tilde{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^N w_t^{(i)}$

2. Resampling step:

– Resample with replacement from $\left\{\mathbf{x}_{0:t}^{(i)}\right\}_{i=1}^N$ with probabilities $\left\{\tilde{w}_t^{(i)}\right\}_{i=1}^N$ to obtain a samples of size N approximately distributed according to $p\left(\mathbf{x}_{0:t}\,|\mathbf{y}_{1:t}\,;\boldsymbol{\theta}\right)$. This new sample is denoted by $\left\{\widehat{\mathbf{x}}_{0:t}^{(i)}\right\}_{i=1}^N$

– In $\left\{\widehat{\mathbf{x}}_t^{(i)}\right\}_{i=1}^N$ we have $w_t^{(i)} = \frac{1}{N}$ for all $i=1,...,N$

3. State estimates

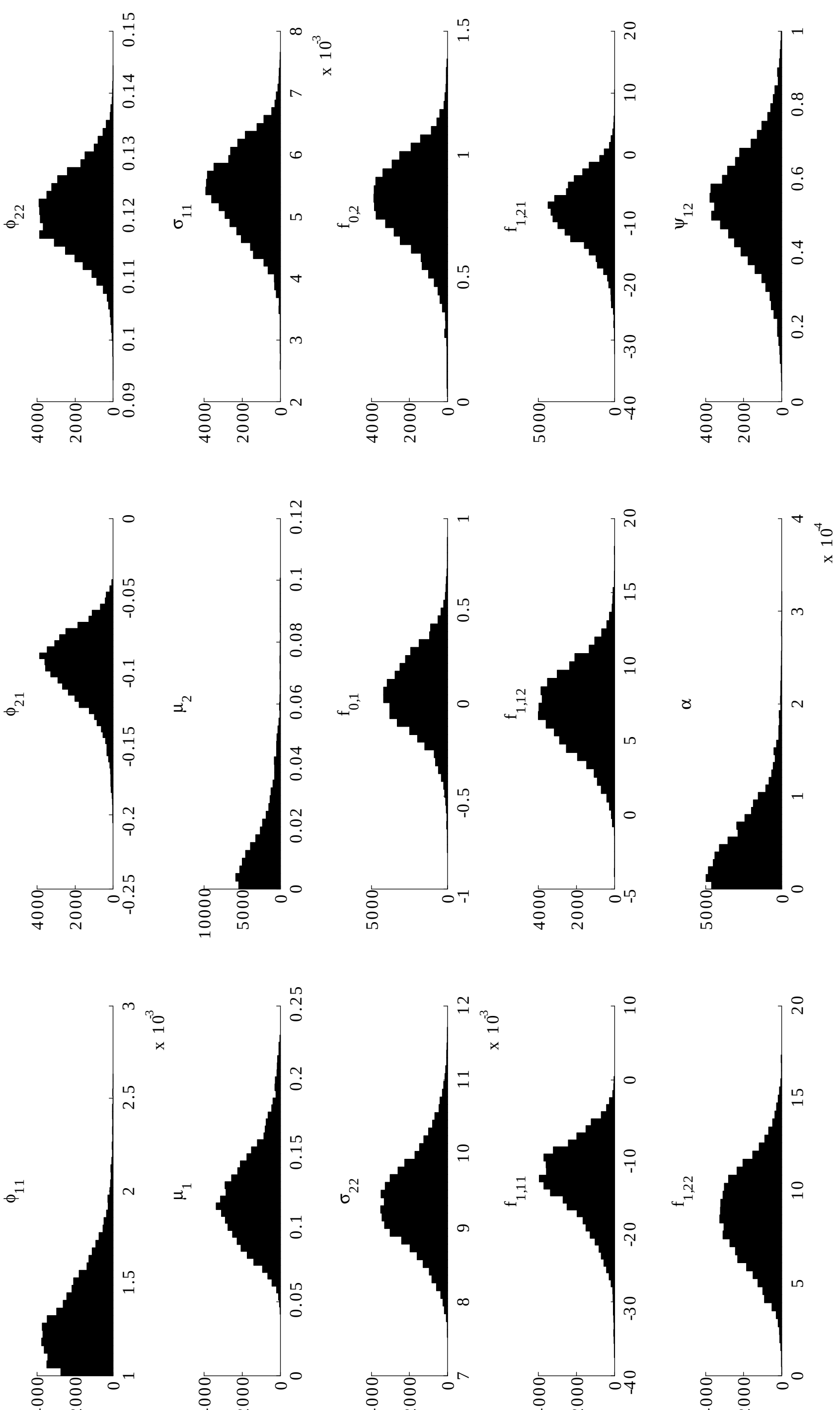
– The posterior state estimate: $\widehat{\mathbf{x}}_t = \frac{1}{N}\sum_{i=1}^N \widehat{\mathbf{x}}_t^{(i)}$

Appendix 2: Parameter estimates

Table 4: Posterior and QML distributions for the QTSM

Parameter	Posterior distribution				QML distribution			
	5^{th} percentile	Mean	95^{th} percentile	Standard deviation	5^{th} percentile	QML estimate	95^{th} percentile	Standard deviation
ϕ_{11}	0.0010	0.0013	0.0018	0.0002	-0.0018	0.0013	0.0044	0.0019
ϕ_{21}	-0.1444	-0.1008	-0.0649	0.0241	-0.1128	-0.0778	-0.0424	0.0212
ϕ_{22}	0.1103	0.1206	0.1311	0.0063	0.1094	0.1251	0.1410	0.0095
μ_1	0.0753	0.1208	0.1770	0.0309	-0.3231	0.1292	0.5804	0.0295
μ_2	0.0012	0.0157	0.0433	0.0135	-0.2761	1.1470×10^{-9}	0.2741	0.0049
σ_1	0.0043	0.0054	0.0065	0.0007	0.0052	0.0062	0.0072	0.0004
σ_2	0.0084	0.0094	0.0104	0.0006	0.0074	0.0087	0.0099	0.0006
$f_{0,1}$	-0.2924	0.0562	0.4002	0.2090	-0.4023	-0.0167	0.3696	0.2003
$f_{0,2}$	0.5057	0.8078	1.0759	0.1765	0.6617	1.1269	1.5934	0.1676
$f_{1,11}$	-23.5141	-13.8368	-5.9447	5.3212	-17.4714	-10.1801	-2.8966	3.4327
$f_{1,12}$	2.4859	7.1827	11.6262	2.7553	1.1863	5.7998	10.4484	0.7040
$f_{1,21}$	-18.3465	-9.3246	-1.2150	5.2499	-10.9993	-4.0737	2.8741	2.7397
$f_{1,22}$	4.5233	8.7389	12.8673	2.5246	4.8903	10.4723	16.0774	0.6294
α	4.7981×10^{-6}	5.2993×10^{-5}	1.3478×10^{-4}	4.2075×10^{-5}	-2.7535×10^{-4}	1.8722×10^{-13}	2.7164×10^{-4}	0.0001
ψ_{21}	0.3005	0.5357	0.7658	0.1400	0.1498	0.4552	0.7621	0.0647
s_w	0.2146	0.2236	0.2331	0.0056	0.2017	0.2224	0.2433	0.0082

Figure 11: Posterior distributions for the QTSM



References

Ahn, D., R. F. Dittmar, and A. R. Gallant (2002). Quadratic term structure models: theory and evidence. *The Review of Financial Studies 15*, 243–288.

Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society 53*, 111–142.

Andreasen, M. M. (2010). Non-linear DSGE models and the Optimised Particle Filter. *CREATES Research Paper 2010-5*.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B 72*, 269–342.

Black, F. (1995). Interest rates as options. *The Journal of Finance 50*, 1371–1376.

Carrasco, M. and J.-P. Florens (2002). Simulation-based method of moments and efficiency. *Journal of Business and Economic Statistics 20*, 482–492.

Chib, S. and B. Ergashev (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association 104*, 1324–1337.

Dai, Q. and K. J. Singleton (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics 63*, 415–441.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer.

Doucet, A., S. Godsill, and C. Andrieu (2000). On Sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing 10*, 197–208.

Duffee, G. R. and R. H. Stanton (2012). Estimation of dynamic term structure models. *Working paper*.

Flury, T. and N. Shephard (2009). Learning and filtering via simulations: smoothly jittered particle filters. *University of Oxford, Department of Economics Discussion Paper Series, number 469*.

Flury, T. and N. Shephard (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory 27*, 933–956.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–681.

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*. Oxford University Press.

Gilks, W. and C. Berzuini (2001). Following a moving target: Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society B 63*, 127–146.

Gordon, N., D. S. Salmon, and A. Smith (1993). A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Institute of Electrical Engineers - Radar, Sonar and Navigation 140*, 107–113.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing 16*, 339–354.

Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive Metropolis algorithm. *Bernoulli 7*, 223–242.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for LDV models using simulation. *Handbook of Econometrics 4*, 2384–2438.

Hansen, N. and S. Kern (2004). Evaluating the CMA Evolution Strategy on multimodal test functions. *Eighth International Conference on Parallel Problem Solving from Nature PPSN VIII Proceedings*, 282–291.

Hansen, N., S. D. Muller, and P. Koumoutsakos (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation 11*, 1–18.

Kantas, N., A. Doucet, S. S. Singh, and J. M. Maciejowski (2009). An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. *15th IFAC Symposium on System Identification Proceedings*.

Kim, D. H. (2007). Spanned stochastic volatility in bond markets: a re-examination of the relative pricing between bonds and bond options. *BIS Working Paper 239*.

Kim, D. H. and K. J. Singleton (2011). Term structure models and the zero bound: an empirical investigation of Japanese yields. *Working paper*.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*, 119–147.

Li, H. and F. Zhao (2006). Unspanned stochastic volatility: evidence from hedging interest rate derivatives. *Journal of Finance 2006*, 341–377.

Lieppold, M. and L. Wu (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis 37*, 271–295.

Lieppold, M. and L. Wu (2003). Design and estimation of quadratic term structure models. *European Finance Review 7*, 47–73.

Lieppold, M. and L. Wu (2007). Design and estimation of multi-currency quadratic models. *Review of Finance 11*, 167–207.

Norgaard, M., N. K. Poulsen, and O. Ravn (2000). Advances in derivative-free state estimation for non-linear systems. *Automatica 36*, 1627–1638.

Pitt, M. K. (2002). Smooth particle filters for likelihood evaluation and maximisation. *Working paper*.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association 94*, 590–599.

Realdon, M. (2006). Quadratic term structure models in discrete time. *Finance Research Letters 3*, 277–289.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*, 1385–1411.

Taulbjerg, J. (2002). Estimating quadratic term structure models by non-linear filtering. *University of Aarhus Centre for Analytical Finance Working Paper 123*.