What does a term structure model imply about very long-term discount rates?

Anne Balter

Maastricht University

a.balter@maastrichtuniversity.nl

Antoon Pelsser

Maastricht University and Netspar

a.pelsser@maastrichtuniversity.nl

Peter Schotman

Maastricht University and Netspar

p.schotman@maastrichtuniversity.nl

February 25, 2014

Abstract: The purpose of the paper is to extrapolate the yield curve. Taking data from time series of observed yields up to 20 years, what are model implied yields for longer maturities? We use a standard Gaussian essentially affine model that we estimate by Bayesian methods. Our data consists of Euro swap rates from 2002 through 2013. We use an informative prior on the time series mean reversion of interest rates and on the cross sectional convergence of the yield curve. In the prior we further impose a zero lower bound on the unconditional mean of nominal interest rates.

Convergence of the yield curve to a constant ultimate forward rate appears very slow, while the ultimate yield (or ultimate forward rate) itself cannot be estimated with any accuracy. An important effect in the extrapolation is convexity, which causes extrapolated yields to move substantially above the ultimate forward rate. We find large differences with alternative curve fitting methods such as Smith-Wilson extrapolation and Nelson-Siegel calibration. Since our Bayesian procedure also estimates the uncertainty around the posterior mean estimate of the extrapolation, the alternatives are, however, mostly within our estimated credibility region.

Keywords: affine term structure, parameter uncertainty, Bayesian, extrapolation

JEL codes:

1 Introduction

Long maturity discount rates are an essential input for valuing the liabilities of pension funds and insurance companies. Life-insurance or pension fund liabilities can be as long as 100 years, whereas the available liquid instruments in the market have much shorter maturities. In most countries market rates can only be observed for maturities up to 20 or 30 years for government debt. Swap rates are available for maturities up to 50 years, but there are doubts about the liquidity of the longest maturities. Fair value, or market consistent, valuation requires discount rates that are close to market rates, but free of liquidity effects.[1]

For this purpose various methods have been proposed to extend an observed yield curve. Assuming that market rates are liquid up to a 'last liquid point' with maturity of 20 years (say), how should such a yield curve be extended to maturities up to 100 years? One option is the use of numerical extrapolation techniques. A prominent example is the Smith-Wilson methodology adopted by EIOPA[2], which extrapolates the forward rate curve using exponential functions.[3] The extrapolation method provides a smooth extension from the yield at the last liquid point to an externally specified ultimate forward rate (UFR) and a chosen convergence speed parameter. Other methods, such as the Nelson-Siegel methodology, would first fit level, slope and curvature factors using data on the liquid part of the yield curve and then extend the yield curve with the parameters of the fitted model.[4] In the Smith-Wilson methodology the yield curve always converges to the same constant, whereas in the Nelson-Siegel model long rates converge to a time-varying level factor estimated from the current term structure.

---

[1] Quoting from From Moody's Analytics (May 2013): 'Fair value is the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction (that is, not a forced liquidation or distressed sale) between market participants at the measurement date under current market conditions.'

[2] European Insurance and Occupational Pensions Authority

[3] See the EIOPA notes at `eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/ceiops-paper-extrapolation-risk-free-rates_en-20100802.pdf` for references and details of the method.

[4] See Diebold and Rudebusch (2013) for a textbook treatment of the Nelson-Siegel model for fitting term structure data.

A problem with these methods is that the extension is based on curve fitting, and not a formal term structure model. The Nelson-Siegel model can be made arbitrage free by adding a yield adjustment term, like in Christensen, Diebold, and Rudebusch (2011), but this adjustment will push very long term yields to minus infinity.[5] The strong downward pressure on long-term yields is caused by the unit root of the level factor under the risk neutral measure that drives the convexity adjustment. The arbitrage-free Nelson-Siegel model is a member of the class of essentially affine Gaussian term structure models (Duffee, 2002). Other models in that class do not necessarily have a unit root in the risk neutral dynamics, and will thus have a smaller convexity adjustment at the very long end of the yield curve.

The existence of a convexity effect is the main argument for using a formal term structure model for the extrapolation. It implies that yields do not always converge monotonically from the last liquid point towards the ultimate yield. Starting from low interest rate levels, the convergence will follow a hump shape, in which yields first overshoot the ultimate yield before a finally decreasing slowly towards the limit. With current low interest rates such an extrapolation will result in a much higher level for long-term yields than either the Nelson-Siegel or the Smith-Wilson extrapolation.

A secondary aim of the paper is to quantify the uncertainty around extrapolated yields given a formal term structure model. How much can we learn from time series of observed yields with medium to long term maturities (5 to 20 years) about the shape of the yield curve at very long maturities (between 20 and 100 years). More explicitly, what can we infer about the three crucial elements at the long end of the yield curve: the ultimate forward rate, the convergence speed towards the UFR, and the convexity?

The simplest model to estimate these quantities is a single factor Vasicek model. The model can be parameterized by three key parameters: the ultimate yield, mean reversion (or convergence speed) and volatility. It therefore directly addresses the main challenges for extrapolating a yield curve. The single factor model can not fit

---

[5] See the leading adjustment term I_1 in Appendix B in Christensen *et al* (2011)

the complex curvatures at the short end (1 month to 2 years) of the yield curve, but performs reasonably for long maturities. We therefore estimate the parameters using data on 5- and 20-year maturities.

In a Bayesian analysis using Euro swap rate data we find that the mean reversion of the level factor is non-zero, but with considerable probability mass very close to the unit root (under the risk neutral measure). That means that convexity effects are important, but not so large that they drive the limiting yield to minus infinity. In our extrapolations the yield curve remains upward sloping for maturities up to 100 years. As expected the Vasicek extrapolation leads to a higher level of very long-term yields than the Nelson-Siegel or Smith-Wilson methods. The uncertainty around the extrapolated yields increases with the maturity. For market conditions prevailing in the fall of 2013 the Nelson-Siegel and Smith-Wilson extrapolations are at the lower end of the 95% highest posterior density region.

2 Data

Our yield curve data consists of a monthly panel of discount rates from the website of the Bundesbank.[6] These yield curve data are constructed from Euro swap rates with maturities ranging from 1 to 50 years. The sample period is from January 2002 to September 2013 resulting in 141 data-points per maturity.

Figure 1 provides an overview of the data. The average term structure is increasing until the 20-year maturity, after which it becomes slightly downward sloping for longer maturities. The yield curve has fluctuated substantially over the twelve year sample period. The figure shows a curve from the beginning of the sample (March 2002) when the curve has a similar shape as the average, but at a 1.5% higher level. The lowest long term yields are from May 2012, where the 50 years maturity yield is just below 2%. With such dominant parallel shifts of yield curves, it will be difficult to fit a long-term common ultimate forward rate to these data. It is definitely not reached

---

[6] `http://www.bundesbank.de/Navigation/EN/Statistics/Time_series_databases`

Figure 1: Euro swap rates

In the left panel the solid blue line shows the average yield curve over the period from January 2002 – September 2013. The two red lines with crosses show the yield curves with the minimum and maximum rate at a 20 years maturity. The right panel shows the volatility of yield level, yield changes and the one month prediction errors from an AR(1) model. The horizontal axis in both panels is the maturity in years. The vertical axis unit is percent per year. The right vertical axis is for levels; the left vertical axis is for changes and prediction errors.
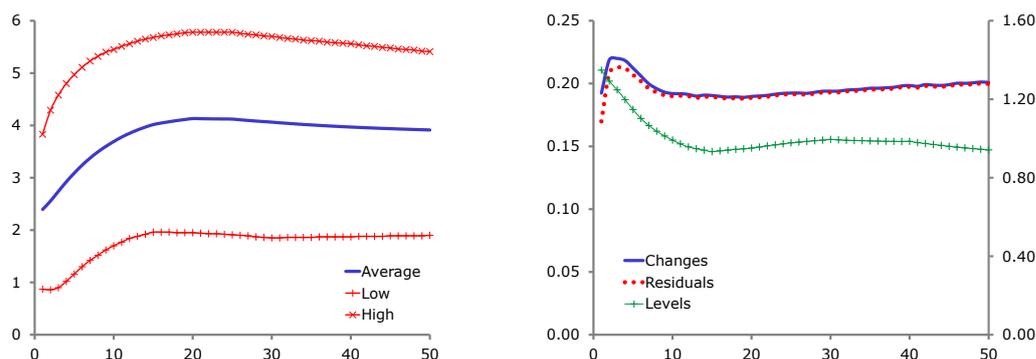

before a 50 years maturity and convergence to a common ultimate yield requires a very low level of implied mean reversion.

The right hand panel of figure 1 shows the volatility of yields. When measuring volatility as the standard deviation of yield levels we see a quickly decreasing volatility structure up to maturities of 12 years, after which the volatility stabilizes. Within an affine term structure model this points to at least two factors. One factor would be a level factor that is close to a random walk under the risk neutral measure, while the second factor is a stationary factor with strong mean reversion. Due to the strong mean reversion the second factor has a negligible influence on yields with maturities longer than 10 years. Since the shortest maturity in this data set is one year, it is difficult to identify a third factor.

The same figure also shows the volatility of yield changes. In a one-factor model it should have the same shape as the level volatility, and only the scaling should be different. In a multi-factor model the shape for short and medium-term maturities can be very different from the shape of the level volatilities. The figure shows the familiar hump-shaped volatility, where the volatility peaks at the three year maturity,

4

Figure 2: Time series data

The figure shows the monthly time series data for the 5- and 20-year maturity discount yields, both as a time series graphs and as a scatter diagram for the first differences of the yields.
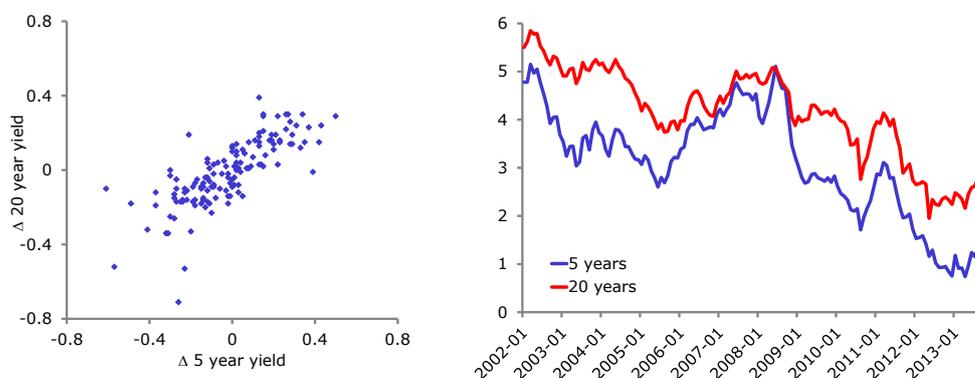
and then starts a gradual decrease. The initial hump shape for shorter to intermediate maturities can be explained by a two-factor model. The gradual downward sloping volatility is consistent with a slowly mean-reverting level factor. Most puzzling is the upward sloping pattern from a maturity of 15 years onwards, which cannot be explained by standard term structure models. The affine model, for example, implies that the volatility curve is downward sloping for longer maturities. Very long-dated swap prices may contain more noise because the market at these long-term maturities is less liquid. This suggests that the 20- years rate may be a good reference for the last liquid point to start the extrapolation.

In our econometric model we will work with time series data and the discount yields for maturities 5 and 20 years. The time series are shown in figure 2. The scatter diagram suggest that the single factor assumption is not too far off. From the time series plot it appears that both rates have a slight negative trend over the last decade, so that it may be hard to estimate an unconditional mean from these time series.

5

3   Term Structure Models

Our basic model is the essentially affine term structure model introduced by Duffee (2002) as an extension of the Duffie and Kan (1996) class of affine term structure models. Dai and Singleton (2000) show that the canonical structure of a Gaussian affine model can be written as

$$y_t(\tau) = a(\tau) + \sum_{j=1}^{K} b_j(\tau) x_{jt} \tag{1}$$

where $y_t(\tau)$ is the yield of a discount bond at time t with time to maturity τ; $a(\tau)$ and $b_j(\tau)$ are function of the time to maturity and the underlying parameters of the model; and x_{jt} are time-varying factors that follow the Ornstein-Uhlenbeck processes

$$\mathrm{d}x_j = \tilde{\kappa}_j(\tilde{\mu}_j - x_j)\mathrm{d}t + \boldsymbol{\sigma}'_j \mathrm{d}\tilde{W} \tag{2}$$

with $\tilde{W}$ a K-dimensional Brownian motion under the risk-neutral density ($\mathbb{Q}$ measure) and $\boldsymbol{\sigma}_j$ a vector of volatilities. Multiple factor term structure models are characterized by different mean reversion parameters $\tilde{\kappa}_j$ that determine the $b_j(\tau)$ functions in (1):

$$b_j(\tau) = \frac{1 - \mathrm{e}^{-\tilde{\kappa}_j \tau}}{\tilde{\kappa}_j \tau}, \tag{3}$$

The larger the $\tilde{\kappa}_j$, the less the impact of the factor on long-term yields. For typical estimates of a three factor model, more than 95% of the variation at maturities longer than 5 years is explained by the first factor, usually referred to as the level factor.

3.1   Vasicek model

Since our aim is to extrapolate the yield curve beyond maturities of 20 years, using data in the segment between 5 and 20 years, we specialize our model to a single factor. For the single factor 'Vasicek' model we drop the subscript j. Using the explicit solution for $a(\tau)$ the Vasicek yield curve takes the form

$$y_t(\tau) = \theta + b(\tau)\left(x_t - \theta\right) + \tfrac{1}{2}\omega^2 \tau b(\tau)^2 \tag{4}$$

6

where

$$b(\tau) = \frac{1 - \mathrm{e}^{-\tilde{\kappa}\tau}}{\tilde{\kappa}\tau}$$

$$\omega^2 = \frac{\sigma^2}{2\tilde{\kappa}} \tag{5}$$

$$\theta = \tilde{\mu} - \frac{\omega^2}{\tilde{\kappa}}$$

The function $b(\tau)$ defines the volatility of long-term yields relative to the level factor x_t, ω^2 is the unconditional variance of the factor, and θ is the limiting yield $y_t(\tau)$ when $\tau \to \infty$. In the Vasicek model the constant θ is both the ultimate yield as well as the ultimate forward rate. It is equal to the unconditional mean of the risk-neutral distribution of the factor minus the infinite horizon convexity adjustment. All zero rates are a weighted average of the factor and the ultimate long term yield plus a convexity adjustment.

Equation (4) provides the cross-sectional relation among yields with different maturities. For the time series dynamics of the yields we need to transform from the $\mathbb{Q}$ dynamics to the physical measure $\mathbb{P}$. This requires an assumption on the price of risk and a stochastic discount factor. Following Duffee (2002) we make the essentially affine assumption and specify the stochastic discount factor as

$$\frac{\mathrm{d}\Lambda}{\Lambda} = -x_t \mathrm{d}t - \lambda_t \mathrm{d}W \tag{6}$$

with

$$\lambda_t = \Lambda_0 + \Lambda_1 x_t \tag{7}$$

With this assumption the time series process for the factor becomes

$$\mathrm{d}x = \kappa(\mu - x)\mathrm{d}t + \sigma \mathrm{d}W \tag{8}$$

and parameters under the $\mathbb{P}$ and $\mathbb{Q}$ measures are related by

$$\tilde{\kappa} = \kappa + \sigma \Lambda_1 \tag{9}$$

$$\tilde{\mu}\tilde{\kappa} = \mu\kappa - \sigma \Lambda_0 \tag{10}$$

Combining (4) and (8) provides an expression for the time series behavior of different yields $y_t(\tau)$, which are the starting point for the econometric analysis.

7

3.2 Extrapolation

In general extrapolation uses forward rates and the identity

$$y_t(s) = \frac{1}{s} \int_0^s f_t(u) \mathrm{d}u, \tag{11}$$

with $f_t(\tau)$ the instantaneous forward rate at time t for time $t + \tau$. If we have reliable data for the term structure up to the reference maturity τ^* ('last liquid point') the extension to maturities $s > \tau^*$ follows as

$$y_t(s) = \frac{1}{s} \left(\tau^* y_t^* + \int_{\tau^*}^s f(u) \mathrm{d}u \right), \tag{12}$$

where $y_t^* = y_t(\tau^*)$ is the observed yield at maturity τ^*.[7] The extrapolation ensures continuity of the extended yield curve at the last liquid point. For the one factor Vasicek model the forward rate curve itself is a function of the single state variable x_t and we can use y_t^* to solve for the state variable using (4). Assuming that y_t^* is exactly on the Vasicek curve, the final result for the extrapolation is available in closed form by simply using (4) twice to first express $y_t(s)$ as a function of x_t and then again to replace x_t by the reference yield y_t^*. The result is

$$y_t(s) = \frac{b(s)}{b^*} y_t^* + \left(1 - \frac{b(s)}{b^*} \right) \theta + C^*(s), \tag{13}$$

with b^* a shorthand notation for $b(\tau^*)$ and where

$$C^*(s) = \tfrac{1}{2}\omega^2 b(s) \left(s b(s) - \tau^* b^* \right) \tag{14}$$

First solving for the forward rate curve and computing (12) leads to the same result as long as we use y_t^* as the only input for the extrapolation. More sophisticated methods allow for a slight measurement error in the observed rate y_t^* and extract the state variables from multiple maturities using a Kalman filter. Assuming that the measurement error is small this will only lead to minor changes in the estimate for the state variable x_t and therefore not cause major changes in the extrapolation.

---

[7] In practice most methods replace the instantaneous forward rate by one year forward rates $F_t(\tau, \tau + 1)$ in which case the integral in (12) becomes the sum $\sum_{i=\tau^*}^{s-1} F_t(i, i+1)$. For expositional purposes we stick with the instantaneous representation.

The extrapolated yields are a weighted average of the last liquid point y_t^* and the ultimate yield θ plus a convexity adjustment. The convergence speed is measured by the relative volatility $b(s)/b^*$. This is a decreasing function of s, starting at one for $s = \tau^*$ and gradually moving towards zero as s increases. The convexity adjustment $C^*(s)$ is always positive. Hence, if $y_t^* = \theta$, the extrapolation first moves $y_t(s)$ above the ultimate yield, before slowly converging downwards to the ultimate yield θ again. The Vasicek extrapolation will thus be markedly different from a simple weighted average of the last liquid point and an ultimate forward rate. It will often imply a steeper upward sloping yield curve before eventually flattening (or decreasing) towards the ultimate yield.

In general all long-term yields have a negative convexity exposure. We obtain the positive term in (14), because both the reference yield y_t^* and ultimate yield θ themselves are negatively affected by convexity and the convexity in (13) is measured relative to these yields. An equivalent expression for the extrapolated yield that explicitly shows the convexity effect in the ultimate yield is

$$y_t(s) = \frac{b(s)}{b^*} y_t^* + \left(1 - \frac{b(s)}{b^*}\right)\left(\tilde{\mu} - \frac{\omega^2}{\tilde{\kappa}}\right) + C^*(s), \qquad (15)$$

which uses (5) to replace θ by the risk neutral mean $\tilde{\mu}$ of the spot rate. The convexity in θ increases with maturity and reaches a minimum of $-\omega^2/\tilde{\kappa}$ at infinite maturity. The convexity term scales with the unconditional variance ω^2 of the risk-neutral factor dynamics. If the mean-reversion $\tilde{\kappa}$ goes to zero, meaning a true level factor, the variance will tend to infinity. But even if $\tilde{\kappa}$ does not move all the way to its limit, the convexity in the ultimate yield can be substantial due to the additional $\tilde{\kappa}$ in the numerator. For small $\tilde{\kappa}$ it can become so big that yields will be negative for large s and converge to a negative θ (with fixed $\tilde{\mu}$). This will occur, for example, in the arbitrage-free version of the Nelson-Siegel model of Christensen, Diebold and Rudebusch (2011).

The convergence towards the ultimate forward rate is measured by how quickly the forward rate converges to θ. For the Vasicek model the forward rates are given

by

$$f_t(s) = \theta + e^{-\tilde{\kappa}s}(x_t - \theta) + \omega^2 e^{-\tilde{\kappa}s} s b(s), \tag{16}$$

and thus θ is the ultimate forward rate. Using to (16) to express the forward rate $f_t(s)$ relative to the forward rate at maturity τ^* we have

$$f_t(s) = \theta + e^{\tilde{\kappa}(\tau^* - s)}(f_t^* - \theta) + \omega^2 e^{-\tilde{\kappa}s}(s b(s) - \tau^* b^*) \tag{17}$$

The mean reversion parameter $\tilde{\kappa}$ can thus alternatively be labeled as the convergence rate of the forward curve. Using the forward rate f_t^* to solve for the state variable x_t is equivalent to using y_t^*, if the Vasicek model would fit perfectly. With measurement errors the yield y_t^* will most likely provide a more accurate estimate of the factor than the forward rate f_t^*. Since the forward rate is a function of the derivative of the yield with respect to maturity, it is more sensitive to measurement error than the yield itself.

4   Econometric model

To extrapolate the yield curve we would only need the parameters $(\tilde{\kappa}, \tilde{\mu}, \sigma^2)$ of the risk neutral density $\mathbb{Q}$. These parameters can be identified from a single cross-section, but this would be very inefficient. If these parameters are time invariant, a panel estimate from multiple cross-sections increases efficiency.

Further information on the parameters σ^2 and $\tilde{\kappa}$ can be obtained from time series data, since these two parameters determine the volatility of long-term yields. Since volatility parameters are typically estimated with more precision than location parameters, time series data may be best for estimating σ^2 and $\tilde{\kappa}$. Since $\tilde{\kappa}$ determines the relative volatility of different yields through the function $b(\tau)$, we need time series data on at least two maturities to identify $\tilde{\kappa}$ from volatility moments. For the estimation we use the time series data for both the 5-year and 20-year maturity discount yields. The 5-year yield is the shortest maturity that seems uncontaminated by additional factors, while the 20-years rate is the longest one that still is on the

downward sloping part of the volatility curve in the empirical data. By choosing the two maturities relatively far apart we also include as much of the cross sectional information as possible.[8]

The parameters on the physical measure $\mathbb{P}$ are fully identified from time series data. For a one-factor model a single time series would be enough to identify the model parameters (κ, μ, σ). The parameters under the $\mathbb{P}$ and $\mathbb{Q}$ measures are connected through the price of risk function λ_t, which contains two free parameters. Hence there is an overlap between the two sets of parameters, which we indicate by including the common parameter σ^2 in both parameters sets. Since the relation between the two measures depends on the market price of risk, knowing Λ_0 and Λ_1 is equivalent to knowing $\tilde{\kappa}$ and $\tilde{\mu}$. With a single interest rate time series it is impossible to identify the cross-sectional parameters. With multiple maturities the parameters are overidentified.

Henceforth, we consider a single factor model for two maturities as following the restricted VAR(1) process

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \end{pmatrix} = \begin{pmatrix} y_{t-h}(\tau_1) \\ y_{t-h}(\tau_2) \end{pmatrix} - \alpha \begin{pmatrix} y_{t-h}(\tau_1) - m(\tau_1) \\ y_{t-h}(\tau_2) - m(\tau_2) \end{pmatrix} + \begin{pmatrix} e_t(\tau_1) \\ e_t(\tau_2) \end{pmatrix}, \qquad (18)$$

where h is the length of the time interval between two observations (one month, $h = 1/12$), $m(\tau)$ is the unconditional mean of a discount rate with maturity τ, and the shocks $e_t(\tau)$ are normally distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}$. The mean reversion parameter α is the discrete time equivalent of the continuous time mean reversion parameter κ,

$$\alpha = 1 - e^{-\kappa h} \qquad (19)$$

In this bivariate process the mean-reversion parameter α should be the same for the two different interest rates. According to (4) and (8) the error covariance matrix

---

takes the form

$$\boldsymbol{\Sigma}^* = s_h^2 \sigma^2 \begin{pmatrix} b_1^2 & b_1 b_2 \\ b_1 b_2 & b_2^2 \end{pmatrix}, \tag{20}$$

where $b_i = b(\tau_i)$ and where s_h^2 is a scaling constant that links the discrete time model to the continuous time parameterization through

$$s_h^2 = \frac{1 - e^{-2\kappa h}}{2\kappa} \tag{21}$$

In an Euler discretization we would have $\alpha = \kappa h$ and $s_h^2 = h$. Since we are using a single factor model, the matrix $\boldsymbol{\Sigma}^*$ has rank one. To avoid the stochastic singularity in the estimation it is common to assume a small measurement (or model) error. This can be done through a formal measurement equation and a Kalman filter model as in De Jong (2000). Since here we only have two time series, we take the simpler approach by adding a small positive variance to the diagonal elements of $\boldsymbol{\Sigma}^*$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^* + s_h^2 \eta^2 \boldsymbol{I} \tag{22}$$

With this specification the covariance matrix is a function of three parameters: σ^2, $\tilde{\kappa}$, and η^2. The parameter $\tilde{\kappa}$ enters through the function $b(\tau)$. The mean reversion under the risk-neutral measure is identified through the covariance matrix, since $b(\tau)$ defines the volatility of a bond with maturity τ and $\tilde{\kappa}$ is thus primarily a volatility parameter. The parameterization with η^2 interpreted as a measurement error variance will only be credible if η^2 is small relative to the overall volatility of the shocks $\boldsymbol{e}_t$. Large measurement error would not only cast doubt on the model specification, but would also imply that the regressors $y_{t-h}(\tau_i)$ are subject to errors-in-variables.

When we estimate the model we obtain estimates of the covariance matrix $\boldsymbol{\Sigma}$. From the three elements in $\boldsymbol{\Sigma}$ we solve for the three model parameters σ^2, η^2 and $\tilde{\kappa}$. Estimates of $\tilde{\kappa}$ are only admissible if the implied $\tilde{\kappa} \geq 0$. Nonnegativeness of $\tilde{\kappa}$ requires the following two conditions to hold:

$$\begin{aligned} \sigma_{11} &\geq \sigma_{22} \\ \sigma_{21} &\geq 0 \end{aligned} \tag{23}$$

12

The first condition imposes that volatility decreases monotonically with maturity. The second condition states that shocks to long-term interest rates are positively correlated. In appendix A we derive the further admissability condition

$$\sigma_{11} - \sigma_{22} \leq \frac{15}{4}\sigma_{21}, \tag{24}$$

which is specific to the maturities $\tau_1 = 5$ and $\tau_2 = 20$.

The intercepts $m(\tau_i)$ in (18) are related to the parameters μ and $\tilde{\mu}$ in the Vasicek model, or equivalently μ and the ultimate forward rate θ,

$$\begin{pmatrix} m(\tau_1) \\ m(\tau_2) \end{pmatrix} = \begin{pmatrix} b_1 & 1 - b_1 \\ b_2 & 1 - b_2 \end{pmatrix} \begin{pmatrix} \mu \\ \theta \end{pmatrix} + \frac{\sigma^2}{4\tilde{\kappa}} \begin{pmatrix} \tau_1 b_1^2 \\ \tau_2 b_2^2 \end{pmatrix} \tag{25}$$

Since $\tilde{\kappa}$ and σ^2 are already identified from the covariance matrix $\boldsymbol{\Sigma}$, these two equations uniquely identify μ and θ. Inverting (25) is problematic when $\tilde{\kappa} \to 0$. For small values of $\tilde{\kappa}$ the system becomes almost singular because $b_1 \to b_2$, while at the same time the intercepts go to infinity (unless $\sigma^2 \to 0$). The ultimate yield may therefore be very difficult to identify from the data if the risk-neutral mean reversion is small.

Assuming normality and time series independence for the error terms, we obtain a normal likelihood function from which we can estimate the six unknown parameters κ, $\tilde{\kappa}$, μ, $\tilde{\mu}$, σ^2 and η^2. The parameters are also exactly identified from the reduced form parameters α, $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$.

5   Maximum Likelihood estimates

For a first formal evaluation of the model we estimate the parameters by conditional maximum likelihood, where we condition on the initial observation. The results are in table 1.

In contrast to many other studies the model is estimated with data on medium- to long-term maturities. Results are similar, however, to what has been found for other sample periods and countries. The time series mean reversion κ, corresponding to a monthly first order autocorrelation of 0.975, is not significantly different from zero.

Table 1: ML parameter estimates

The table reports conditional Maximum Likelihood estimates for the bivariate model using interest rates with maturities 5 and 20 years. The asymptotic standard errors ('se') are from the Hessian of the log-likelihood function. The columns on the right side of the table are estimated under the restriction $\tilde{\kappa} = \kappa$. All parameters are reported in their natural units with time measured in years. The last line is the value of the log likelihood function.

	Estimate	se	Estimate	se
$\tilde{\kappa}$	0.0202	0.0101	0.0210	0.0103
κ	0.3023	0.1685	$= \tilde{\kappa}$	
$\tilde{\mu}$	0.1338	0.2677	0.2034	0.4720
μ	0.0155	0.0102	-0.1317	0.1392
θ	0.0717	0.0324	-0.1090	0.4541
Λ_0	0.2940	0.7668	-1.0199	1.5508
Λ_1	-40.556	24.157	0	
σ^2	4.710×10^{-5}	1.313×10^{-5}	4.744×10^{-5}	8.374×10^{-6}
ω^2	1.190×10^{-3}	4.928×10^{-4}	1.113×10^{-3}	4.568×10^{-4}
η^2	1.099×10^{-5}	1.315×10^{-6}	1.093×10^{-5}	1.306×10^{-6}
$\ln L$	1665.10		1663.67	

A unit root can not be rejected. We also find, like *e.g.* De Jong (2000), that the risk neutral mean reversion parameter is much smaller than the time series mean reversion. It is also not significantly different from zero, even though the asymptotic standard error is very small. Due to the near unit root for the time series, the estimate of the unconditional mean μ at the short end of the yield curve is very imprecise. With a standard error of around 1% it is difficult to anchor an average interest rate. The long-term yield θ is even more imprecise, partly because of the same unit root problem, and partly because of the additional uncertainty on the long-term convexity in the term structure. The implied parameters for the price of risk are both insignificant. The imprecision in Λ_0 and Λ_1 can be reduced by setting $\Lambda_1 = 0$, which would also imply that $\tilde{\kappa} = \kappa$. Since $\Lambda_1 = (\tilde{\kappa} - \kappa)/\sigma$ the insignificance of Λ_1 suggests that this restriction can not be rejected.

Re-estimating under the restriction $\tilde{\kappa} = \kappa$ results in the same estimate for mean reversion for the $\mathbb{Q}$ dynamics. Therefore the time series mean reversion κ is much closer to the unit root than in the unrestricted case. Consistent with the estimates

of the unrestricted model, the likelihood ratio statistic shows that the restriction can not be rejected. A consequence of moving much closer to the unit root under the time series measure $\mathbb{P}$ is that the unconditional means are much harder to estimate and in fact almost unidentified. The ultimate yield θ becomes negative with a much larger asymptotic standard error than in the unrestricted model where κ is much larger.[9] The restriction does not help in learning about the shape of the yield curve at very long maturities.

The model error variance is so small that results can not be affected by errors-in-variables problems. Assuming the measurement error to uncorrelated over time, the measurement error in y_{t-h} is of the order $\frac{1}{2}s_h^2\eta^2 \approx 5 \times 10^{-7}$, which is negligible relative to its true variance $b_i^2\omega^2 \approx 10^{-3}$.

Statistically it is impossible to distinguish between the two sets of parameter estimates. Implications of the parameters are, however, very different due to the differences in the estimate of the ultimate yield. One element on which both estimates agree is the convexity of the yield curve. The parameters $\tilde{\kappa}$ and σ^2 are both estimated with large precision and are independent of the estimates of the unconditional means.

Our estimated cross-sectional mean reversion $\tilde{\kappa} = 0.02$ implies a convergence of forward rates towards the ultimate forward rate θ that is much slower than the convergence assumed in the Smith-Wilson methodology adopted by EIOPA, which for most periods will be based on a convergence rate of 0.10.

6   Bayesian analysis

Since it is hard to decide which parameters to use, and since the unconditional mean parameters are also highly uncertain, we turn to a Bayesian analysis of the model. This provides a way to account for the parameter uncertainty by computing the extrapolation as a weighted average of different sets of parameters with weights given

---

[9] Standard errors reported in the table are from the Hessian of the log-likelihood function. Robust standard errors allowing for heteroskedasticity and non-normality do not make a difference to the conclusions in this case.

by the posterior density of the parameters. The Bayesian analysis also allows for informative priors, by which we can impose stationarity of the dynamics under both $\mathbb{P}$ and $\mathbb{Q}$ and add a prior view on the unconditional mean of the interest rates that we analyze.

6.1 Priors

We will use mildly informative priors on the long-term means of the interest rate data and on the mean reversion parameters under $\mathbb{P}$ and $\mathbb{Q}$. We require all these parameters to be non-negative. For the time series mean reversion we use a truncated normal prior

$$p(\alpha) \sim \text{TN}(h\mu_a, h^2\psi_a^2) \tag{26}$$

with $\mu_a = 0, \psi_a = 0.2$, such that the truncation implies a prior mean and standard deviation of $\text{E}[\alpha] = 0.16h$ and $\text{s}[\alpha] = 0.12h$, respectively. With monthly data the prior is centered around a first order autocorrelation of 0.987. The prior is centered close to a unit root, but the relatively tight precision also ensures that the posterior will be away from the unit unless the data are very informative on the dynamics.

To impose that the long term means $m_i = m(\tau_i)$ of the interest rates are positive we also use a truncated normal. We assume independent priors for m_1 and m_2 that are both specified as

$$p(m_i) \sim \text{TN}\left(\mu_m, h_m^2\right) \tag{27}$$

with $\mu_m = -0.923$ and $h_m^2 = 0.2$ implying $\text{E}[m_i] = 0.04$ and $\text{s}[m_i] = 0.039$. The prior ensures that the unconditional means are positive at maturities τ_1 and τ_2, but it does not guarantee that the unconditional mean is positive for all maturities. Most problematic could be the ultimate yield θ, since it is extremely sensitive to a near unit root in the risk-neutral process.[10]

For the covariance matrix $\boldsymbol{\Sigma}$ we assume a truncated inverted Wishart distribution.

$$p\left(\boldsymbol{\Sigma}^{-1}\right) \sim \text{TW}(\boldsymbol{\Psi}, \nu) \tag{28}$$

---

[10] Imposing $\theta > 0$ introduces a highly nonlinear dependence of $\boldsymbol{m} = (m_i \ m_2)'$ on $\boldsymbol{\Sigma}$, which complicates the numerical analysis.

where

$$\boldsymbol{\Psi} = 0.01^2 \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}$$

and the degrees of freedom parameter is set to $\nu = 3$, which is slightly above the minimum value of 2. The prior is truncated to the region that satisfies the inequalities (23) and (24). The prior for $\tilde{\kappa}$ is implicit in the prior for $\boldsymbol{\Sigma}$. Even though the prior is almost non-informative for $\boldsymbol{\Sigma}$, the prior for $\tilde{\kappa}$ is mildly informative, since it only depends on the ratio $S = \frac{\sigma_{11} - \sigma_{22}}{\sigma_{21}}$. Accounting for the truncation by the inequality constraints, the prior for $\tilde{\kappa}$ has a mean of 0.033 and standard deviation of 0.049 (based on a numerical evaluation).

Since all priors are proper and have well-defined means and variances for the reduced form parameters, the posterior moments for the reduced form parameters also exist. The posterior is not available in closed form due to the truncation and the non-linear parameterization involving the product $\alpha \boldsymbol{m}$. Numerically the posterior can be easily obtained through Gibbs sampling, since all conditional posteriors are straight-forward except for the truncation. When we sample from the conditional posteriors we reject a draw if it is outside the admissible region. In some cases the probability of accepting a draw can be extremely low. This happens when we need to draw the unconditional means $\boldsymbol{m} = (m_1 \ m_2)'$ at a point where the mean reversion parameter α is close to zero. In this case the data are uninformative about the unconditional mean, meaning that we need to draw $\boldsymbol{m}$ from a distribution that is approximately equal to the prior. Since this is a truncated normal with a negative mean, the probability of obtaining a positive number by drawing from a normal distribution becomes very small. For small α we therefore use the exponential rejection sampling algorithm suggested by Geweke (1991).

6.2 Posterior densities

Results of the Bayesian analysis are reported in table 2. Posterior moments are based on 1 million draws from the MCMC sampler.

Table 2: Posterior moments of parameters

|	Mean	St. Dev.	HPD95 lb	HPD95 ub
$\tilde{\kappa}$	0.0203	0.0096	1.647×10^{-3}	0.0382
κ	0.1687	0.0915	7.534×10^{-7}	0.3332
$\tilde{\mu}$	0.2017	1.2164	3.149×10^{-2}	0.4101
μ	0.0139	0.0044	7.232×10^{-3}	0.0212
θ	-7.313	473.8	-0.4038	0.2106
Λ_0	-0.0007	0.2388	-0.4957	0.4881
Λ_1	-21.6018	13.5835	-47.0256	2.7775
σ^2	4.850×10^{-5}	8.395×10^{-6}	3.301×10^{-5}	6.520×10^{-5}
ω^2	0.0021	0.0119	4.9004×10^{-4}	0.0039
η^2	1.086×10^{-5}	1.300×10^{-6}	8.449×10^{-6}	1.347×10^{-5}

The table shows the posterior means, standard deviations and 95% highest posterior density intervals for the parameters of the Vasicek term structure model estimated on interest rates with maturities $\tau_1 = 5$ and $\tau_2 = 20$. Results are based on one million draws from the Gibbs sampler for the parameters α, $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$, which are solved for the structural parameters in the table.

The posterior moments for κ and $\tilde{\kappa}$ are close to the maximum likelihood estimates. Due to the prior specification the posterior mean for the time series mean reversion is a bit closer to the unit root and also a bit more precise. Similar to the ML estimates the mean reversion under $\mathbb{P}$ is still substantially larger than under $\mathbb{Q}$. The risk parameter Λ_1 provides a direct comparison on the equality of the two parameters and, as for the ML estimates, the 95% credible interval for Λ_1 contains zero.

Most of the differences with the ML results are in the unconditional means. Although the prior imposes that the unconditional means of the 5- and 20-year yields exist and are positive, this does not guarantee that the unconditional means at other maturities are also positive. The ultimate yield θ is hard to identify from the data. Its posterior mean is negative with a huge standard deviation that is due to a few extremely negative outliers when $\tilde{\kappa}$ is close to zero. Figure 3 shows a scatter plot of the posterior draws for θ conditional on $\tilde{\kappa}$. The left-side of the figures zooms in on the 1% smallest draws for θ. All of these occur conditional on very small values for $\tilde{\kappa} < 10^{-4}$. Given the large uncertainty on the ultimate yield, it is doubtful if the posterior mean of θ exists with our prior specification. A clear sign that the posterior

Figure 3: Conditional posterior draws of θ given $\tilde{\kappa}$

The figure shows a scatter plot of the draws $\theta^{(i)}$ conditional on $\tilde{\kappa}^{(i)}$. The left-hand panel only shows the 1% smallest values of $\theta^{(i)}$. The right-hand panel shows all draws. The vertical axis is in natural units: -2 therefore means -200%.
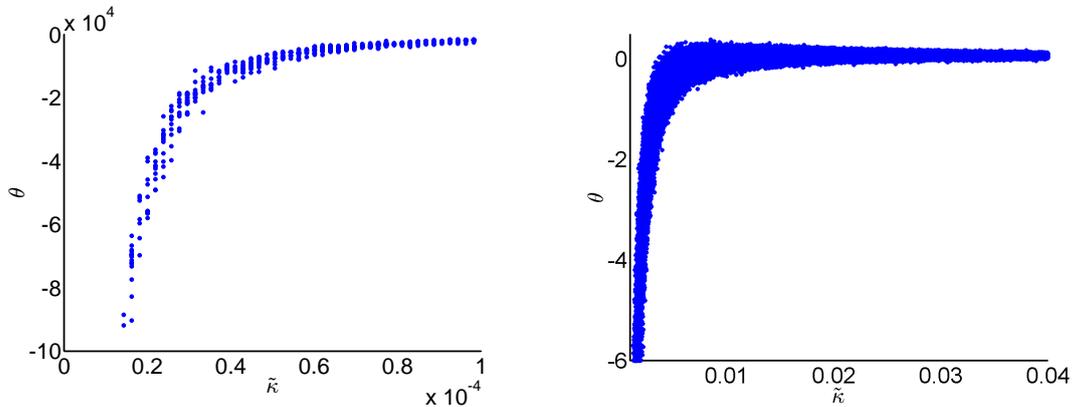
mean may not exist is the fact the average of the simulated $\theta^{(i)}$ is far below the lower bound of the 95% HPD. The posterior density is extremely skewed with a very long left tail.[11]

In contrast to the ultimate yield, the posterior on the unconditional variance under $\mathbb{Q}$ is well-behaved. It does have a fat right tail, but the posterior simulation does not produce any of the severe outliers that we encountered for θ. The unconditional variance depends on $\tilde{\kappa}^{-1}$, whereas the ultimate yield depends on $\tilde{\kappa}^{-2}$.
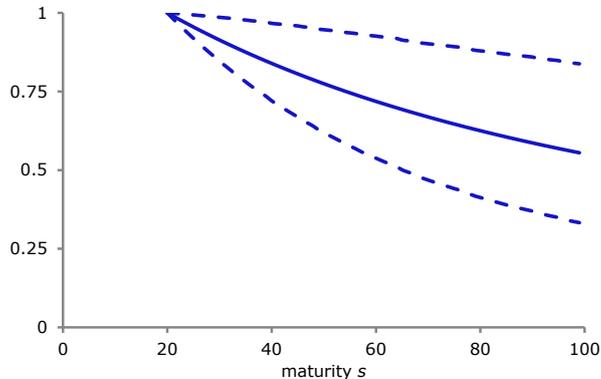
7   Extrapolation results

For the extrapolation we set the reference maturity as $\tau^* = 20$ years. This is the longest maturity in our model and is also the choice of the 'last liquid point' made by

---

[11] The parameter $\tilde{\kappa}$ depends on the error covariance matrix $\boldsymbol{\Sigma}$, which follows a Wishart distribution. Since we have sufficiently many time series observations, low order moments of $\boldsymbol{\Sigma}$ clearly exist. But $\tilde{\kappa}$ is an implicit non-linear function of $\boldsymbol{\Sigma}$ and therefore its properties can not be determined analytically. What matters for the existence of the mean of θ are the properties of the ratio $\sigma^2/\tilde{\kappa}^2$, which is even more complicated, since σ^2 and $\tilde{\kappa}$ are dependent functions of the same matrix $\boldsymbol{\Sigma}$.

Figure 4: Convergence speed

The figure shows the posterior mean of $b(s)/b^*$ for different values of $s > \tau^*$. The dashed lines define the 95% HPD interval.

EIOPA. Below we discuss our Vasicek extrapolation results and compare them with the Smith-Wilson methodology adopted by EIOPA and an extrapolation based on the Nelson-Siegel model.

7.1 Vasicek extrapolation

The extrapolation formula (13) consists of three terms. The ratio $b(s)/b^*$ in the first term defines the dependence of the extrapolated yield on the last liquid yield y_t^* and is solely a function of $\tilde{\kappa}$. Figure 4 shows the posterior of this ratio as a function of s. At $s = \tau^*$ the ratio is equal to one by construction and therefore not subject to any uncertainty. The posterior mean of $b(s)/b^*$ is decreasing in s. The uncertainty in the ratio increases with s, however. The distribution is right skewed for large s, meaning that there is considerable probability mass for $b(s)/b^*$ remaining very close to one at long maturities.

The second term is the effect of the ultimate yield. It is the product of θ and the weight of the ultimate yield. The parameter θ is very poorly determined by the data, but the large negative outliers for θ occur when $\tilde{\kappa}$ is very close to zero. For small $\tilde{\kappa}$ the weight $(1 - b(s)/b^*)$ will also go to zero, so that the overall effect is unclear. Figure 5

Figure 5: Extrapolation towards the ultimate yield

The left panel shows the posterior mean of $(1 - b(s)/b^*)\theta$ for $s > \tau^* = 20$. The left panel adds 95% highest posterior density region.
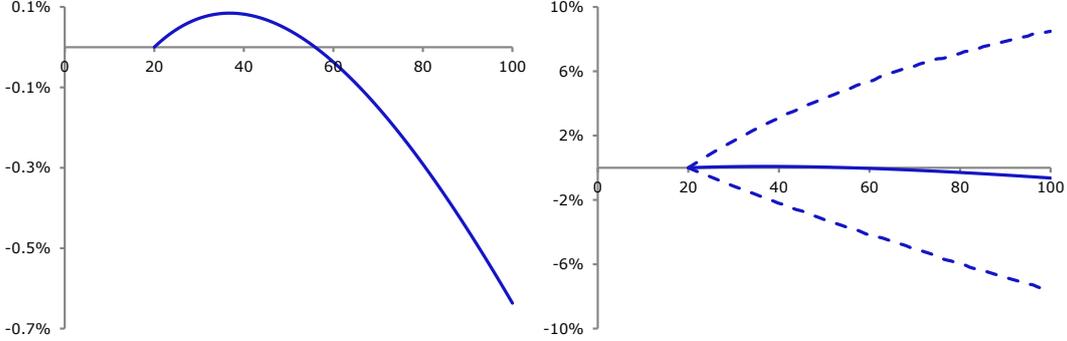
shows the posterior mean of the second term. Despite the outliers in θ the product $(1 - b(s)/b^*)\theta$ has a very small posterior mean. The total effect of θ on yields up to maturity of 60 years is less than 10 basis points on the overall extrapolated yields $y_t(s)$. The effect becomes negative for longer maturities. Combining the results in figures 4 and 5 implies that the weighted average of the reference yield y_t^* and the ultimate yield will thus generally be below y_t^*.

The right-hand panel in figure 5 adds the 95% HPD bounds to the posterior mean. The scaling is very different. Because the bounds are wide, the posterior mean looks like a flat line around zero. Still the range of uncertainty in $(1 - b(s)/b^*)\theta$ is much smaller than the 95% HPD region for θ itself (see table 2). The interval is also much more symmetric, showing once more that θ and $\tilde{\kappa}$ are highly correlated in the tails.

The convexity term $C^*(\tau)$ adds positively to the extrapolated yield. Figure 6 shows the posterior mean of $C^*(s)$ for $\tau^* = 20$. Therefore, convexity at $\tau^* = 20$ is zero by construction and positive for all $s > \tau^*$. The posterior mean increases over the entire range to the maturity of 100 years (even though it must decrease to zero by construction as $s \to \infty$). At the 60 years maturity it will contribute about 2% to the yield curve on top of the weighted average of y_t^* and θ; at 100 years the effect increases to 4%. As with the ultimate yield, the uncertainty in this term is large. At the lower end of the 95% HPD region the convexity effect is negligible. Small

Figure 6: Convexity

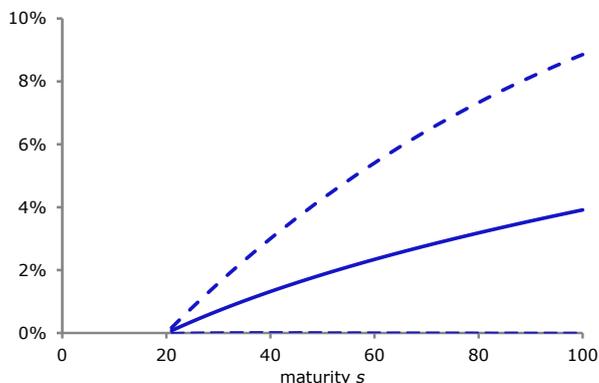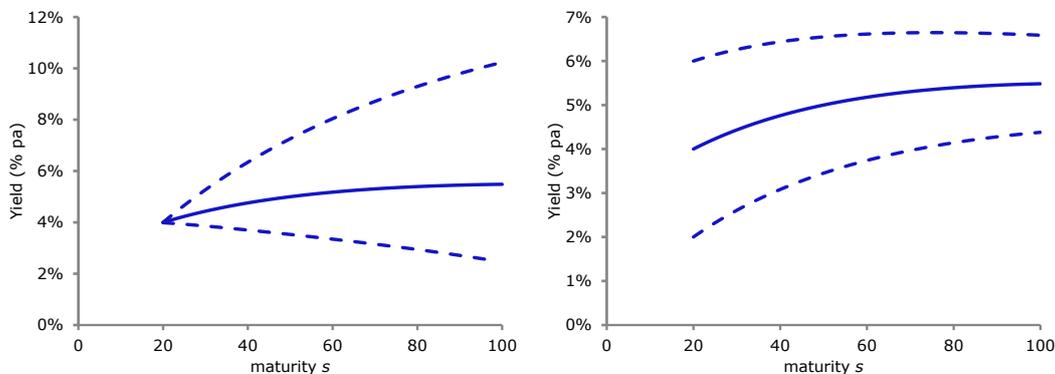The figure shows the posterior mean and 95% HPD region for the convexity term $C^*(s)$ in (14).

Figure 7: Posterior extrapolation $y_t(s)$

The left panel shows the posterior mean of $y_t(s)$ for $s > \tau^* = 20$ given $y_t^* = 4\%$. The dashed lines define the 95% highest posterior density region. The right panel shows the posterior mean of the extrapolated term structure for different last liquid points y_t^*, being 2% (dashed), 4% (solid) and 6% (dashed) respectively.

convexity effects coincide with relatively large values of $\tilde{\kappa}$. In these cases the ultimate yield will contribute positively to the extrapolation.

Taking all terms in (13) together, figure 7 shows the posterior mean and the 95% HPD interval for the extrapolated yield curve conditional on a 20-years rate equal to $y_t^* = 4\%$. The posterior mean shows a relatively flat yield curve at 4%. In the posterior mean the downward effect of the ultimate yield and the upward convexity

effect balance each other. For small $\tilde{\kappa}$ we will often have a very negative θ pushing long-term yields downward, and at the same time a very large positive $C^*(s)$ pulling yields upwards. For large $\tilde{\kappa}$ the opposite happens: negligible convexity and a large θ with substantial weight.

The error bounds show the uncertainty in the extrapolation. The bounds ar wide for practical purposes: at 60 years maturity the 95% region covers an interval from 3.5% to 8%. The uncertainty is the joint effect of the uncertainty in all parameters that enter the extrapolation formula. The basic uncertainty relates to $\tilde{\kappa}$, which has strong effects on both θ and ω^2 and the weights of these terms.

The extrapolation is similar for different values of y_t^*. The curves in figure 7 show the posterior means for three different values of y_t^*. The extrapolated curves are almost parallel consistent with a very small convergence rate. Even at the 100 year maturity the curves are still far apart. All curves are slowly upward sloping because of the strong convexity effect (relative to θ).

7.2 Alternative extrapolations

For pension funds and insurance companies recent developments about pricing of long-term obligations is under debate. In some countries the UFR is applied by central banks as explained in Solvency II. We apply the Smith-Wilson smoothing technique using Thomas and Maré (2007) and implementation notes from Norway (2010)[12] to the swap curve input data with an UFR of 4.2%, a last liquid point of 20 years and the aim of reaching the UFR in 60 years from now by approaching it by a deviation of at most 3 basis points.

To implement the Nelson-Siegel model we run a cross-sectional regression for each month to estimate the factors x_{it} in the model

$$y_t(\tau) = x_{1t} + x_{2t}\frac{1 - \mathrm{e}^{-\lambda\tau}}{\lambda\tau} + x_{3t}\left(\frac{1 - \mathrm{e}^{-\lambda\tau}}{\lambda\tau} - \mathrm{e}^{-\lambda\tau}\right) + \epsilon_t(\tau) \qquad (29)$$

---

[12] Financial Supervisory Authority of Norway: `www.finanstilsynet.no/Global/Forsikring` `%20og%20pensjon/Skadeforsikring/Tilsyn%20og%20overv%C3%A5king/Rapportering/A_` `Technical_Note_on_the_Smith-Wilson_Method_100701.pdf`.

Figure 8: Alternative extrapolations

The figure shows extrapolated yield curves for September 2013 using alternative extrapolation methods. The solid blue line shows the actual discount rates extracted from the swap curve. ML refers to extrapolation based on the maximum likelihood parameter estimated in the unrestricted model in table 1.
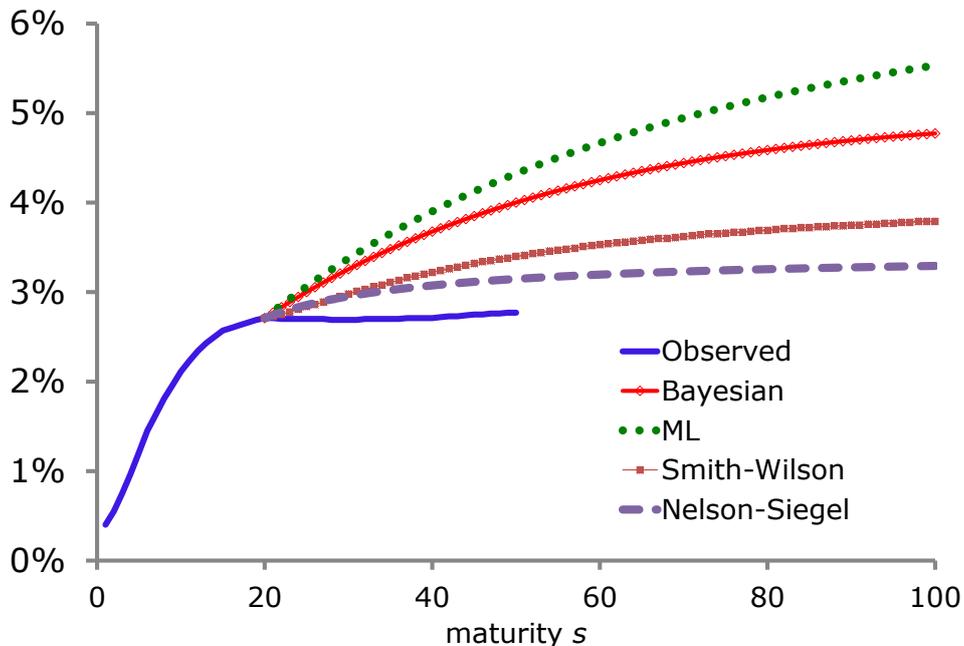
The model is estimated using data with maturities $\tau = 1, \ldots, 20$ years. We estimate the convergence parameter λ once as a constant for all months in the sample. The value $\lambda = 0.51$ minimizes the overall sum of squared errors over all maturities and months. Conditional on the estimated factors $\hat{x}_{it}$ we construct the Nelson-Siegel forward curve as

$$f_t(s) = \hat{x}_{1t} + \hat{x}_{2t}\mathrm{e}^{-\lambda s} + \hat{x}_{3t}\lambda s e^{-\lambda s} \tag{30}$$

for maturities $s > \tau^*$ and use these to extrapolate the yield curve using (12).

Figure 8 compares the different extrapolations for September 2013, the last month in our sample. Both the Smith-Wilson and Nelson-Siegel curve provide a very smooth extrapolation for maturities beyond 20 years. The Nelson-Siegel curve does not move far from the last observed yield, since the effect of the second and third factors are

Figure 9: Extrapolation Uncertainty

The figure shows extrapolated yield curves for September 2013 using alternative extrapolation methods. The solid blue line shows the actual discount rates extracted from the swap curve. The dashed lines are the p5% HPD upper and lower bounds from the Bayesian extrapolation.
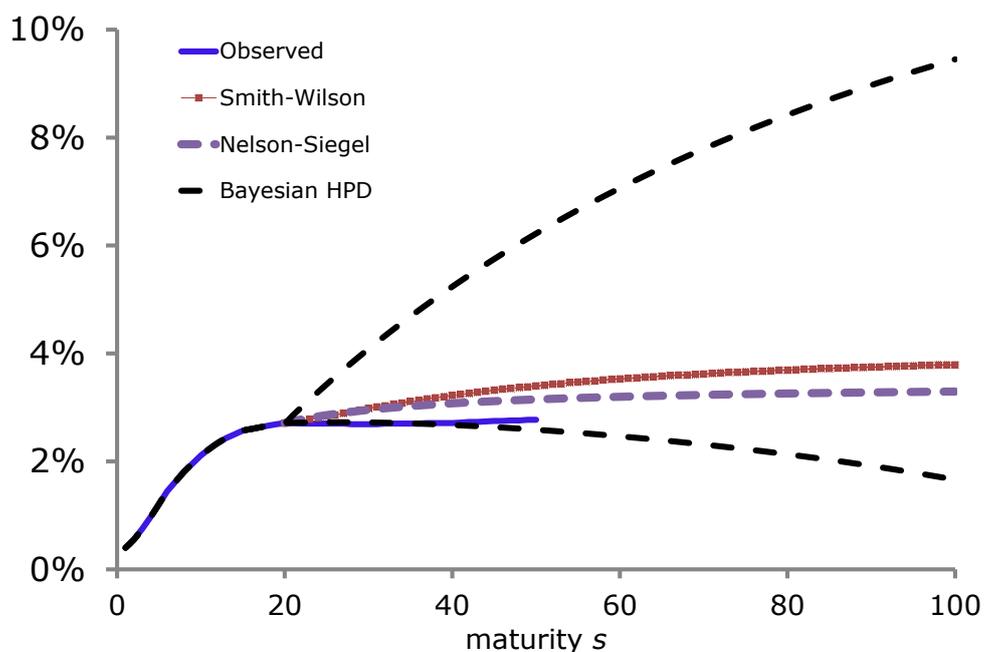
already small at these long maturities. The Smith-Wilson curve is slightly higher in order to converge to the level of 4.2%, which is above the estimated value of the Nelson-Siegel level factor for this month. The two Vasicek extrapolations are above the two alternatives. As discussed before, this is related to the convexity terms in the extrapolation. The very high level of the ML extrapolation is also due to the large, but inaccurate, estimate for the ultimate yield θ.

The Vasicek extrapolation shows a kink at the last liquid point, because it does not take into account the local curvature at this point. It just calibrates the single factor to the level of the 20-year rate for continuity. The kink can be smoothed away by using the alternative calibration of the factor by the forward rate at τ^* instead of the yield y_t^*.

Figure 9 adds 95% HPD bounds to the graph, which puts the differences in perspective. Scaling on this graph is different, because the uncertainty around the extrapolation is large. Both two alternative extrapolations as well as the observed data are within the bounds.

<div style="text-align:center">TO BE COMPLETED</div>

8 Conclusion

We extrapolated the yield curve using the Vasicek model. The Vasicek model produces extrapolated yield curves that are almost parallel and in most cases slightly upward sloping at very long maturities. The most important difference compared to alternative extrapolation techniques is the convexity effect in very long term yields. The convexity effect is an important element in no-arbitrage term structure models and can be a large component due to the slow mean reversion of the dominant interest rate level factor under the risk neutral measure. The second important element of our extrapolation method is the construction of a credible interval to quantify the uncertainty in the extrapolation.

A Parameter transformations

The bivariate restricted VAR has error covariance matrix $\boldsymbol{\Sigma}$. We solve $\tilde{\kappa}$, σ^2 and η^2 from $\boldsymbol{\Sigma}$. For $\tilde{\kappa}$ we have

$$\frac{b_1^2 - b_2^2}{b_1 b_2} = \frac{\sigma_{11} - \sigma_{22}}{\sigma_{21}} \equiv S, \tag{31}$$

where the last equivalence defines the variable S. The condition can be rewritten as

$$\frac{b_2}{b_1} = \tfrac{1}{2} \left(\sqrt{S^2 + 4} - S \right) \tag{32}$$

For $S > 0$ the right-hand side varies between 0 and 1, meaning that $b_2 < b_1$ as required by positive mean reversion $\tilde{\kappa} > 0$. Since in our application $\tau_2 = 4\tau_1$, the

left-hand side can be rewritten to obtain

$$\frac{b_2}{b_1} = \tfrac{1}{4}\left(1 + e^{-x}\right)\left(1 + e^{-2x}\right) \tag{33}$$

with $x = \tilde{\kappa}\tau_1$. For positive x this is a monotone decreasing function in x, and hence the equation has a unique solution for $\tilde{\kappa}$ if it exist. Since the left-hand side has a lower bound of $\tfrac{1}{4}$, we must also restrict the right-hand side to be above $\tfrac{1}{4}$. The largest S such that a solution exists equals $S = 3.75$. Larger values of S are not admissible.

References

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, 164(1):4–20.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, 55(5):1943–1978.

De Jong, F. (2000). Time series and cross-section information in affine term-structure models. *Journal of Business & Economic Statistics*, 18(3):300–314.

Diebold, F. and Rudebusch, G. (2013). *Yield Curve Modeling and Forecasting: the Dynamic Nelson-Siegel Approach*. Princeton University Press.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57(1):405–443.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pages 571–578.

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

Smith, A. and Wilson, T. (2001). *Fitting Yield curves with long Term Constraints*. Research Notes, Bacon and Woodrow.

Thomas, M. and Maré, E. (2007). Long term forecasting and hedging of the south african yield curve. In *2007 ASSA Convention, The Actuarial Society of South Africa*.