Long-run priors for term structure models

Andrew Meldrum† †\hspace{1em} Matt Roberts-Sklar† †
Bank of England \hspace{1em} Bank of England

First version: 18 December 2015
This version: 22 June 2016

Abstract

Dynamic no-arbitrage term structure models are popular tools for decomposing bond yields into expectations of future short-term interest rates and term premia. But there is insufficient information in the time series of observed yields to estimate the dynamics of yields accurately or precisely. This can result in implausibly low estimates of long-term expected future short-term interest rates, considerable uncertainty around those estimates and instability in estimates of term premia across different samples. This paper proposes a tractable Bayesian approach for incorporating prior information about the unconditional means of yields, which we calibrate on the basis of a simple time-series model of nominal GDP. We apply it to UK data and find that with reasonable priors it results in more plausible estimates of the long-run average of yields, lower estimates of term premia in long-term bonds and substantially reduced uncertainty around decompositions and improved out-of-sample forecasting performance.

Keywords: affine term structure model, Gibbs sampler, shifting end-point.

JEL: C11, E43, G12.

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of England or members of its committees. The authors would like to thank Michiel De Pooter, Iryna Kaminska, Don Kim, Wolfgang Lemke, Canlin Li, Marcel Priebsch, Peter Spencer, Kostas Theodoridis, Thomas Werner, participants in seminars at the Bank of England and European Central Bank and an anonymous referee for the Bank of England Staff Working Paper series for helpful comments on this paper.
†Macro Financial Analysis Division, Bank of England, Threadneedle Street, London EC2R 8AH, UK. e-mail: andrew.meldrum@bankofengland.co.uk.
‡Macro Financial Analysis Division, Bank of England, Threadneedle Street, London EC2R 8AH, UK. e-mail: matt.roberts-sklar@bankofengland.co.uk.
1 Introduction

Dynamic no-arbitrage term structure models are popular tools for analyzing the joint dynamics of bond yields of different maturities. Policymakers routinely use these models to decompose long-term bond yields into expectations of future short-term interest rates and a term premium that reflects the additional expected return for investing in long-term bonds rather than rolling over a series of short-term bonds.\(^1\) But the uncertainty around the decompositions obtained using these models is substantial. Confidence intervals are wide and point estimates can be sensitive to modest variations in the sample period or in the specification of the model.

This paper provides a tractable method for incorporating prior information about the long-run average level of interest rates in a Bayesian setting, which substantially alleviates these problems and which results in substantial improvements in the out-of-sample forecasting performance of the models. Using UK data, we show that our approach results in more plausible term structure decompositions, reduces the estimated uncertainty around those decompositions substantially and results in greater sub-sample stability. This should have obvious appeal to policymakers and others concerned with the long-horizon properties of these models.

In maximally flexible no-arbitrage term structure models, the accuracy and precision of term premium estimates is primarily determined by the accuracy and precision with which we can estimate the time-series dynamics of the pricing factors driving yields. Bond yields, including short-term risk-free rates, are affine functions of a small number of pricing factors, which follow a first-order Gaussian Vector Autoregression (VAR). Term premia are defined as the difference between model-implied yields (which tend to be very close to observed yields) and the model-implied average expected short-term interest rate over the relevant horizon, which is given by a projection from the VAR. Unfortunately, however, the short samples of yields typically available, together with general declines in yields over those

\(^1\)For example, as the then Chairman of the Federal Reserve, Ben Bernanke, explained in a speech on long-term interest rates in March 2013: "It is useful to decompose longer-term yields into three components: one reflecting expected inflation over the term of the security; another capturing the expected path of short-term real, or inflation-adjusted, interest rates; and a residual component known as the term premium. Of course, none of these components is observed directly, but there are standard ways of estimating them."
periods, means that there is little sample information with which to estimate the dynamics of yields - a point made previously by a number of studies, including Kim and Orphanides (2012), Bauer et al. (2012) and Wright (2014).\textsuperscript{2}

While there is little information in the data with which to estimate the unconditional means of yields, it is nevertheless reasonable to suppose that we do have relevant prior information. Ignoring that information, and estimating the model with flat priors, implies that we attach a higher prior weight on a steady state value of the short rate that is less than (say) zero than (say) between zero and 10\% - which is clearly inconsistent with any reasonable prior beliefs. We show that incorporating prior information about the long-run average yield curve, using a framework based on that of Villani (2009), substantially alleviates these problems. Specifically, we rotate the pricing factors of the models such that they are equal to observed bond yields and specify priors on the unconditional means of those yields, calibrated using a very simple time-series model of nominal GDP. The unconditional means of yields can then be drawn directly within a Gibbs sampling procedure that is otherwise very similar to the approach for estimating no-arbitrage affine term structure models proposed by Bauer (2016).

A number of alternative approaches have been proposed previously to address the underlying problem of weak identification of the time-series dynamics in dynamic term structure models. One option is to incorporate additional information in the form of survey expectations of professional economists about future short-term policy interest rates (proposed by Kim and Orphanides (2012)). In the case of the UK, there are unfortunately no long-horizon surveys of Bank Rate expectations available and the time-series of short-horizon surveys has also fallen over time. Moreover, Malik and Meldrum (2016) show that incorporating short-term surveys for the UK can result in markedly inferior performance of affine term structure models against standard specification tests.

A second approach, taken by Cochrane and Piazzesi (2008) among others, is to impose zero restrictions on the price of risk, in order that estimates of the risk-neutral factor dynamics can inform the time-series dynamics. One option is to impose zero restrictions

\textsuperscript{2}The uncertainty around estimates of UK term premia is discussed by Malik and Meldrum (2016) and Guimarães (2016).
on any parameters that are not significantly different from zero in an initial unrestricted estimation. Bauer (2016) proposes a Bayesian approach for weighting models with different zero restrictions on the price of risk, in which the prior is specified to shrink towards more parsimonious models, and finds a substantial impact on estimated risk premia. In this paper, we work entirely with maximally flexible models, although in principle it would be possible to combine a long-run prior for the mean of yields with restrictions on the prices of risk.

A related problem with small samples is that OLS estimates of autoregressive models are biased. Bauer et al. (2012) use statistical techniques to correct for small-sample bias in a classical setting. That approach is focussed more on the persistence of the factors, rather than their average level. However, bias corrections are typically applied to demeaned data, so the intercept is effectively calibrated to match the sample mean.\(^3\) This may reduce the problem of underestimating the mean in some samples but in general the sample average may also be unlikely \emph{a priori}. Moreover, by calibrating the intercept in this way we are likely to \emph{understate} our true uncertainty about term premium estimates. Estimating the intercept but allowing for prior information to inform that estimate is likely to result in more reasonable estimates of the true uncertainty.\(^4\)

Given the findings of Bauer et al. (2012) about the importance of allowing for estimation bias, however, we also consider a variant of our model which adopts a Minnesota prior, under which the dynamics of yields of are shrunk towards independent random walks. As with bias corrections, this will tend to raise the estimated persistence of yields. Estimates of term premia are slightly lower at the beginning of the sample and higher towards the end of the sample. The reason is intuitive: if yields are estimated to be more persistent, they take longer to return to their average levels. For example, when yields were high in the early 1990s, the model with the Minnesota prior implies a higher expected path of future short rates and therefore a lower term premium. The Minnesota prior also has the effect of

\(^3\)Adrian et al. (2013) and Malik and Meldrum (2016) do not apply a small-sample bias correction but do nevertheless calibrate the intercept in the VAR so that the unconditional mean of the pricing factors matches the sample average.

\(^4\)Although we do not address the issue of classical small-sample bias, in a variant of our approach, we explore a prior that shrinks the persistence of the pricing factors towards a unit root, which has a similar qualitative effect to classical bias adjustment.
reducing the parameter uncertainty around estimates of term premia at times when yields are further from their average levels and leads to further reductions in out-of-sample forecast errors.

While our ‘long-run priors’ approach goes a long way to addressing the parameter uncertainty that plagues efforts to decompose bond yields into expectations of future short rates and term premia, it does not address the issue of model uncertainty. We consider the impact of two forms of model uncertainty in this paper. First, affine term structure models do not allow for a lower bound on nominal interest rates. We show how to apply our long-run prior in the lower-bound-consistent ‘shadow rate’ term structure model proposed by Black (1995). We show that allowing for the lower bound can result in posterior mean estimates of term premia that are lower than in affine models and that have narrower probability intervals. This contrasts with some previous studies (Kim and Priebsch (2013) for the US and Malik and Meldrum (2016) for the UK), which have found that long-maturity term premium estimates from shadow rate models are very similar to those from affine models. In studies that use frequentist estimators, term premium estimates are generally reported as point estimates at (e.g.) the maximum likelihood parameters, whereas here we report posterior probability intervals for term premia. In affine models a considerable proportion of the posterior probability interval for expectations of short-term interest rates is in the negative region, which suggests that affine models place a material posterior probability of term premia that are implausibly high.

Another form of uncertainty is that standard dynamic term structure models, with a constant unconditional means, may not be appropriate. Structural breaks in the mean level of yields - for example, related to changes in the institutional framework for monetary policy - may have resulted in time-varying infinite-horizon conditional expectations of yields - sometimes referred to as a ‘shifting end-point’ (Kozicki and Tinsley (2001)). We therefore explore the implications of dropping the assumption of a constant unconditional mean and replacing it with a shifting end-point. Our approach is similar to Van Dijk et al. (2014), in that we allow long-horizon survey forecasts of bond yields to inform estimates of the endpoint - but has two key differences (aside from using UK data and a Bayesian estimation approach). First, we incorporate the survey-based shifting end-point within a no-arbitrage
term structure model; and second, we allow the survey-based measure to be measured with error, such that they provide only a noisy signal about the end-point.\footnote{Kim and Orphanides (2012) allow long-horizon surveys to inform the dynamics of yields in a no-arbitrage model and allow for measurement error on surveys - but in a model with a constant end-point for bond yields.} Perhaps encouragingly, the effect of allowing for a shifting end-point on estimate of long-horizon term premia is broadly similar to the effect we get from imposing the long-run prior over a constant mean combined with a Minnesota prior: in practice, a process with a shifting end-point is difficult to distinguish from a stationary but extremely persistent process.

Section 2 of this paper describes our benchmark no-arbitrage affine term structure model. Section 3 describes the techniques we use to estimate it and the choice of priors and Section 4 reports results, including on how the choice of priors affects out-of-sample forecasts. Section 6 discusses how the results are affected if we impose a zero lower bound on nominal interest rates or allow for a shifting end-point in long-horizon expectations of yields. Section 7 concludes.

2 Model

2.1 Affine term structure model

This section sets out our (entirely standard) benchmark affine term structure model of nominal bond yields. Similar models have been applied to UK data previously, including by Joyce et al. (2010), Guimarães (2016) and Malik and Meldrum (2016).\footnote{As far as we are aware, ours is the first study to use Bayesian methods for the estimation of such a model using UK data.} A nominal $n$-period zero-coupon bond pays £1 at its maturity after $n$ periods. In the absence of arbitrage, the time-$t$ price ($P_t^{(n)}$) is equal to the expected discounted present value of the price at time $t+1$:

$$P_t^{(n)} = \mathbb{E}_t^Q \left[ \exp (-i_t) P_{t+1}^{(n-1)} \right], \quad (1)$$

where $i_t$ is the one-period nominal risk-free rate and expectations are formed with respect to the risk-neutral probability measure, denoted $\mathbb{Q}$. The short-term rate is an affine function
of a $K \times 1$ vector of pricing factors $x_t$:

$$i_t = \delta_0 + \delta'_1 x_t. \quad (2)$$

The factors follow a first-order Gaussian Vector Autoregression (VAR) under $Q$:

$$x_{t+1} = \mu^Q + \Phi^Q x_t + v^Q_{t+1}, \quad (3)$$
$$v^Q_t \sim i.i.d. N(0, \Sigma).$$

Given the above assumptions, nominal bond yields are affine functions of the factors:

$$y^{(n)}_t = \frac{1}{n} (a_n + b'_n x_t), \quad (4)$$

where $a_n$ and $b_n$ follow the standard recursive equations

$$a_n = a_{n-1} + b'_{n-1} \mu^Q + \frac{1}{2} b'_{n-1} \Sigma b_{n-1} - \delta_0 \quad (5)$$
$$b'_n = b'_n \Phi^Q - \delta'_1. \quad (6)$$

with the initial conditions $a_0 = 0$ and $b_0 = 0$.

As has been discussed widely elsewhere (e.g. Dai and Singleton (2000), Joslin et al. (2011) and Hamilton and Wu (2012)) the model is not identified without additional parameter restrictions. We adopt the normalization $\delta_1 = 1_{(K \times 1)}$, $\mu^Q = 0_{(K \times 1)}$ and $\Phi^Q = diag \{[\phi_1, \phi_2, ..., \phi_K]\}$, with $1 > \phi_1 > \phi_2 > ... > \phi_K > 0$.

Following Duffee (2002), we assume that the market prices of risk are affine in the pricing factors, which implies that the pricing factors also follow a first-order Gaussian VAR under the real-world probability measure:

$$x_{t+1} = \mu + \Phi x_t + v_{t+1} \quad (7)$$
$$v_t \sim i.i.d. N(0, \Sigma).$$

As is standard (e.g. Dai and Singleton (2002)), we define the term premium component
of an $n$-period yield as the difference between the model-implied yield and the average expected short-term rate over the lifetime of the bond:

$$TP_t^{(n)} = y_t^{(n)} - \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}_t [t_{t+i}].$$

(8)

3 Estimation

3.1 Data

The UK nominal government zero-coupon yields we use to estimate the model have maturities of 3, 12, 24, 36, 48, 60, 84 and 120 months. All except the three-month rate are estimated using the cubic spline technique of Anderson and Sleath (2001).\(^7\) As this dataset does not consistently include nominal maturities shorter than one year, we augment it by using a three-month Treasury bill yield.\(^8\) Our sample period starts in October 1992, when the UK first introduced an inflation targeting framework for monetary policy, and runs until December 2014. The choice of sample start date is the same as or similar to those chosen by most other studies using UK data (e.g. Joyce et al. (2010) also use a sample starting in October 1992, while D’Amico et al. (2014) use a sample starting in January 1993).

Figure 1 plots yields of selected maturities. UK nominal yields generally fell through this period, in common with those in other advanced economies.\(^9\) As discussed above, the lack of ‘mean reversions’ in the sample means that we cannot reliably estimate the dynamics of yields (Kim and Orphanides (2012)). As well as leading to substantial uncertainty around estimates of term premia, this can also result in implausibly low estimates of the unconditional mean of yields, which in turn results in estimates of long-maturity term premia that are too high. To illustrate why this is the case, the solid line on Figure 2 plots the UK ten-year yield, while the dotted line overlays a projection starting in October 1992 from a univariate first-order autoregressive model estimated using the full sample. As pointed out by Sims (2000), OLS estimates of autoregressive models using small samples

\(^8\) Available from: http://www.bankofengland.co.uk/statistics/Pages/default.aspx.
\(^9\) For example, the majority of studies using US data also tend to use a sample that starts in the 1980s or early 1990s. US nominal bond yields have also fallen over much of this period.
have a tendency to exaggerate the component of the sample variation that is deterministic conditional on the initial observation. In broad terms, the autoregressive model interprets most of the fall in the 10-year yield over the sample as an initial observation a long way above the unconditional mean and a subsequent deterministic reversion, lasting around 20 years, towards that mean. The estimated unconditional mean - a little over 2% (shown by the thick black line) - is below almost all of the sample data and the initial point is well outside the 90% confidence interval for the unconditional mean.\footnote{We bootstrap the confidence interval for the unconditional mean at the OLS parameter estimates, using a 10,000 draws in the bootstrap.}

3.2 Factor structure and measurement error

As is standard in the dynamic term structure literature, our benchmark model has three pricing factors. We assume that three yields (collected in the vector $x_t^* = \left[ y_t^{(3)}, y_t^{(60)}, y_t^{(120)} \right]'$) are observed without error and the remaining 5 yields ($y_t^w = \left[ y_t^{(12)}, y_t^{(24)}, y_t^{(36)}, y_t^{(48)}, y_t^{(84)} \right]'$) are observed with errors $w_t$.\footnote{The assumption that three yields can be measured without error is purely to simplify the estimation and is not important for the way we implement the long-run prior.} This means that the measurement equations of the model can be written as

\begin{align*}
x_t^* &= A_1 + B_1 x_t \\
y_t^w &= A_2 + B_2 x_t + w_t \\
w_t &\sim i.i.d. \mathcal{N}(0, R_w)
\end{align*}

where the definitions of $A_1$, $B_1$, $A_2$ and $B_2$ follow from (4). Conditional on values of $\delta_0$, $\delta$, $\mu^Q$, $\Phi^Q$ and $\Sigma$, we can use the procedure of Chen and Scott (1993) to invert (9) and recover the pricing factors, i.e. $x_t = B_1^{-1}(x_t^* - A_1)$. 

\begin{figure*}[ht] <Figure 1 about here> 
\end{figure*}

\begin{figure*}[ht] <Figure 2 about here> 
\end{figure*}
3.3 Gibbs sampler

We estimate the model using Bayesian methods, splitting the parameters into six blocks and using a Gibbs sampler to draw from the conditional posteriors of each in turn: (i) the parameters governing the dynamics of the factors under the time series measure (Φ); (ii) the intercepts under the time-series dynamics (μ); (iii) the parameters governing the dynamics of the factors under Q (ΦQ); (iv) the intercept in the short rate equation (δ0); (v) the factor shock covariance matrix Σ; and (vi) the covariance matrix of measurement errors Rw. The approach for drawing blocks (iii)-(vi) is very similar to that proposed by Bauer (2016).12 The most substantial methodological innovation in this paper is therefore the process for drawing the parameters of the time-series dynamics ((i) and (ii)), which are addressed in detail immediately below.13

We first estimate the parameters by maximum likelihood to obtain initial values for the chain, as explained in Appendix A. We then draw 10,000 times from the Gibbs sampler, discarding the first 5,000 draws as burn-in.

3.3.1 Time series dynamics (μ and Φ)

A typical approach to specifying a prior for a Bayesian VAR such as (7) would be to assume that (conditional on Σ) μ and Φ are jointly and independently Normally distributed under the prior. In our case, however, it is not obvious how to specify a meaningful prior over the factors, particularly given the fact that they depend on the precise normalization (a point made by Kim (2009)). But, as discussed above, it is reasonable to believe that we have prior information about the long-run mean of bond yields. To implement such a long-run prior in an affine term structure model, we assume there are K independent linear combinations of the N bond yields about which we have some prior information, which we can write as

\[ x^*_t = W'y_t = W' (A + Bx_t), \]  

12 Other studies that have estimated dynamic term structure models using Bayesian methods include Chib and Ergashev (2009), Ang et al. (2011) and Andreasen and Meldrum (2013).

13 Whereas Bauer draws the parameters of the market prices of risk which relate the time-series and risk-neutral factor dynamics, we instead draw the time-series dynamics directly.
where $W$ is a $K \times N$ matrix of full rank; $y_t$ is a vector of all model-implied yields; and the definitions of $A$ and $B$ follow from (5) and (6). We can write the reduced-form time-series dynamics of these yields as

$$x^*_{t+1} = \mu^* + \Phi^* x^*_t + v^*_{t+1},$$

$$v^*_t \sim i.i.d. N(0, \Sigma^*).$$

Using (7), (11) and (12), note that we can recover the structural parameters $\mu$, $\Phi$ and $\Sigma$ in terms of $\mu^*$, $\Phi^*$ and $\Sigma^*$:

$$\mu = (W'B)^{-1} (\mu^* - W'A + \Phi^* W'A)$$

$$\Phi = (W'B)^{-1} \Phi^* W'B$$

$$\Sigma = (W'B)^{-1} \Sigma^* (B'W)^{-1}.$$  

We can also re-write (12) in terms of deviations from the unconditional mean of $x^*_t$, $\gamma = E[x^*_t] = (I - \Phi^*)^{-1} \mu^*$:

$$\tilde{x}^*_{t+1} = x^*_{t+1} - \gamma = \Phi^* \tilde{x}^*_t + v^*_{t+1}. $$

Stacking this equation across $t$ gives

$$\tilde{X}_+ = \tilde{X}_- \Phi^* + V,$$

where $\tilde{X}_+ = [\tilde{x}^*_2, \tilde{x}^*_3, \ldots, \tilde{x}^*_T]'$ and $\tilde{X}_- = [\tilde{x}^*_1, \tilde{x}^*_2, \ldots, \tilde{x}^*_T_{-1}]'$. If we assume an independent Normal prior for $\phi^* = vec(\Phi^*)$:

$$\phi^*, \Sigma^*, \gamma \sim N(\phi, \Sigma^*)$$

it is straightforward to generate a draw from the posterior:

$$\phi^*, \Sigma^*, X, \gamma \sim N(\tilde{\phi}, \tilde{V}_\phi).$$
where

\[
\mathbf{V}_\phi = \left( \mathbf{V}_\phi^{-1} + \Sigma^* \otimes \mathbf{X}_- \mathbf{X}_- \right)^{-1},
\]

\[
\mathbf{\bar{\phi}} = \mathbf{V}_\phi \left( \mathbf{V}_\phi^{-1} \mathbf{\phi} + \Sigma^* \otimes \mathbf{I}_K \right) vec \left( \mathbf{X}_- \mathbf{X}_+ \right).
\]

The baseline results reported in Section 4 assume a diffuse prior (i.e., \(\mathbf{V}_\phi^{-1} = 0\)), although we also explore the impact of assuming a Minnesota prior for \(\mathbf{\Phi}^*\). In all our models we also impose a prior that yields are stationary by rejecting any draws that imply eigenvalues that are outside the unit circle.

Turning to the intercept, we can re-write (12), substituting \((\mathbf{I} - \mathbf{\Phi}^*) \gamma\) for \(\mu^*\):

\[
(I - \Phi^*)^{-1} (x_{t+1}^* - \Phi^* x_t^*) = \gamma + (I - \Phi^*)^{-1} v_{t+1}^*.
\] (20)

Stacking across \(t\), we can re-write this as

\[
\mathbf{\Xi} = (\mathbf{X}_+ - \mathbf{X}_- \Phi^*) (I - \Phi^*)^{-1} = \nu_T \gamma' + \mathbf{V} (I - \Phi^*)^{-1} \gamma',
\] (21)

where \(\mathbf{X}_+ = [x_2^*, x_3^*, ..., x_T^*]'\), \(\mathbf{X}_- = [x_1^*, x_2^*, ..., x_{T-1}^*]'\) and \(\nu_T\) is a \(T \times 1\) vector of ones. As proposed by Villani (2009), we assume an independent Normal prior for \(\gamma\):

\[
\gamma | \Sigma^*, \Phi^* \sim \mathcal{N} \left( \gamma, \mathbf{V}_\gamma \right).
\] (22)

It is straightforward to draw from the posterior, which is given by:

\[
\gamma | \Sigma^*, \Phi^*, \mathbf{X}_+ \sim \mathcal{N} \left( \bar{\gamma}, \bar{\mathbf{V}}_\gamma \right),
\] (23)

where

\[
\bar{\mathbf{V}}_\gamma = \left( \mathbf{V}_\gamma^{-1} + T \left( (I - \Phi^*)^{-1} \Sigma^* (I - \Phi^*)^{-1} \right)^{-1} \right)^{-1},
\]

\[
\bar{\gamma} = \mathbf{V}_\gamma \left( \mathbf{V}_\gamma^{-1} \gamma + \left( (I - \Phi^*)^{-1} \Sigma^* (I - \Phi^*)^{-1} \right) vec (\nu_T \mathbf{\Xi}) \right).
\]
Given that the information required to identify the unconditional mean of yields is simply not in the data, the choice of prior moments $\gamma$ and $\mathbf{V}_\gamma$ is clearly key. We base our prior on perhaps the simplest possible consumption CAPM, calibrated using only macroeconomic data. In a standard setting in which households have time-separable log utility over consumption, the pricing equation for a long-term nominal bond is

$$P_t^{(n)} = \mathbb{E}_t \left[ \prod_{i=1}^{n} \beta \frac{C_{t+i} - 1}{C_t} \frac{Q_{t+i} - 1}{Q_t} \right], \quad (24)$$

where $\beta$ is the rate of time preference; $C_t$ is consumption at time $t$; and $Q_t$ is the consumer price level at time $t$. We assume that log nominal consumption growth $g_{t+1} = \frac{C_{t+1} - 1}{C_t}$ follows an AR(1) process:

$$g_{t+1} = \mu_g + \phi_g g_t + \varepsilon_{g,t}, \quad (25)$$

$$\varepsilon_{g,t} \sim \mathcal{N}(0, \sigma_g^2).$$

For simplicity we also assume that nominal consumption growth is the same as log nominal GDP growth and estimate $(25)$ using UK quarterly GDP growth for 1992Q4-2014Q4. We fix $\sigma_g^2$ equal to the OLS estimate $\hat{\sigma}_g^2$ and estimate $\mu_g$ and $\phi_g$ with flat priors. We do not estimate the distribution for $\beta$ but assume that it comes from an independent beta distribution with mean 0.9982 (the posterior mean from Doh (2013) and a standard deviation of 0.002.

We compute bond prices numerically. We sample from the posterior distribution for $\mu_g$ and $\phi_g$ 1,000 times; for each parameter draw we draw a value of $\beta$ and simulate 100 samples (each of 40 quarters) of nominal GDP growth to approximate the expectation in $(24)$. We then compute the mean and variance of bond prices across the 1,000 parameter draws, which we use as our prior mean and variance for our term structure model. The prior moments given by this procedure are:

$$\gamma = \begin{bmatrix} \frac{5.17}{1200} & \frac{5.17}{1200} & \frac{5.17}{1200} \end{bmatrix},$$

$$\mathbf{V}_\gamma = \text{diag} \left\{ \left[ \frac{0.82}{1200^2}, \frac{0.20}{1200^2}, \frac{0.20}{1200^2} \right] \right\}.$$
3.3.2 $\mathcal{Q}$ parameters ($\delta_0$ and $\Phi^\mathcal{Q}$)

We draw the parameters governing the $\mathcal{Q}$ dynamics of the factors ($\Phi^\mathcal{Q}$) and the short-term interest rate ($\delta_0$) using Metropolis-within-Gibbs steps, very similar to those proposed by Bauer (2016). We parameterize $\Phi^\mathcal{Q}$ as $\Phi^\mathcal{Q} = I + \text{diag}\{\phi^\mathcal{Q}\}$, where $\phi^\mathcal{Q}_i = \sum_{j=1}^i \theta_j$ and restrict $-1 < \theta_j < 0$. We assume an independent beta prior over $1 + \theta_j$:

$$1 + \theta_j \sim B(a, b)$$

where $B$ denotes the density of a beta distribution and we set $a = 1000$ and $b = 10$. In initial investigations with a flat prior (as used by Bauer (2016)), we found that the posterior distributions for a number of parameters became extremely wide and the Gibbs sampler spent extremely long periods exploring regions with $\theta_j$ close to zero, where the likelihood surface becomes extremely flat. Our prior is nevertheless consistent with all factors being highly persistent under $\mathcal{Q}$ (the prior mean of $\theta_j$ is approximately -0.01) but relative to a flat prior downweights the possibility that $\theta_j$ is greater than about $-10^{-5}$ and greatly speeds the convergence of the sampler without having a material impact on the model’s ability to fit the cross-section of bond yields (so has a negligible impact on estimated term premia).

As proposed by Bauer (2016), at the $i^{th}$ draw in the chain we draw a candidate parameter vector $\theta_p$ according to

$$\theta_p \sim T_5\left(\theta^{(i-1)}, \Omega_\theta\right),$$

where $T_5$ denotes the density of a multivariate Student’s t-distribution with five degrees of freedom; $\theta^{(i-1)}$ is the $i-1^{th}$ draw in the chain; and the proposal covariance $\Omega_\theta$ is set equal to minus the inverse hessian of the likelihood function with respect to $\theta$ (evaluated at the initial values of the chain), scaled to achieve a reasonable Metropolis acceptance rate. The procedure for sampling $\delta_0$ is exactly analogous, with the exception that the prior is flat.
3.3.3 Factor covariance ($\Sigma$)

The procedure for drawing $\Sigma$ using another Metropolis-within-Gibbs step is also very similar to Bauer (2016). We assume an independent inverse Wishart prior over $\Sigma$:

$$\Sigma \sim \mathcal{IW}(\nu_\Sigma, \Psi_\Sigma).$$  \hspace{1cm} (27)

where we choose the prior hyperparameters using a training sample, as explained in Appendix B.

At the $i^{th}$ draw in the chain, we draw a proposal $\Sigma_p$ according to

$$\Sigma_p \sim \mathcal{IW}\left(\nu_{\Sigma,p}, \Psi_{\Sigma,p}^{(i)}\right),$$  \hspace{1cm} (28)

where $\mathcal{IW}$ denotes the density of an inverse Wishart distribution; the shape parameter $\nu_{\Sigma,p}$ is tuned to achieve a reasonable acceptance rate; and the scale parameters $\Psi_{\Sigma,p}^{(i)}$ are set such that the mean of the proposal distribution is equal to $\Sigma^{(i-1)}$.

3.3.4 Measurement error covariance ($R_w$)

Finally, we assume an independent inverse Wishart prior for $R_w$:\footnote{This differs slightly from Bauer (2016), who assumes that the measurement error is independent across yields and has the same variance for all maturities (i.e. $R_w = \sigma^2 I_N$).}

$$R_w \sim \mathcal{IW}(\nu_w, \Psi_w)$$  \hspace{1cm} (29)

with $\nu_w = N + 2$ and $\Psi_w = \frac{0.05}{1200} \times I$ (i.e. mean variances of five basis points for each bond yield expressed in annualized percentage points). The posterior is given by

$$R_w|Y, X, \delta_0, \theta, \Sigma \sim \mathcal{IW}\left(\bar{\nu}_w, \bar{\Psi}_w\right),$$  \hspace{1cm} (30)
where

\[
\begin{align*}
\bar{\gamma}_w &= \gamma_w + T \\
\bar{\Psi}_w &= \Psi_w + \sum_{t=1}^{T} w_t w'_t.
\end{align*}
\]

4 Results

4.1 Parameter estimates

Table 1 reports parameter estimates for the benchmark model with the long-run prior.\(^{15}\) As is standard, the factors are highly persistent under the risk-neutral dynamics (the largest eigenvalue of \(\Phi^Q\) is very close to one). The factors are also persistent under the time-series measure, but the posterior distributions for the parameters \(\Phi\) are much wider than those of the risk-neutral equivalent, reflecting the lack of information in the time-series of yields relative to the cross-section.

Table 2 shows estimates of the long-run mean parameters from the model with the long-run prior (panel (b)) and from the model with the flat prior (panel (c)). Percentiles of the long-run prior distribution are shown for reference in panel (a). With a flat prior over \(\gamma\), the unconditional means of yields are implausibly low - for example, at the posterior mean, the average 3-month Treasury bill rate is -3.7%, rising to only -1.0% for the 10-year yield. The posterior distributions are also extremely wide - for example, the central 90% of the posterior distribution for the average 3-month Treasury bill rate covers the region between -15.9% and 6.0%. In contrast, in the model with the long-run prior, estimates of the unconditional yield curve are much more reasonable and probability intervals are much narrower - for example, the posterior mean of the long-run mean of the Treasury bill rate in the model with the long-run prior is 4.7%, with a 90% probability interval covering the range 3.7% to 5.7%. The posterior 90% probability intervals from the model with the long-run prior are all somewhat narrower than the equivalent prior intervals, showing that

\(^{15}\)In the interests of space, the table omits the parameters of the measurement error covariance matrix \(R_w\). Parameter trace plots and posterior histograms for the elements of \(\mu, \Phi\) and \(\delta_0\), which are the most important parameters for determining the decomposition of yields, are reported in Appendix E.
4.2 Yield curve decompositions

In the model with flat priors over the long-run mean, the fact that yields revert to implausibly low long-run averages is likely to lead the models to underestimate the component of yields that reflects expected future policy rates. Between October 1992 and December 2014, the model-implied average expected short-term interest rate over ten-year horizons (shown in panel (b) of Figure 3) were on average around 2.8%, which seems rather low given the MPC’s inflation target and average UK real GDP growth (see above). In late 2012 it was close to zero, implying that the 10-year yield of around 3% (panel (a) shows the 10-year yields) was entirely made up of a term premium (panel (c)). Moreover, the uncertainty around these point estimates is extremely wide - for example, the average width of the 90% posterior probability interval for the 10-year term premium is almost always above 3 percentage points.

The broad dynamics of the posterior mean term premium are similar in the model with the long-run prior. But the average expected short rate over a 10-year horizon between October 1992 and December 2014 is around a percentage point higher compared with the model with the flat prior (panel (b) of Figure 4) and the term premium is correspondingly lower (panel (c)). And the 90% posterior probability interval is also considerably narrower (Figure 5), except over the first few years of the sample. Note that the width of the probability intervals for both models rises towards the end of the sample period, a point which we return to when discussing the variant of the model with a Minnesota prior below.\textsuperscript{16}

\textsuperscript{16}One potential concern with a Gaussian affine term structure model is that it is inconsistent with the zero lower bound on nominal interest rates. In Appendix C we extend our framework to the ‘shadow rate’ setting proposed by Black (1995) - which is consistent with the zero lower bound - and show that the main results reported here are not substantively affected. If anything, the probability interval is even narrower in the shadow rate model over the recent period of low nominal interest rates.
4.3 Sub-sample stability

One consequence of the substantial uncertainty around estimates of the time-series dynamics of bond yields in the model with flat priors is that model-implied term premia may vary substantially across sub-samples - see e.g. Guimarães (2016). This is illustrated by panel (a) of Figure 6, which shows estimates of UK 10-year term premia obtained using the model with flat priors over the long-run mean over different sub-samples. The sample periods correspond to the dates plotted for each of the five lines on the chart: October 1992-December 2014 (i.e. the full sample); October 1992-December 2004; October 1992-December 1999; October 2002-December 2014; and October 2007-December 2014. The dispersion across the different estimates is substantial, with estimates from the shortest sub-samples differing markedly from the full-sample estimates. In contrast, in the model with the long-run prior, the dispersion across the sub-samples is generally much smaller (panel (b)).

<Figure 6 about here>

This is likely to be a considerable practical benefit to users of these models because the appropriate choice of sample period is often not entirely straightforward. For example, as discussed above, we have chosen a sample that starts in October 1992, on the basis that this coincided with the introduction of an inflation targeting framework in the UK and that making use of previous data would increase the likelihood of a structural break in the sample. But it is also plausible that there are other structural breaks in our sample - such as the introduction of Bank of England operational independence for setting monetary policy in May 1997. The fact that providing additional information in the form of the long-run prior results in estimates of term premia is therefore a distinct advantage of our approach.
4.4 A Minnesota prior for $\Phi$

As explained in Section 3, the models reported above assume a diffuse prior over $\Phi^*$ (i.e. $V_\phi^{-1} = 0$). An obvious alternative is to consider a Minnesota prior, i.e.:

$$\phi_{ij}^* = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Under this prior, coefficients on own lags are shrunk towards one and those on the lags of other yields towards zero. One rationale for such a prior is that previous studies using US data have shown that restricting the factors in term structure models to be independent (i.e. off-diagonal elements of $\Phi$ equal to zero) improves out-of-sample forecasts of bond yields (e.g. Christensen et al. (2011)). A second rationale is that, as discussed above, OLS estimates of autoregressive processes are biased, with the bias generally tending to push down on the estimated persistence of yields; Bauer et al. (2012) show that bias-correcting the dynamics of US yields can have a substantial impact on point estimates of US term premia. Shrinking the diagonal elements of $\Phi^*$ towards one in a Bayesian setting will have a similar impact as bias correction in a classical framework - in that both will tend to increase the estimated persistence of the pricing factors.\(^\text{17}\)

Clearly, much depends on the tightness of the Minnesota prior. If $V_\phi$ is extremely small, the posterior will be shrunk very aggressively towards a random walk; whereas if the prior is diffuse the Minnesota prior will have little effect on the posterior. We specify the prior variance as $V_\phi = 0.001 \times I_{K^2}$ on the basis of a preliminary calibration exercise to maximize the marginal likelihood of a VAR of bond yields, which is very much in the spirit of the approach taken by Del Negro and Schorfheide (2004), who calibrate the tightness of a Minnesota prior in order to maximize the marginal likelihood of a Bayesian VAR. Appendix B provides further details of this exercise.

Figure 7 shows the resulting decomposition of the UK 10-year bond yield into average expected short-term rates and term premium from a model that imposes both the long-run and Minnesota priors. Relative to the model with only the long-run prior (Figure 4), \(^\text{17}\)Jarocinski and Marcet (2010) discuss the difference between Bayesian and classical interpretations of bias in OLS estimates of autoregressive models.
estimates of term premia are somewhat lower in the early part of the sample and a little higher in the later part, similar to the results reported by Malik and Meldrum (2016) for a UK term structure model with bias-corrected dynamics in a classical setting. The reason is intuitive: like classical bias corrections, the Minnesota prior raises the persistence of yields. The higher the estimated persistence of yields, the longer it will take for yields to revert back to their long-run average. When short rates are unusually low, the expected future path of short rates is therefore lower in a model with the Minnesota prior - and the term premium component commensurately higher. The Minnesota prior also has the effect of narrowing the 90% probability interval at the beginning and particularly the end of the sample (Figure 5). At times when yields are further away from the long-run averages, the impact of uncertainty about the persistence of yields - i.e. how long it will take for yields to revert back to the mean - on the uncertainty around term premia will be greater.

<Figure 7 about here>

4.5 Out-of-sample forecasting

Ultimately, if we are concerned with the ability of models to estimate the time-series dynamics of yields - and hence term premia - accurately, an obvious way to discriminate between different priors is to assess their impact on the out-of-sample forecasting performance of the model. Moreover, previous studies have shown that both long-run and Minnesota priors (Villani (2009)) can result in improved forecasting performance of Bayesian VARs, so it is natural to consider whether a similar result arises in this context.

For these purposes we consider a simplified version of the model - specifically, a first-order VAR of the 3-month and 5- and 10-year yields i.e. equation (12) above. This means that we only have to estimate the time-series dynamics of yields, and can ignore the parameters that determine the cross-sectional loadings of the yield curve on the factors, which greatly speeds the convergence of the estimation and makes a recursive out-of-sample forecasting exercise more tractable. Although a VAR of these yields is not exactly the same as the ATSM we report above, this is very unlikely to make a material difference when it comes to the

\footnote{For example, Giannone et al. (2015) set the tightness of priors for Bayesian VARs in order to optimize out-of-sample forecast performance.}
forecasting performance of the model. First, as discussed above and by Joslin et al. (2011), we can rotate the factors of an ATSM into any linear combination of yields without affecting either the contemporaneous cross-section of yields or expected future yields. Second, in the model reported above, we have already assumed that these three yields are observed without error, which is in effect what the simplified VAR of yields implies. And third, Duffee (2011) shows that imposing no-arbitrage restrictions makes a negligible impact on forecasting performance relative to unrestricted factor models.

We estimate (12) separately with (i) flat priors over $\gamma$; (ii) our long-run prior; and (iii) both the long-run and Minnesota prior. We initially estimate the models using the first 10 years of data (i.e. October 1992-September 2002) and produce out-of-sample forecasts of the three yields in the VAR at 1-, 3-, 6- and 12-month horizons. We then add an additional month of data to the estimation period and produce new forecasts with the same horizons and repeat until the end of the sample period (reserving 12 months of data for forecast evaluation, the final forecasts are produced in December 2013).

For each estimation period $t = 1, 2, \ldots, \tau$ we draw 5,000 times using the Gibbs sampling procedure for $\gamma$, $\Phi^*$ and $\Sigma^*$ outlined in Section 3 and discard the first 2,500 draws as a burn-in. After each iteration of the Gibbs sampler, we produce a single draw for $\{x^*_{i,t+h}\}_{h=1}^{12}$. We then computed estimated root mean squared forecast error (RMSFE) statistics for the $i^{th}$ element of $x^*$ at an $h$-month forecast horizon as:

$$RMSFE_{i,h} = \sqrt{\frac{1}{2500 \times T - \tau + 11} \sum_{j=1}^{2500} \sum_{t=\tau}^{T-12} \left( E_t \left[ x^*_{i,t+h} \right] - x^*_{i,t+h} \right)^2}, \quad (31)$$

where $E_t \left[ x^*_{i,t+h} \right]$ denote the time-$t$ conditional expectation of $x^*_{i,t+h}$ and $x^*_{i,t+h}$ denotes the out-turn.

Table 3 reports these RMSFEs for the three different types of prior. Using the long-run prior results in improved forecasting performance (i.e. smaller RMSFEs) for all the considered yields and forecast horizons. The differences are particularly great for longer maturity yields and at longer forecast horizons. This provides evidence that the long-run prior indeed results in gains in out-of-sample forecasting performance, consistent with results
found for the application of these priors in other settings (e.g. Villani (2009)). Using the Minnesota prior as well as the long-run prior further improves the forecasting performance, which suggests that there are forecasting gains from shrinking the persistence of the factors towards random walks. This seems broadly consistent with the findings of Duffee (2011), who shows that imposing a unit root on the level of the term structure improves the out-of-sample forecasting performance of dynamic term structure models.

5 Alternative specifications

The results reported above demonstrate that we can alleviate the problem of parameter uncertainty in term structure models substantially by introducing some very reasonable prior information about the unconditional mean of the yield curve. But users of these models should remain mindful of the fact that this does nothing to address the substantial model uncertainty associated with term structure models. In this section, we explore the implications of two forms of model uncertainty: first, allowing for a lower bound on nominal interest rates, using the shadow rate framework proposed by Black (1995); and second, extending the model to allow for time-variation in the long-horizon expectation of yields.

5.1 Long-run priors in a shadow rate term structure model

One potential drawback of a Gaussian affine term structure model over our sample is that the model is not consistent with a lower bound on nominal interest rates. When interest rates are close to zero, as has been the case towards the end of our sample, this means that the model can imply a significant probability of negative nominal interest rates (a point made previously by a number of studies, including Andreasen and Meldrum (2013) and Bauer and Rudebusch (2014)). A potential concern could therefore be that the results reported in Section 4 are driven by the fact that we were estimating an affine model over a period that ended with very low short-term interest rates. To demonstrate that this is not likely to be the case, this section shows that we can apply a similar long-run prior in a model that does impose the zero bound on nominal interest rates, with only minimal changes to the specification, and that term premium estimates from such a model are actually even
lower than in the benchmark model.

5.1.1 Specification

In the shadow rate model, as proposed by Black (1995), the short-term interest rate is the maximum of zero and a ‘shadow rate’ of interest ($s_t$):

$$i_t = \max \{0, s_t\},$$

which is affine in the pricing factors

$$s_t = \delta_0 + \delta'_1 x_t.$$

The risk-neutral (3) and time-series (7) dynamics of the pricing factors are the same as in the affine model. While the shadow rate specification ensures that bond yields are non-negative, unfortunately there are no closed-form expressions for yields as functions of the pricing factors and structural parameters of the model. We therefore use the second-order approximation to yields proposed by Priebsch (2013), applied previously in a discrete-time setting by Andreasen and Meldrum (2015a) (for the US) and (for the UK) by Andreasen and Meldrum (2015b).

Since the mapping between yields and factors is non-linear in the shadow rate model is non-linear, we cannot simply specify priors about the long-run values of bond yields by inverting the pricing factors. We can, however, specify priors on the ‘shadow term structure’, which is defined as

$$s_t^{(n)} = -\frac{1}{n} \left(a_n + b'_n x_t\right),$$

where $a_n$ and $b'_n$ follow the same recursive equations as in the affine model, i.e. (5) and (6) above. We can think of the shadow term structure as the bond yields that would apply if there were no lower bound on nominal interest rates. This is convenient, since it means that we can specify a long-run prior about the shadow term structure in exactly the same way as before.
5.1.2 Estimation

A related complication when working with the shadow rate model (given the non-linear relationship between yields and factors) is that we can no longer extract the factors using the Chen and Scott (1993) inversion.\textsuperscript{19} We instead assume that all \( N \) yields \( (y_t) \) are observed with additive measurement error, i.e.

\[
y_t = g(x_t; \delta_0, \delta_1, \mu^Q, \Phi^Q, \Sigma) + w_t
\]

\[
w_t \sim \text{i.i.d.} \mathcal{N}(0, R_w)
\]

where \( g(x_t; \delta_0, \delta_1, \mu^Q, \Phi^Q, \Sigma) \) is the non-linear function given by the Priebsch (2013) approximation, and estimate the factors using an adaptation of the single-move procedure proposed by Jacquier et al. (1994). At the \( i^{th} \) step in the Gibbs sampler, for each time period in turn we construct a proposal \( x_t' \) according to

\[
x_t' \sim \mathcal{N}(x_t^{(i-1)}, R^{CDKF}_{x_t})
\]

where \( x_t^{(i-1)} \) is the \( i-1^{th} \) draw of the factors at time \( t \) and \( R^{CDKF}_{x_t} \) is the filtered covariance matrix for \( x_t \) obtained using the Central Difference Kalman Filter of Norgaard et al. (2000) evaluated at the initial parameter values. We assume a flat prior over \( x_t \) and initialise the chain at the filtered values obtained by running a single pass of the Central Difference Kalman Filter, again at the initial parameter values.

5.1.3 Results

Figure 8 shows estimates of the 10-year term premium from the shadow rate model with the long-run prior. Until the period of near-zero short-term interest rates towards the end of the sample, the posterior mean term premium estimates from the model are very similar to those from the affine model (4). More recently, however, the estimated term premium from the shadow rate model has been lower than that from the affine model. This contrasts\textsuperscript{19}

\textsuperscript{19}In the classical literature on shadow rate models, the factors are typically estimated using a non-linear extension of the Kalman filter (e.g. Christensen and Rudebusch (2013), Kim and Priebsch (2013) and Bauer and Rudebusch (2014)) or using non-linear regression (e.g. Andreasen and Meldrum (2015a)).
slightly with previous findings by Kim and Priebsch (2013) (for the US) and Malik and Meldrum (2016) (for the UK) that long-maturity term premia from shadow rate models are similar to those from affine models. If anything, the model-implied uncertainty around the term premium estimates is much narrower than in the affine model during the recent period of very low nominal interest rates.

<Insert Figure 8 here.>

5.2 Allowing for time-variation in long-horizon expectations

Standard Gaussian affine term structure models assume that the unconditional mean of yields is time-invariant. In contrast, Kozicki and Tinsley (2001) and Van Dijk et al. (2014) find that allowing for changing expectations of infinite-horizon conditional expectations of short rates (‘shifting end-points’) improves forecasts of bond yields. We therefore illustrate the potential impact of this form of model uncertainty by extending the standard model presented above to allow for shifting end-points in a Bayesian setting.

5.2.1 Specification

In our model with shifting end-points, the time-series dynamics of the pricing factors from the standard model (7) are modified to allow for a time-varying intercept:

\[ x_{t+1} = x_{t}^{(\infty)} + \Phi \left( x_{t} - x_{t}^{(\infty)} \right) + v_{t+1}, \]

where \( x_{t}^{(\infty)} \) denotes the limit of the conditional expectation of the pricing factors \( E_t [x_{t+h}] \) as \( h \to \infty \), the time-\( t \) end-point for the pricing factors.\(^{20}\) Similar to the approach taken in Section 3, where we define \( x_t^* = \begin{bmatrix} y_t^{(3)} & y_t^{(60)} & y_t^{(120)} \end{bmatrix}' = W'y_t \), we can re-write (33) as

\[ x_{t+1}^* - x_{t+1}^{(\infty)} = \Phi^* (x_t^* - x_t^{(\infty)}) + v_{t+1}^* \]

\[ v_{t+1}^* \sim N(0, \Sigma^*), \]

\(^{20}\)In the standard model, the end-point is constant, \( x_{t}^{(\infty)} = (I - \Phi)^{-1} \mu \).
where

\[
\begin{align*}
\mathbf{x}_t^{(\infty)} &= \lim_{h \to \infty} \mathbb{E}_t [\mathbf{x}_{t+h}] \\
\Phi &= (\mathbf{W}'\mathbf{B})^{-1} \Phi \mathbf{W}' \mathbf{B} \\
\Sigma &= (\mathbf{W}'\mathbf{B})^{-1} \Sigma^* (\mathbf{B}'\mathbf{W})^{-1}.
\end{align*}
\]

Previous studies allowing for shifting end-points fall into one of three categories. First, some model the end-point using purely statistical time-series approaches - for example, Kozicki and Tinsley (2001) and Van Dijk et al. (2014) consider models in which end-points are weighted averages of past out turns, while Bianchi et al. (2009) allow for time-varying parameters in the time-series dynamics of the pricing factors. Second, Dewachter and Lyrio (2006) link the end-points to stochastic trends in GDP and inflation, which are included as observed factors within a dynamic term structure model. Our model falls into a third category, which is to link estimates of end-points to long-horizon survey expectations of future interest rates, similar to another approach proposed by Van Dijk et al. (2014). Specifically, we use long-horizon surveys of interest rates from Consensus Economics to construct observed proxies for the end-points for \( \mathbf{x}_t^{\star} \), using the method explained in Appendix C. We assume that these proxies \((\mathbf{s}_t)\) are measured with error \((\varepsilon_t)\), such that

\[
\begin{align*}
\mathbf{s}_t &= \mathbf{x}_t^{(\infty)} + \varepsilon_t \\
\varepsilon_t &\sim \mathcal{N}(0, \mathbf{R}_\varepsilon).
\end{align*}
\]

Finally, as in Van Dijk et al. (2014), we assume that the end-points follow a random walk:

\[
\begin{align*}
\mathbf{x}_{t+1}^{(\infty)} &= \mathbf{x}_t^{(\infty)} + \eta_{t+1} \\
\eta_{t+1} &\sim \mathcal{N}(0, \mathbf{R}_\eta).
\end{align*}
\]

5.2.2 Estimation

The Gibbs sampler for the model with shifting end-points has three blocks of parameters that do not appear in the standard model: (i) the unobserved end-points \( \mathbf{x}_t^{(\infty)} \) for \( t = \)]
1, 2, ..., T; (ii) the covariance matrix of shocks to the end-points (R_\eta); and (iii) the variance of end-point measurement errors (R_\varepsilon). The remainder of this section explains how draws from these blocks are obtained in turn.\footnote{In addition, the procedure for drawing the time-series dynamics of the pricing factors is modified trivially to account for the shifting end-point.}

**Unobserved end-points (x_{1:T}^{*(\infty)})** Equation (36) can be written in companion form as:

\begin{equation}
\begin{bmatrix}
    \begin{bmatrix}
        x_t^{*(\infty)} \\
        x_{t-1}^{*(\infty)}
    \end{bmatrix} \\
    x_{t-1}^{*(\infty)}
\end{bmatrix} = \begin{bmatrix}
    I & 0 \\
    I & 0
\end{bmatrix} \begin{bmatrix}
    \begin{bmatrix}
        x_{t-1}^{*(\infty)} \\
        x_{t-2}^{*(\infty)}
    \end{bmatrix} \\
    0
\end{bmatrix} + \begin{bmatrix}
    \eta_t \\
    0
\end{bmatrix},
\end{equation}

(37)

while equations (34) and (35) can be written as:

\begin{equation}
\begin{bmatrix}
    \begin{bmatrix}
        x_t^{*} - \Phi^* x_{t-1}^{*} \\
        s_t
    \end{bmatrix} \\
    \begin{bmatrix}
        x_t^{*} - \Phi^* x_{t-1}^{*} \\
        s_t
    \end{bmatrix}
\end{bmatrix} = \begin{bmatrix}
    I & -\Phi^* \\
    I & 0
\end{bmatrix} \begin{bmatrix}
    \begin{bmatrix}
        z_t \\
        0
    \end{bmatrix} \\
    \begin{bmatrix}
        v_t \\
        \varepsilon_t
    \end{bmatrix}
\end{bmatrix}.
\end{equation}

(38)

Conditional on \Phi^*, R_\varepsilon, R_\eta and \Sigma, (37) and (38) form a linear-Gaussian state-space system, with the end-points as unobserved state variables. We therefore use the method of Durbin and Koopman (2002) to draw the end-points. This requires an independent prior mean and variance for the initial state vector, which we set to be diffuse:

\[ z_1 \sim \mathcal{N}(0, 10^6 I_{2K}) . \]

**Covariance of end-point shocks (R_\eta)** For simplicity, we specify the covariance matrix of end-point shocks to be a diagonal matrix, i.e. \[ R_\eta = \text{diag}\left\{\sigma_\eta_1^2, ..., \sigma_\eta_K^2\right\} . \] For each of the diagonal elements, we assume an independent inverse gamma prior, i.e.:

\[ \sigma_\eta_i^2 \sim IG\left(\nu_\eta, \nu_\eta\right), \]

where we set \( \nu_\eta = 2 \) and \( \nu_\eta = \frac{5.7}{1200} \times 10^{-9} \). This ensures that the prior mean is equal to the sample variance of changes in the survey-based measure of the end-point for the ten-year yield, \( Var_t (\Delta s_{3,t}) \). Drawing from the posterior is then straightforward as the prior is conjugate.
Covariance of end-point measurement errors (R_e) For simplicity, we specify the variance of end-point measurement errors to be the same for all maturities and that the measurement errors are uncorrelated across maturities, i.e. \( R_e = \sigma_e^2 \times I \). We again assume an inverse gamma prior, i.e.

\[
\sigma_e^2 \sim IG\left(\nu_e, \psi_e\right),
\]

where we set \( \nu_e = 2 + 10^{-6} \) and \( \psi_e = 0.001 \), which are chosen such that the prior mean for \( \sigma_e^2 \) is approximately 0.1 percentage points and the variance is approximately 1 percentage point. This is consistent with our belief that the survey-based measure is likely to provide a noisy measure of the true end-point but that we are uncertain exactly how noisy.

5.2.3 Results

Figure 9 shows a decomposition of the UK 10-year yield from the model with shifting end-points. Comparing this with the equivalent decomposition from the model with the long-run prior over a constant end-point (Figure 4) shows that allowing for shifting end-points has a material impact. Most strikingly, early in the sample estimates of average expected short rates over the next ten years from the model with shifting end-points are higher. This is because estimates of the end-points for yields (Figure 10) were substantially higher - between 6-8% - during the early 1990s than more recently, when they have been closer to 4% (which is broadly similar to the estimates of the constant unconditional mean from the model with the long-run prior). As a result, estimates of term premia from the model with shifting end-points do not exhibit the usual pattern of falling over time, which is a feature of models that have a constant mean.

<Figure 9 about here>

<Figure 10 about here>

Comparing term premium estimates from the model with the shifting end-point with those from the model with the constant end-point but with both the long-run and Minnesota priors is instructive. While there are differences, there are some common features; in particular, that estimates of term premia are low at the beginning of the sample and high at the end in comparison with the other models. In terms of long-horizon term premia, a
model with a shifting end-point (i.e. a random walk in the long-run mean) is quite similar to a model in which yields follow a stationary - but extremely persistent - process.

6 Conclusions

Classical estimation of maximally flexible dynamic, no-arbitrage term structure models is beset with problems associated with the weak identification of the time-series dynamics. We argue that to a large extent this stems from the lack of sample information about the unconditional mean of the yield curve. This paper develop a tractable approach for incorporating prior information about the unconditional mean of yields in a Bayesian setting, which goes a long way to reducing problems of parameter uncertainty. We build on the work of Villani (2009) who proposes a way to specify a prior about the unconditional mean in Bayesian VAR models, and Bauer (2016) who uses a similar Bayesian method for estimating affine term structure models. We rotate the term structure model pricing factors into bond yields and specify priors on the unconditional means of those yields, calibrated using a very simple macroeconomic model. Parameters of the time-series dynamics are then drawn within a Gibbs sampling procedure. We find that with reasonable priors we obtain more plausible estimates of the long-run average of yields, lower estimates of term premia in long-term bonds, substantially reduced uncertainty, greater sub-sample stability of term structure decompositions and substantial improvements in out-of-sample forecast performance.

References


Appendix A: Maximum likelihood estimation

We initialise our Gibbs sampler at maximum likelihood estimates of the three-factor affine term structure model. As explained in Section 3, conditional on values of the parameters $\Psi = \{\delta_0, \delta_1, \mu^Q, \Phi^Q, \Sigma\}$, we can use the procedure of Chen and Scott (1993) to invert (9) and recover the pricing factors, i.e. $x_t = B^{-1}_1 (x_t^* - A_1)$. We can then compute the likelihood of observing data $y_{2,1:T} = \{y_{2,1}, y_{2,2}, \ldots, y_{2,T}\}$ as

$$p (y_{2,1:T} | \Psi, \mu, \Phi, R_{w2}) = \prod_{t=1}^T p (x_{t+1} | x_t; \Psi, \mu, \Phi, R_{w2}) \prod_{t=1}^T p (y_{2,t} | x_t; \Psi, R_{w2}).$$

Appendix B: Calibration of priors for $\Phi$ and $\Sigma$

As explained in Section 3, we calibrate the prior distributions for the factor persistence under the time-series measure ($\Phi$) and the covariance of disturbances to the factors ($\Sigma$) on the basis of a preliminary estimation of the model using a training sample.

We first estimate the standard three-factor Gaussian ATSM by maximum likelihood (using the procedure set out in Appendix A) using month-end yields for the period January 1975-September 1992, which gives an maximum-likelihood estimates:

$$\left\{ \tilde{\gamma}^{\text{train}}, \tilde{\phi}_0^{\text{train}}, \tilde{\Sigma}^{\text{train}}, \tilde{\mu}^{\text{train}}, \tilde{\Phi}^{\text{train}} \right\}.$$

The prior hyperparameters for $\Sigma$ (27) are given by

$$\nu_{\Sigma} = K + 2$$
$$\Psi_{\Sigma} = \tilde{\Sigma}^{\text{training}}$$

In the second step, we choose the tightness of the Minnesota prior on $\Phi$ in order to maximise the marginal data density of a time-series VAR of yields. Since we can always rotate the pricing factors in our model into a vector of yields, our approach is very much in the spirit of the method proposed by Del Negro and Schorfheide (2004) for choosing the tightness of priors for Bayesian VARs. The remainder of this appendix explains the
procedure in further detail.

The prior calibration is based on a first-order VAR of $x_t^* = \left[ y_t^{(3)}, y_t^{(60)}, y_t^{(120)} \right]$:

$$
\tilde{x}_{t+1}^* = \Phi^* \tilde{x}_t^* + v_{t+1}
$$

where $\tilde{x}_{t+1}^* = x_t^* - \bar{x}^*$ and $\bar{x}^*$ is the sample mean of $x_t^*$ over $t = 1, 2, ..., T$. The prior for $\Sigma^*$ is given by

$$
\Sigma^* \sim \mathcal{IW} (\nu_{\Sigma^*}, \Psi_{\Sigma^*}),
$$

where

$$
\nu_{\Sigma^*} = K + 2
$$

$$
\Psi_{\Sigma^*} = \left( W^T \hat{B}^{train} \right) \hat{\Sigma}^{train} \left( \hat{B}^{train} W \right),
$$

where $\hat{B}^{train}$ is computed at the maximum likelihood estimates obtained for the training sample. As in the full affine term structure model, the prior for $\phi^* = \text{vec} (\Phi^*)$ is independent Normal:

$$
\phi^* | \Sigma^* \sim \mathcal{N} (\bar{\phi}, \mathbf{V}_\phi),
$$

where

$$
\phi^*_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}
$$

and $\mathbf{V}_\phi = \kappa \mathbf{I}_{n_x^2}$.

We compute the marginal likelihood of the model using the method of Chib (1995) for values of $\kappa = \{ 10^{-10}, 10^{-9}, ..., 10^0, ..., 10^9, 10^{10} \}$. The marginal likelihood is maximised for $\kappa = 10^{-3}$, which is the value we use for the estimations reported in the main text.
Appendix C: A survey-based measure of shifting end-points

This appendix explains our proxy for shifting end-points in the term structure based on Consensus Economics surveys of professional economists, which we use to estimate the model set out in Section 5. Every six months (in April and October), Consensus Economics surveys a panel of professional economists about their expectations for the UK ten-year bond yield at forecasts of two, three, four and five years, in addition to average expectations of the ten-year yield between six and ten-years ahead.

To estimate an infinite-horizon conditional expectation for the ten-year yield, we fit the parametric curve of Nelson and Siegel (1987) to the term structure of survey expectations (we use the mean forecast across respondents), separately for each survey date. The $h$-period ahead expectation is modelled as:

$$E_t \left[ y_{t+n}^{(120)} \right] = \beta_0 + \beta_1 \exp \left( -\frac{n}{\lambda} \right) + \beta_2 \left( \frac{n}{\lambda} \right) \exp \left( -\frac{n}{\lambda} \right) + u_t^{(n)}$$

we set the parameter $\lambda$ to maximise the loading on $\beta_2$ at a maturity of $n = 36$ months. We can then estimate $\beta_0$, $\beta_1$ and $\beta_2$ by ordinary least squares. This functional form implies that

$$s^{(120)}_t = \lim_{n \to \infty} E_t \left[ y_{t+n}^{(120)} \right] = \beta_0,$$

so we take the estimated value of $\beta_0$ to be our end-point for the ten-year yield.

Since Consensus Economics only asks respondents for expectations about the ten-year yield, we separately regress the three-month and five-year yields on the ten-year yield using end-month data during our sample period of October 1992 to December 2014. For example, for the three-month yield, we estimate the following equation by ordinary least squares:

$$y_{t}^{(3)} = \alpha_0 + \alpha_1 y_{t}^{(120)} + \zeta_{t}^{(3)}.$$

We use the estimated coefficients, to obtain an end-point proxy for the three-month yield:

$$s^{(3)}_t = \tilde{\alpha}_0 + \tilde{\alpha}_0 s^{(120)}_t.$$
Finally, we linearly interpolated between the six-monthly surveys to obtain monthly observations of the end-points for all maturities.
Table 1: Posterior parameter estimates for benchmark affine model of nominal yields with long-run prior

| Parameter | (a) Flat priors | | | | | (b) Long-run priors | | |
|-----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|           | 5th percentile | Mean           | 95th percentile | 5th percentile | Mean           | 95th percentile |
| $\delta_0$ | 0.0026         | 0.0039         | 0.0050         | 0.0022         | 0.0035         | 0.0048         |
| $\phi_{10}$ | 0.9995         | 0.9997         | 0.9998         | 0.9995         | 0.9997         | 0.9998         |
| $\phi_{11}$ | 0.9812         | 0.9825         | 0.9837         | 0.9813         | 0.9826         | 0.9838         |
| $\phi_{12}$ | 0.9289         | 0.9348         | 0.9405         | 0.9283         | 0.9341         | 0.9404         |
| $\sigma_{11}$ | $2.695 \times 10^{-4}$ | $3.013 \times 10^{-4}$ | $3.372 \times 10^{-4}$ | $2.683 \times 10^{-4}$ | $3.019 \times 10^{-4}$ | $3.373 \times 10^{-4}$ |
| $\sigma_{21}$ | $-5.019 \times 10^{-4}$ | $-4.132 \times 10^{-4}$ | $-3.308 \times 10^{-4}$ | $-5.093 \times 10^{-4}$ | $-4.156 \times 10^{-4}$ | $-3.325 \times 10^{-4}$ |
| $\sigma_{22}$ | $3.846 \times 10^{-4}$ | $4.367 \times 10^{-4}$ | $4.985 \times 10^{-4}$ | $3.782 \times 10^{-4}$ | $4.317 \times 10^{-4}$ | $4.860 \times 10^{-4}$ |
| $\sigma_{31}$ | $0.327 \times 10^{-4}$ | $1.045 \times 10^{-4}$ | $1.795 \times 10^{-4}$ | $0.369 \times 10^{-4}$ | $1.063 \times 10^{-4}$ | $1.805 \times 10^{-4}$ |
| $\sigma_{32}$ | $-4.871 \times 10^{-4}$ | $-4.201 \times 10^{-4}$ | $-3.607 \times 10^{-4}$ | $-4.830 \times 10^{-4}$ | $-4.170 \times 10^{-4}$ | $-3.575 \times 10^{-4}$ |
| $\sigma_{33}$ | $1.603 \times 10^{-4}$ | $1.797 \times 10^{-4}$ | $2.018 \times 10^{-4}$ | $1.601 \times 10^{-4}$ | $1.790 \times 10^{-4}$ | $2.020 \times 10^{-4}$ |
| $\gamma_{1 \times 1200}$ | $-15.893$ | $-3.710$ | $6.029$ | $3.705$ | $4.677$ | $5.740$ |
| $\gamma_{2 \times 1200}$ | $-12.892$ | $-2.492$ | $6.394$ | $4.637$ | $5.161$ | $5.706$ |
| $\gamma_{3 \times 1200}$ | $-9.208$ | $-0.975$ | $6.355$ | $4.669$ | $5.195$ | $5.704$ |
| $\phi_{11}$ | $0.9422$ | $0.9669$ | $0.9912$ | $0.9439$ | $0.9686$ | $0.9932$ |
| $\phi_{12}$ | $-0.0152$ | $0.0015$ | $0.0185$ | $-0.0157$ | $0.0004$ | $0.0170$ |
| $\phi_{13}$ | $-0.0457$ | $-0.0144$ | $0.0174$ | $-0.0476$ | $-0.0159$ | $0.0154$ |
| $\phi_{21}$ | $-0.0140$ | $0.0353$ | $0.0849$ | $-0.0205$ | $0.0278$ | $0.0756$ |
| $\phi_{22}$ | $0.9494$ | $0.9818$ | $1.0122$ | $0.9551$ | $0.9862$ | $1.0139$ |
| $\phi_{23}$ | $-0.0509$ | $0.0120$ | $0.0740$ | $-0.0490$ | $0.0112$ | $0.0703$ |
| $\phi_{31}$ | $-0.0655$ | $-0.0274$ | $0.0103$ | $-0.0580$ | $-0.0211$ | $0.0146$ |
| $\phi_{32}$ | $-0.0222$ | $0.0021$ | $0.0271$ | $-0.0221$ | $-0.0003$ | $0.0232$ |
| $\phi_{33}$ | $0.8934$ | $0.9429$ | $0.9915$ | $0.8986$ | $0.9469$ | $0.9947$ |
Table 2: Prior and posterior estimates of long-run mean parameters in benchmark affine model of nominal yields

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5&lt;sup&gt;th&lt;/sup&gt; percentile</th>
<th>Mean</th>
<th>95&lt;sup&gt;th&lt;/sup&gt; percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Long-run prior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_t^{(3)})</td>
<td>3.685</td>
<td>5.171</td>
<td>6.656</td>
</tr>
<tr>
<td>(y_t^{(60)})</td>
<td>4.430</td>
<td>5.171</td>
<td>5.592</td>
</tr>
<tr>
<td>(y_t^{(120)})</td>
<td>4.441</td>
<td>5.173</td>
<td>5.905</td>
</tr>
<tr>
<td>(b) Model with long-run prior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_t^{(3)})</td>
<td>3.705</td>
<td>4.677</td>
<td>5.740</td>
</tr>
<tr>
<td>(y_t^{(60)})</td>
<td>4.637</td>
<td>5.161</td>
<td>5.706</td>
</tr>
<tr>
<td>(y_t^{(120)})</td>
<td>4.669</td>
<td>5.195</td>
<td>5.704</td>
</tr>
<tr>
<td>(c) Model with flat prior over (\gamma)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_t^{(3)})</td>
<td>(-15.893)</td>
<td>(-3.710)</td>
<td>6.029</td>
</tr>
<tr>
<td>(y_t^{(60)})</td>
<td>(-12.892)</td>
<td>(-2.492)</td>
<td>6.394</td>
</tr>
<tr>
<td>(y_t^{(120)})</td>
<td>(-9.208)</td>
<td>(-0.975)</td>
<td>6.355</td>
</tr>
</tbody>
</table>

Estimates of the long-run means of yields under the long-run prior (panel (a)), in the model with a long-run prior (panel (b)) and in the model with a flat prior over \(\gamma\). All numbers are annualized percentage points.
Table 3: Out of sample forecasting exercise

<table>
<thead>
<tr>
<th></th>
<th>Root mean square error</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( y_t^{(3)} )</td>
<td>( y_t^{(60)} )</td>
<td>( y_t^{(120)} )</td>
<td></td>
</tr>
<tr>
<td>(a) BVAR with flat priors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-month forecast</td>
<td>0.209</td>
<td>0.227</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td>3-month forecast</td>
<td>0.533</td>
<td>0.477</td>
<td>0.414</td>
<td></td>
</tr>
<tr>
<td>6-month forecast</td>
<td>0.894</td>
<td>0.758</td>
<td>0.658</td>
<td></td>
</tr>
<tr>
<td>12-month forecast</td>
<td>1.340</td>
<td>1.228</td>
<td>1.033</td>
<td></td>
</tr>
<tr>
<td>(b) BVAR with long-run prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-month forecast</td>
<td>0.208</td>
<td>0.218</td>
<td>0.226</td>
<td></td>
</tr>
<tr>
<td>3-month forecast</td>
<td>0.529</td>
<td>0.450</td>
<td>0.395</td>
<td></td>
</tr>
<tr>
<td>6-month forecast</td>
<td>0.886</td>
<td>0.697</td>
<td>0.612</td>
<td></td>
</tr>
<tr>
<td>12-month forecast</td>
<td>1.301</td>
<td>0.968</td>
<td>0.850</td>
<td></td>
</tr>
<tr>
<td>(c) BVAR with long-run prior and Minnesota prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-month forecast</td>
<td>0.207</td>
<td>0.216</td>
<td>0.223</td>
<td></td>
</tr>
<tr>
<td>3-month forecast</td>
<td>0.515</td>
<td>0.437</td>
<td>0.378</td>
<td></td>
</tr>
<tr>
<td>6-month forecast</td>
<td>0.849</td>
<td>0.662</td>
<td>0.571</td>
<td></td>
</tr>
<tr>
<td>12-month forecast</td>
<td>1.276</td>
<td>0.885</td>
<td>0.754</td>
<td></td>
</tr>
</tbody>
</table>

Root mean square forecast errors for out-of-sample forecasts from Bayesian VAR of yields. The forecasts are computed recursively, with an initial estimation period of October 1992-September 2002 and the final forecasts being produced in December 2013. All numbers are annualized percentage points.
Figure 1: UK end-month zero-coupon bond yields, October 1992-December 2014

Figure 2: UK 10-year zero-coupon bond yield with AR(1) model projection from October 1992. The 90% confidence interval (CI) for the unconditional mean is obtained using a bootstrap procedure, as explained in the text.
Figure 3: Decomposition of UK 10-year bond yield from the affine model with flat priors over the time-series dynamics.
Figure 4: Decomposition of UK 10-year bond yield from the affine model with the long-run prior

(a) Yield

(b) Average expected short rates

(c) Term premium

Mean  90% probability interval
Figure 5: Width of the 90% probability interval for the 10-year term premium in a model with the long-run prior and a model with a flat prior over the time-series dynamics.
Figure 6: Estimates of UK ten-year term premia using different sample periods from the model with flat priors over $\gamma$ (panel (a)) and the model with the long-run prior (panel (b)).
Figure 7: Decomposition of UK 10-year bond yield from the affine model with long-run and Minnesota priors.
Figure 8: Decomposition of UK 10-year bond yield from the shadow rate model with the long-run prior
Figure 9: Decomposition of UK 10-year bond yield from the model with shifting end-points

(a) Yield

(b) Average expected short rates

(c) Term premium

---

Mean

90% probability interval
Figure 10: Fitted end-points from the model with shifting end-points. The chart shows the posterior mean with a 90% posterior probability interval (PI), alongside the survey-based proxy described in Appendix D.
Figure 11: Trace plots for elements of $\mu$ and $\Phi$ from the model with the long-run prior

![Trace plots for elements of $\mu$ and $\Phi$ from the model with the long-run prior](image)

Figure 12: Posterior histograms for elements of $\mu$ and $\Phi$ from the model with the long-run prior

![Posterior histograms for elements of $\mu$ and $\Phi$ from the model with the long-run prior](image)
Figure 13: Trace plot and posterior histogram for $\delta_0$ from the model with the long-run prior