Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Pricing long-dated cash flows of pension funds

Determination of a good long-term interest rate for discounting very long-dated pension cash flows

Tom Kasperski

MSc 05/2020-006

Pricing long-dated cash flows of pension funds

Determination of a good long-term interest rate for discounting very long-dated pension cash flows

Maastricht University
Master thesis in Econometrics & Operations Research
Name: T.M.J. Kasperski
Student ID: i6131669
Supervisor: prof. dr. A.A.J. Pelsser
Date submitted: June 5, 2020

Abstract

Pension funds are promising very long-dated cash flows to their participants (up to 100 years), but financial instruments are only available up to 30 years. However, for calculating their funding ratio pensions funds must use a discount rate. In this paper it will be shown how we can find a good long-term interest rate for discounting very long-dated pension cash flows.

1 Introduction

The funding ratio reflects a pension fund's current financial position, indicating the ratio between available assets and liabilities. The current value of the long-dated liabilities of a pension fund is computed using the long-term interest rate. On the one hand, poor funding ratios could be caused by low levels of interest rates. As a consequence, a pension fund must take action to improve its financial position in order to pay out pension benefits to its current and future participants. A pension fund could choose to set higher contributions so it can improve its financial situation. In exceptional situations, a pension fund might have to decrease pension benefits. On the other hand, favorable funding ratios could be caused by high levels of interest rates. Then, fewer reserves are required to be hold in order to pay out future pension benefits. All in all, pension funds use the long-term interest rate to find out the amount of reserves they have to hold so they can pay out current and future pension benefits in 20, 30, 40, 50 and even more years time.

The main aim of this paper is to investigate what would be appropriate model-based yield rates to discount very long-dated liabilities.

In section 2, it will be investigated whether the systematic downward slope of yield curves is consistent with convexity in no-arbitrage term structure models. Moreover, the yield curve beyond 20 years to maturity based on the Vasicek model will be determined using maximum likelihood estimation. Then, it will be compared to the regulatory discount curve prescribed by the European Insurance and Occupational Pensions Authority (EIOPA). In section 3, the two factor Fong-Vasicek model with stochastic volatility will be defined. Next, maximum likelihood estimation will again be used to estimate the parameters of this model. These parameters will be used to determine the yield curves beyond 20 years to maturity. Furthermore, the concept of fast-scaling volatility will be introduced. Based on this concept, an asymptotic approximation of the two factor Fong-Vasicek model will be constructed. It will be shown that the approximated bond price does not depend on the volatility. Then, the corresponding yield curves will be determined. Besides, the yield curves based on all three models discussed in this paper will be compared. In section 4, a derivation of the UFR method proposed by Committee Parameters will be provided. In addition, the yield curve based on this method will be compared to the EIOPA yield curve as well as to the yield curves based on the Vasicek and Fong-Vasicek model. Next, in section 5, the present value of the pension obligations and corresponding funding ratio of the Dutch pension fund ABP in the fourth quarter of 2018 will be computed using the yield curves based on the models proposed by Vasicek, Fong-Vasicek and Committee Parameters. Section 6 wraps up this paper with a conclusion and discusses what would be appropriate model-based yield rates to discount very long-dated liabilities. In the end, four appendices can be found, which contain extensive derivations of some of the developed models and equations.

2 Determining the yield curve based on the Vasicek model

In this section, it will be investigated whether the systematic downward slope of yield curves is consistent with convexity in no-arbitrage term structure models. Evidence will be found for an 'excess' downward slope than cannot be explained by no-arbitrage models. Furthermore, a model-based extrapolation of the yield curve will be constructed and compared to the regulatory discount curve prescribed by the EIOPA. The extrapolated curve differs from the regulatory curve because of a significantly lower level of mean-reversion, which causes slower convergence towards an 'Ultimate Forward Rate', which itself is also much lower due to convexity effects. The UFR is the level to which the yield curve converges for very long maturities.

In general, the very long end of the yield curve is downward sloping. Brown and Schaefer (2000) show that a downward sloping yield curve is regarded as a convexity effect. Convexity indicates how the duration of a bond changes as the interest rate changes. It incorporates the non-linear relationship between prices and yields. Brown and Schaefer (2000) demonstrate that long-term bonds will command a lower yield than short-term bonds. Therefore, the convexity effect should be incorporated. As a consequence, investors will pay more for long-term bonds. This result can be verified by applying Jensen's inequality to the bond price formula in (80). This equation is of the form $f(x) = e^x$. Then, Jensen's inequality is used to show that $f(x) = e^x$ is a convex function:

$$\begin{aligned}\mathbb{E}[e^x] = \mathbb{E}\big[e^{x+\mathbb{E}[x]-\mathbb{E}[x]}\big] = e^{\mathbb{E}[x]}\mathbb{E}\big[e^{x-\mathbb{E}[x]}\big] &\geq e^{\mathbb{E}[x]}\big(\mathbb{E}\big[1+x-\mathbb{E}[x]\big]\big) \\ &= e^{\mathbb{E}[x]}\big(1+\mathbb{E}[x]-\mathbb{E}[x]\big) = e^{\mathbb{E}[x]}\end{aligned}$$

Therefore, (80) is convex and investors will pay more for long-term bonds, which corresponds to a lower long-term yield.

The EIOPA extrapolation method is the standard under Solvency II (a harmonised prudential framework for insurance firms). The discount curve follows swap rates until the 'last liquid' point at 20 years to maturity. The forward rate term structure is extrapolated towards a constant UFR beyond 20 years to maturity. This method is known as the Smith-Wilson extrapolation method. It results in a smooth convergence from the yield at the 'last liquid' point to a predefined UFR. The UFR and the speed of convergence to the UFR are used as input parameters (CEIOPS, 2010). The Smith-Wilson extrapolation method uses the data that is available to exactly fit bond prices where data is available and to extrapolate them by using a weighted average of the predetermined UFR and the last observable data point (CEIOPS, 2010). The main advantage of this method is its simplicity. The yield curves based on the Vasicek model differ for two main reasons from the EIOPA curve. First of all, the UFR is constant in the Vasicek model, but it has convexity and a risk premium as two

more elements. Convexity has a negative effect on the yield curve, whereas the risk premium has a positive effect on the yield curve. Moreover, a lack of mean reversion at the long end of the yield curve based on the Vasicek model is a difference with the EIOPA curve.

The data that is used in this section includes zero rates of 202 months (starting at January 2002 and ending at October 2018) with yearly maturities ranging from 1 year to maturity until 50 years to maturity. This data has been obtained from the Bundesbank.

2.1 Term structure model

2.1.1 The specification

Arbitrage-free term structure models are used to generate the true stochastic interest rate generating process by using real market data. These models are known as arbitrage-free, because they rely on the assumption that the market term structure is correct and that there are no arbitrage opportunities. There are no opportunities for arbitrage in a financial market if at least one equivalent martingale measure exists.

The aim is to construct the yield/zero curve beyond maturities of 20 years. In general, a short rate model can be used to construct the yield curve by constructing the future evolution of the short rate r_t. The short rate is the interest rate at which money can be borrowed for an infinitesimally short period of time from time t. If the evolution of the short rate is modelled as a stochastic process under risk-neutral measure $\mathbb{Q}$, the price at time t of a zero-coupon bond with maturity T can be defined. Then, the yield rates corresponding to the zero-coupon bonds form a yield curve. In this section, a single factor Vasicek model that is estimated on long-term yields will be used. A single factor model is used, because the first factor explains more than 95% of the variation at maturities longer than five years. Moreover, it results in a transparent and parsimonious model. Medium and short maturities are omitted, because this single factor model can not fit the more complicated dynamics and shapes of the medium-term and the short-term.

The yield $y_t(\tau)$ of a discount bond at time t with time to maturity τ is given by

$$y_t(\tau) = \theta + b(\kappa\tau)(r_t - \theta) + \frac{b(\kappa\tau)^2\tau\sigma^2}{4\kappa}, \tag{1}$$

where

$$b(z) = \frac{1 - e^{-z}}{z} \tag{2}$$

$$\theta = \mu - \frac{\sigma^2}{2\kappa^2} \tag{3}$$

The Vasicek model defines that r_t follows the stochastic differential equation

$$dr = \kappa(\mu - r)dt + \sigma dW, \tag{4}$$

where κ is a mean reversion parameter, μ is the unconditional mean of the level factor r under the risk-neutral measure $\mathbb{Q}$, σ is the volatility, and dW is a standard Brownian Motion. This process has the mean reverting property. Hence, the short rate will tend to move to the average short rate over time. The derivation of (1) can be found in appendix A.

The strongest convexity appears as $-\frac{\sigma^2}{2\kappa^2}$ in the UFR definition. The convexity adjustment is small when κ is large. Then, long term yields are fairly close to the ultimate yield and a weighted average of the factor r. The other way around, the convexity adjustment is large when κ is small. Alternative representations of the yield curve can be studied to look at the behavior of the curve when κ is small. An alternative representation looks like

$$dr = (\lambda - \kappa r)dt + \sigma dW, \tag{5}$$

where λ is a free parameter.
The case for $\kappa \to 0$ is known as the Merton model.
Duffee (2002) links the time series properties under risk measures $\mathbb{P}$ and $\mathbb{Q}$ by the change of measure

$$d\tilde{W} = dW + (\Lambda_0 + \Lambda_1 r)dt, \tag{6}$$

where $\tilde{W}$ is a Brownian Motion under risk measure $\mathbb{P}$. Moreover, dr under risk measure $\mathbb{P}$ is given by

$$dr = \tilde{\kappa}(\tilde{\mu} - r)dt + \sigma d\tilde{W}, \tag{7}$$

with parameters

$$\begin{aligned} \tilde{\kappa} &= \kappa + \sigma\Lambda_1 \\ \tilde{\kappa}\tilde{\mu} &= \kappa\mu - \sigma\Lambda_0 \end{aligned} \tag{8}$$

The change of measure indicates that $\Lambda_0 + \Lambda_1\tilde{\mu}$ is the average Sharpe ratio.

2.1.2 The econometric model

The parameters under the risk-neutral measure are used to determine the yield curve. These parameters are estimated from data at the longer end of the yield curve. Assuming that the parameters κ and θ are constant, time series data for the 5-year maturity as well as for the 20-year maturity discount rate are used. The 5-year discount rate is the shortest maturity

that seems to be uninfluenced by additional factors. Furthermore, the 20-year discount yield is the 'last liquid' point in the extrapolation from the EIOPA. Various maturities are needed in order to distinguish the parameters under risk measure $\mathbb{P}$ from the parameters under risk measure $\mathbb{Q}$. In order to achieve this, sufficient cross-sectional variation is needed. Therefore, two maturities relatively far apart from each other $(\tau_1, \tau_2) = (5,20)$ are chosen.
Hence, the following two restricted AR(1) processes for maturities $(\tau_1, \tau_2) = (5,20)$ are chosen:

$$y_t(\tau_i) - y_{t-h}(\tau_i) = \alpha m_i - \alpha y_{t-h}(\tau_i) + e_t(\tau_i), \tag{9}$$

where m_i is the mean of a discount yield with maturity τ_i, and $h = \frac{1}{12}$ is the length of a time interval between two observations. Moreover, the mean reversion parameter $\alpha = 1 - e^{-\tilde{\kappa}h}$ is the discrete time variant of the continuous time variant $\tilde{\kappa}$. Then, the two AR(1) processes are given by

$$y_t(5) = \alpha m_1 + e^{-\frac{1}{12}\tilde{\kappa}}\, y_{t-\frac{1}{12}}(5) + e_t(5) \tag{10}$$

$$y_t(20) = \alpha m_2 + e^{-\frac{1}{12}\tilde{\kappa}}\, y_{t-\frac{1}{12}}(20) + e_t(20) \tag{11}$$

Besides, the shocks $e_t(\tau_i)$ for $i = 1, 2$ are normally distributed with zero mean and covariance matrix Σ.

$$\Sigma = s_h^2 \sigma^2 bb' + s_h^2 \eta^2 I, \tag{12}$$

where b is a vector with elements $b_i = b(\kappa\tau_i)$ for $i = 1, 2$. Furthermore, $s_h^2 = \frac{1-e^{-2\tilde{\kappa}h}}{2\tilde{\kappa}}$ and η^2 is a small measurement error. Then, $\Sigma_{1,1}, \Sigma_{1,2}$ and $\Sigma_{2,2}$ are defined by

$$\Sigma_{1,1} = \sigma^2 \frac{1 - e^{-\frac{1}{6}\tilde{\kappa}}}{2\tilde{\kappa}} \frac{(1 - e^{-5\kappa})^2}{25\kappa^2} + \frac{1 - e^{-\frac{1}{6}\tilde{\kappa}}}{2\tilde{\kappa}} \eta^2 \tag{13}$$

$$\Sigma_{1,2} = \sigma^2 \frac{1 - e^{-\frac{1}{6}\tilde{\kappa}}}{2\tilde{\kappa}} \frac{(1 - e^{-5\kappa})(1 - e^{-20\kappa})}{100\kappa^2} \tag{14}$$

$$\Sigma_{2,2} = \sigma^2 \frac{1 - e^{-\frac{1}{6}\tilde{\kappa}}}{2\tilde{\kappa}} \frac{(1 - e^{-20\kappa})^2}{400\kappa^2} + \frac{1 - e^{-\frac{1}{6}\tilde{\kappa}}}{2\tilde{\kappa}} \eta^2 \tag{15}$$

The elements in Σ identify the model parameters κ, σ^2 and η^2.
The means m_i for $i = 1, 2$ are linked to $\tilde{\mu}$ and θ. m_i is given by

$$m_i = b(\kappa\tau_i)\tilde{\mu} + \big(1 - b(\kappa\tau_i)\big)\theta + \frac{b(\kappa\tau_i)^2\ \tau_i\ \sigma^2}{4\kappa} \tag{16}$$

Two maturities are enough to obtain $\tilde{\mu}$ and θ, because m_i can be retrieved from the data and σ^2 and κ are already identified from Σ. After, μ can be acquired, as σ^2, κ and θ are identified. Then, $\lambda = \kappa\mu$ could be obtained. Finally, Λ_0 and Λ_1 can be acquired as all other model parameters are identified. Throughout this section, time series independence within each equation and normality are assumed.

2.1.3 The parameter estimates

In this subsection, the model parameters will be estimated by maximum likelihood. Equations (10) and (11) will be modelled as a system of two seemingly unrelated regression equations. Next, the log-likelihood of this system of two seemingly unrelated regression equations looks like

$$L = -N \ \ln(2\pi) - \frac{N}{2} \ \ln|\Sigma| - \frac{1}{2}(Y - X\beta)'(\Sigma^{-1} \otimes I)(Y - X\beta) \tag{17}$$

The parameter estimates are represented in table 1.

parameter	coefficient	standard error
κ	0.00943	0.00645
θ	-0.0451	0.2918
σ^2	0.00003821	0.000009492
$\tilde{\kappa}$	0.1476	0.0896
$\tilde{\mu}$	0.006713	0.0277
η^2	0.00000846	0.00000436
λ	0.0016	0.003874
$\Lambda_0 + \Lambda_1\tilde{\mu}$	0.2487	0.3185
Log-Likelihood	2239.60	

Table 1: maximum likelihood parameter estimates including standard errors

The obtained maximum likelihood parameter estimates and corresponding standard errors are pretty close to the estimates in table 2 of Balter et al. (2019). The estimates of the mean reversion parameters are near to zero under risk measure $\mathbb{Q}$ as well as under risk measure $\mathbb{P}$ just like in Balter et al. (2019). This suggests very slow convergence towards the UFR θ. Moreover, the estimate of the UFR is poor due to its large standard error, which is also the case in Balter et al. (2019).

2.2 Extrapolation of the yield curve beyond 20 years to maturity

2.2.1 Extrapolation of the yield curve up to 50 years to maturity

Given the maximum likelihood parameter estimates obtained in the previous subsection, the yield curve is fitted for maturities longer than 20 years. A perfect fit for the level of the yield with a maturity of twenty years (the 'last liquid' point) is assumed. In this way, a

continuous extension of the observed yield curve is produced. Let τ^* be the maturity at this 'last liquid' point with $y_t^* = y_t(\tau^*)$. For maturities $s > \tau^*$, under representation λ, the very long-term yield relative to the reference rate is fitted by

$$y_t(s) = \frac{b(\kappa s)}{b(\kappa\tau^*)} y_t^* + R(s)\lambda - C(s)\sigma^2, \tag{18}$$

where

$$R(s) = \frac{1}{2}\Big[\ g(\kappa s)s - \frac{b(\kappa s)}{b(\kappa\tau^*)} g(\kappa\tau^*)\tau^*\ \Big] \tag{19}$$

$$C(s) = \frac{1}{6}\Big[\ h(\kappa s)s^2 - \frac{b(\kappa s)}{b(\kappa\tau^*)} h(\kappa\tau^*)(\tau^*)^2\ \Big] \tag{20}$$

$$b(z) = \frac{1 - e^{-z}}{z} \tag{21}$$

$$g(z) = 2\Big[\ \frac{1 - b(z)}{z}\ \Big] \tag{22}$$

$$h(z) = 6\Big[\ \frac{1 - b(z) - \frac{1}{2}b(z)^2 z}{2z^2}\ \Big] \tag{23}$$

Equations (19) and (20) remain well-defined for $\kappa \to 0$, in which they simplify to

$$R(s) = \frac{1}{2}(s - \tau^*) \tag{24}$$

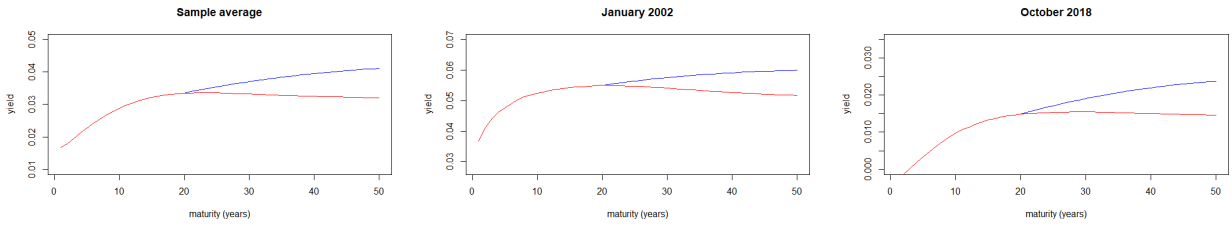
$$C(s) = \frac{1}{6}\big(s^2 - (\tau^*)^2\big) \tag{25}$$

Therefore, $y_t(s)$ for $\kappa \to 0$ is given by

$$y_t(s) = \frac{b(\kappa s)}{b(\kappa\tau^*)} y_t^* + \frac{1}{2}\lambda(s - \tau^*) - \frac{1}{6}\sigma^2\big(s^2 - (\tau^*)^2\big) \tag{26}$$

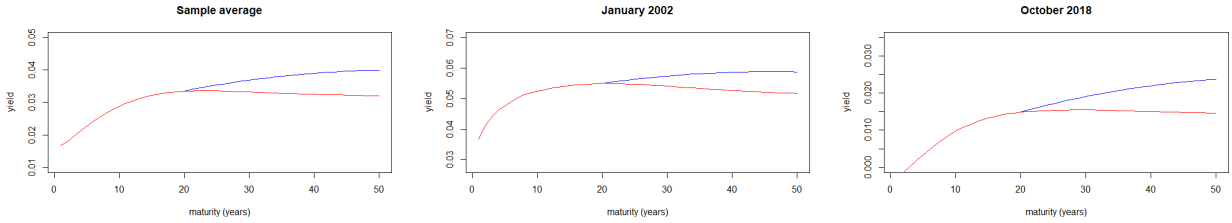
Equation (26) is quadratic in s. Hence, the prediction will create a hump-shaped yield curve with its maximum at $\tau = \frac{3}{2}\frac{\lambda}{\sigma^2} \approx 62.81$.

First, the general case is investigated, which results are represented in figure 1.
Equation (18) is used to fit the yield curves for maturities longer than 20 years. The focus will be on the sample average, January 2002 and October 2018. The maximum likelihood estimates of previous subsection are used. Next, the case where $\kappa \to 0$ is investigated, which results are represented in figure 2. Equation (26) is used to fit the yield curves for maturities longer than 20 years. The focus will again be on the sample average, January 2002 and October 2018.



(a) (b) (c)

Figure 1: fitted yield curves (blue) vs observed yield curves (red)



(a) (b) (c)

Figure 2: fitted yield curves when $\kappa \to 0$ (blue) vs observed yield curves (red)

In figure 1 and figure 2, it can be noted that the fitted values for the yield with maturity between 20 and 50 years are above the observed yield values. Furthermore, the differences between the observed yield values and the fitted yield values increase with maturity. The fitted values imply an upward slope at least up to maturities of 50 years. The downward sloping convexity effect is too weak to explain the low long-term yield rates relative to the 20-year yield rate. Moreover, it can be perceived that the yield level for October 2018 is substantially lower than the sample average yield level, while the yield level for January 2002 is considerably higher than the sample average yield level. The main reason for this is the pattern of decreasing yield rates over the years. However, the financial crisis plays a crucial role as well. As of the start of the crisis in 2007, a significant drop in the yield levels is noticeable. Besides, when comparing figure 1 and figure 2, one can conclude that the figures look very similar. Therefore, it can be concluded that the fitted yield curves for the general case are almost identical to the fitted yield curves when $\kappa \to 0$. Finally, the plots of the sample average and October 2018 of figure 1 and figure 2 look similar to figure 4 of Balter et al. (2019).

2.2.2 Convexity

The downward sloping convexity effect is too weak to explain the low long-term yield rates relative to the 20-year yield rate. Stronger convexity is needed for the model to construct lower fitted values at the longer end of the yield curve. Convexity is most powerful if no mean reversion is present. Then, the spread $s_t(s) = y_t(s) - y_t^*$ at the long end of the yield curve will follow a quadratic form. The convexity equation C(s) in (25) is quadratic. By observing that (26) is linear in its parameters λ and σ^2, these parameters can be estimated using cross-sectional regressions of $s_t(s)$ on the slope and curvature variables. Then, using spreads that correspond to the observed maturities at 5, 10, 12, 15, 20, 25, 30, 40 and 50 years result in cross-sectional regressions using eight observations. Consequently, for each time period, the estimates of the implied risk premium λ and implied volatility σ^2 that would be required to explain the convexity as observed in the data are determined. Afterwards, a time series average is computed.

parameter	coefficient	standard error
λ	0.00216352	0.00026574
σ^2	0.00009310	0.00001343

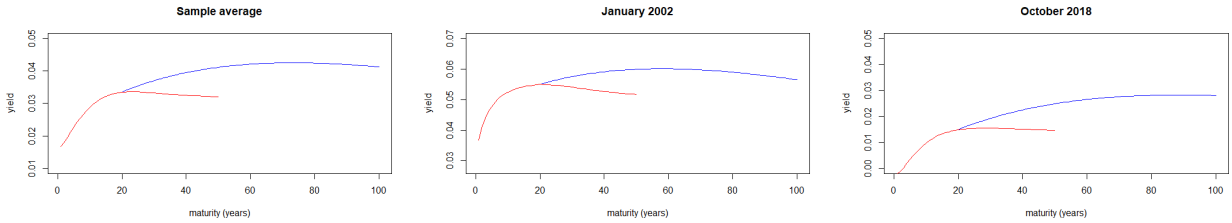
Table 2: time series average of the cross-sectional estimates of σ^2 and λ with standard errors

It can be concluded from the regression output that both coefficients are highly significant. Moreover, the estimate for λ is only a bit higher than the maximum likelihood estimate computed in the previous subsection. Therefore, the linear part is reasonably consistent. However, the σ^2 in table 2 is much larger than the σ^2 obtained in the subsection before. As a consequence, much larger volatility is needed in order to explain why yields at the very long end of the yield curve are at a lower level. To conclude, the observed low long-term interest rate levels can not be entirely explained by convexity only. Volatility has to be way higher than the σ^2 obtained in the subsection before to obtain the low long-term observed yield rates. The results of table 2 coincide with the results and conclusions from table 4 of Balter et al. (2019).

2.2.3 Extrapolation of the yield curve up to 100 years to maturity

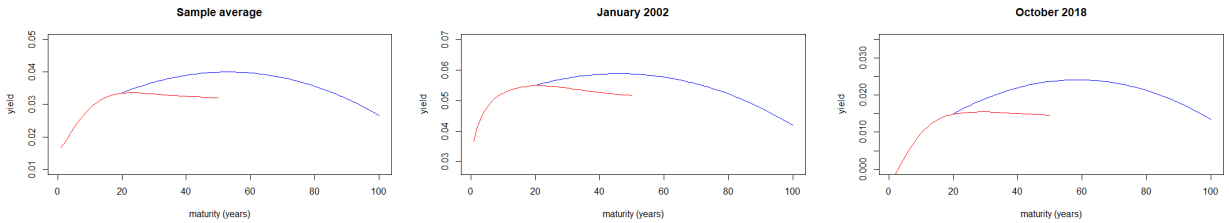
In figure 3 and figure 4, the fitted yield curves are extended to maturities very far ahead in the future up to 100 years. Maturities beyond 50 years can still be interesting for pension funds as they are promising very long-dated cash flows to their participants (up to 100 years). There, no actual data is available to compare the fitted values that have a maturity between

50 and 100 years to observed data.



(a) (b) (c)

Figure 3: fitted yield curves extrapolated up to 100 years (blue) vs observed yield curves (red)



(a) (b) (c)

Figure 4: fitted yield curves when $\kappa \to 0$ extrapolated up to 100 years (blue) vs observed yield curves (red)

Convexity dominates at maturities very far ahead in the future and causes downward sloping curves. Equation (26) is quadratic in s as mentioned earlier. Hence, the prediction will create a hump-shaped yield curve with its maximum at $\tau = \frac{3}{2}\frac{\lambda}{\sigma^2} = \frac{3}{2}\frac{0.0016}{0.00003821} \approx 62.18$. This hump-shape is clearly visible in figure 4. Furthermore, the fitted yield curves are consistent with a rather slow convergence rate. Moreover, both figures look pretty similar. However, a difference between the fitted yield curves in figure 3 and figure 4 is that the ones in figure 4 are more downward sloping for maturities very far ahead in the future than the ones in figure 3. This is in line with $\kappa \to 0$ in figure 4.

2.3 Discussion

The main purpose of this section has been to investigate the following two questions. First, are very long yield rates consistent with standard no-arbitrage term structure models? Sec-

ond, what would be appropriate model-based yield rates to discount very long-term liabilities? For the first question, it has been noticed that the one factor no-arbitrage term structure model, for which monthly zero rate data with a maturity up to 20 years was used, predicts that yield rates with a maturity between 20 and 50 years should be higher than the observed yield rates in this same period. However, although the model used in this section would not predict the behavior of these long-term yield rates, it can also not be repudiated that the observed behavior of the yield rates is consistent with a no-arbitrage term structure model. For the second question, it has been established that a pretty low level of mean reversion is present at the long end of the yield curve. Consequently, a constant UFR is impossible to estimate from long-term maturity data of yields. Although a lack of mean reversion suggests very strong convexity effects at the long end of the yield curve, the fitted yield curves are not suitable to explain the relatively low observed yield rates up to at least 50 years to maturity. However, the downward pattern is better suited to explain the relatively low observed yield rates at the very long end of the yield curve. Moreover, the model extrapolation of subsection 2.2.3 is compared to the EIOPA yield curve based on the Smith-Wilson extrapolation. The Smith-Wilson extrapolation lets the forward rate converge to the UFR θ much faster than the case is for the extrapolated yield curves based on the Vasicek model. At the beginning, the two extrapolations are not very distinct. However, eventually they will diverge completely, because of the extreme convexity at the very long end of the yield curve, which is reflected in the extrapolated yield curves based on the Vasicek model.

3 Determining the yield curve based on the Fong-Vasicek model

In this section, the two factor Fong-Vasicek model with stochastic volatility will be defined. A two factor model, for which the short rate and volatility are stochastic, is better suited to describe the behavior of markets than a one factor model, because it allows for more diversity in terms of potential outcomes than a one factor model. In short, the Fong-Vasicek model is an affine term structure model. The zero rate is a function of $A(\tau), B(\tau), C(\tau)$, the short rate $r(t)$ and the variance $y(t)$. $A(\tau), B(\tau)$ and $C(\tau)$ depend on the model parameters. Hence, if zero rates of two maturities are observed, a system of two linear equations with two unknowns can be created. As a consequence, $r(t)$ and $y(t)$ can still be obtained, despite that they are not observable in the market. In the next stage, maximum likelihood estimation can be used on the bivariate $\big(r(t), y(t)\big)$ time series to estimate the model parameters. Finally, the yield curves beyond 20 years to maturity will be determined. The data that is used in this section is similar to the data that is used in previous section. It includes zero rates of 202 months with yearly maturities ranging from 1 year to maturity until 50 years to maturity.

3.1 The Fong-Vasicek model

In section 2, the one factor Vasicek short rate model has been used implying that the yield curve $R(\tau, r)$ only depends on the evolution of the short rate. There, the short rate has been defined as a random process defined by the stochastic differential equation $dr = \kappa(\mu - r)dt + \sigma dW$. In section 2, it has been noticed that this model insufficiently captures the behavior that is detected in practice.

Longstaff and Schwarz (1992) argue that stochastic volatility allows for multiple shapes of the yield curve, which also seems to be the case when observations are looked at. Volatility of interest rates changes over time. In some periods, the market is very volatile with interest rates changing dramatically in a relatively short period. However, in other periods, the market is quite stable with only marginal changes in interest rates. Consequently, a pattern that is often observed is periods of high volatility being alternated with periods of low volatility. It is not possible to get a meaningful volatility exposure from a term structure model if volatility is a deterministic parameter. However, varying the parameter would violate the assumption of the one factor model that the parameter is constant.

The type of behavior that the volatility shows can be modelled by a random process with mean reverting property, like the short rate. Furthermore, another property of volatility is that it can not be negative. A Bessel process satisfies both properties and will therefore be used to model volatility. As a result, the stochastic volatility term structure is defined by

$$dr = \kappa_1(\theta_1 - r)dt + \sqrt{y}dZ_1^{\mathbb{P}} \tag{27}$$

$$dy = \kappa_2(\theta_2 - y)dt + v\sqrt{y}\left(\rho dZ_1^{\mathbb{P}} + \sqrt{1-\rho^2}dZ_2^{\mathbb{P}}\right), \tag{28}$$

where,

- dr: change in the short rate
- κ_1: speed of reversion to the long-term average short rate θ_1
- θ_1: long-term average short rate
- dt: change in time
- y: variance
- $dZ_1^{\mathbb{P}}$: random element under risk measure $\mathbb{P}$ that is uncorrelated with $dZ_2^{\mathbb{P}}$
- dy: change in the variance
- κ_2: speed of reversion to the long-term average variance θ_2
- θ_2: long-term average variance
- v: volatility of the variance
- $dZ_2^{\mathbb{P}}$: random element under risk measure $\mathbb{P}$ that is uncorrelated with $dZ_1^{\mathbb{P}}$
- ρ: correlation coefficient between two correlated Wiener processes

Cholesky's decomposition has been applied to define the stochastic volatility term structure in terms of two uncorrelated Wiener processes $dZ_1^{\mathbb{P}}$ and $dZ_2^{\mathbb{P}}$. Next, a risk-neutral probability measure $\mathbb{Q}$ consistent with the no-arbitrage condition is needed to convert the discounted bond prices into martingales. In order to apply this change of probability measure, a Radon-Nikodym derivative needs to be defined and the Girsanov theorem needs to be applied. The details about this change of probability measure and subsequent derivation of (29) can be found in appendix B. If $\lambda_1\sqrt{y}$ and $\lambda_2\sqrt{y}$ are the market prices of risk, then the expression for the price of a zero coupon bond $P = P(\tau, r, y)$ with time to maturity $\tau = T - t$ can be obtained by applying Ito's lemma:

$$\begin{aligned} -\frac{\partial P}{\partial \tau} + \big(\kappa_1(\theta_1 - r) - \lambda_1 y\big)\ \frac{\partial P}{\partial r} + \big(\kappa_2(\theta_2 - y) - \lambda_2 v y\big)\ \frac{\partial P}{\partial y} \\ + \frac{y}{2}\frac{\partial^2 P}{\partial r^2} + \frac{v^2 y}{2}\frac{\partial^2 P}{\partial y^2} + \rho v y \frac{\partial^2 P}{\partial r \partial y} - rP = 0 \end{aligned} \tag{29}$$

Fong and Vasicek (1991) express the solution of this PDE, taking into account the boundary condition $P(0, r, y) = 1$, as

$$P(\tau, r, y) = A(\tau)\ e^{-rB(\tau) - yC(\tau)} \tag{30}$$

The functions $A(\tau), B(\tau)$ and $C(\tau)$ are the solutions to a system of Ordinary Differential Equations obtained from (29) by plugging in a solution of this form.
This system of Ordinary Differential Equations looks like

$$A'(\tau) = -A(\tau)\Big(\kappa_1\theta_1 B(\tau) + \kappa_2\theta_2 C(\tau)\Big) \tag{31}$$

$$B'(\tau) = 1 - \kappa_1 B(\tau) \tag{32}$$

$$C'(\tau) = -\lambda_1 B(\tau) - \kappa_2 C(\tau) - \lambda_2 v C(\tau) - \frac{1}{2}B^2(\tau) - \frac{v^2}{2}C^2(\tau) - v\rho B(\tau)C(\tau), \tag{33}$$

where $A(0) = 1$ and $B(0) = C(0) = 0$, which corresponds to $P(0, r, y) = 1$. The derivation of (31), (32) and (33) can be found in appendix C. This system could be solved in the following way. First, B could be found analytically. Then, this solution is inserted into (33), which can be solved numerically. Finally, both solutions are inserted into the integrated version of (31) to find the values of A. Then, this scheme could be summarized by equations

$$B(\tau) = \frac{1}{\kappa_1}\big(1 - e^{-\kappa_1\tau}\big) \tag{34}$$

$$C'(\tau) = -\lambda_1 B(\tau) - \kappa_2 C(\tau) - \lambda_2 v C(\tau) - \frac{1}{2}B^2(\tau) - \frac{v^2}{2}C^2(\tau) - v\rho B(\tau)C(\tau), \quad C(0) = 0 \tag{35}$$

$$A(\tau) = \exp\left(-\kappa_1\theta_1 \int_0^\tau B(s)ds - \kappa_2\theta_2 \int_0^\tau C(s)ds \right) \tag{36}$$

3.2 Parameter estimates

In this subsection, maximum likelihood estimation will be used to estimate the model parameters. The starting point is (30). Then, the yield curves can be defined by

$$R(\tau, r, y) = -\frac{1}{\tau}\ \ln\big[P(\tau, r, y)\big] = -\frac{1}{\tau}\big(\ \ln\big[A(\tau)\big] - B(\tau)r(t) - C(\tau)y(t)\ \big) \tag{37}$$

Next, the monthly zero rate data is used. The two maturities $(\tau_1, \tau_2) = (5{,}20)$ are selected just like in section 2. These maturities are chosen to include as much of the cross-sectional information as possible. Then, (37) is used to create a system of two linear equations with two unknowns:

$$R(5, r, y) + \frac{1}{5}\ \ln\big[A(5)\big] = \frac{1}{5}\ B(5)r(t) + \frac{1}{5}\ C(5)y(t) \tag{38}$$

$$R(20, r, y) + \frac{1}{20}\ \ln\big[A(20)\big] = \frac{1}{20}\ B(20)r(t) + \frac{1}{20}\ C(20)y(t) \tag{39}$$

The functions $A(\tau), B(\tau)$ and $C(\tau)$ are the solutions to a system of Ordinary Differential Equations obtained from (29). Then, $A(5), B(5), C(5), A(20), B(20)$ and $C(20)$ are obtained. $R(5, r, y)$ and $R(20, r, y)$ are two vectors collected from the monthly zero rates data. Next, this system of two linear equations with two unknowns is solved for the two vectors $r(t)$ and $y(t)$. In the next stage, it is necessary to obtain the bivariate $\big(r(t), y(t)\big)$ time series. The underlying stochastic differential equations (27) and (28) of the Fong-Vasicek model need to be discretized. dr and dy are the changes in the short rate r and variance y respectively. These changes are approximated by

$$dr \approx r(t + \Delta) - r(t) \tag{40}$$

$$dy \approx y(t + \Delta) - y(t), \tag{41}$$

where $\Delta = \frac{1}{12}$. Then, it is possible to rewrite, and in this way discretize, (27) and (28):

$$r(t + \frac{1}{12}) = r(t) + \kappa_1\big(\theta_1 - r(t)\big)\frac{1}{12} + \sqrt{y(t)}\Delta Z_1(t) \tag{42}$$

$$y(t + \frac{1}{12}) = y(t) + \kappa_2\big(\theta_2 - y(t)\big)\frac{1}{12} + v\sqrt{y(t)}\Big(\rho\Delta Z_1(t) + \sqrt{1 - \rho^2}\Delta Z_2(t)\Big), \tag{43}$$

where

$$\begin{bmatrix}\Delta Z_1 \\ \Delta Z_2\end{bmatrix} \sim N\Bigg(\ \begin{bmatrix}0 \\ 0\end{bmatrix},\ \frac{1}{12}\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\ \Bigg)$$

Next, it is helpful to rewrite the discretized equations, because $\begin{bmatrix}\Delta Z_1 \\ \Delta Z_2\end{bmatrix}$ follows a bivariate normal distribution.

$$\Delta Z_1(t) = \frac{r(t+\frac{1}{12}) - r(t) - \kappa_1\big(\theta_1 - r(t)\big)\frac{1}{12}}{\sqrt{y(t)}} \qquad \text{for } t = 1, ..., T \tag{44}$$

$$\Delta Z_2(t) = \frac{y(t+\frac{1}{12}) - y(t) - \kappa_2\big(\theta_2 - y(t)\big)\frac{1}{12} - v\sqrt{y(t)}\rho\Delta Z_1(t)}{v\sqrt{y(t)}\sqrt{1-\rho^2}} \qquad \text{for } t = 1, ..., T \tag{45}$$

Then, the log-likelihood function can be defined by

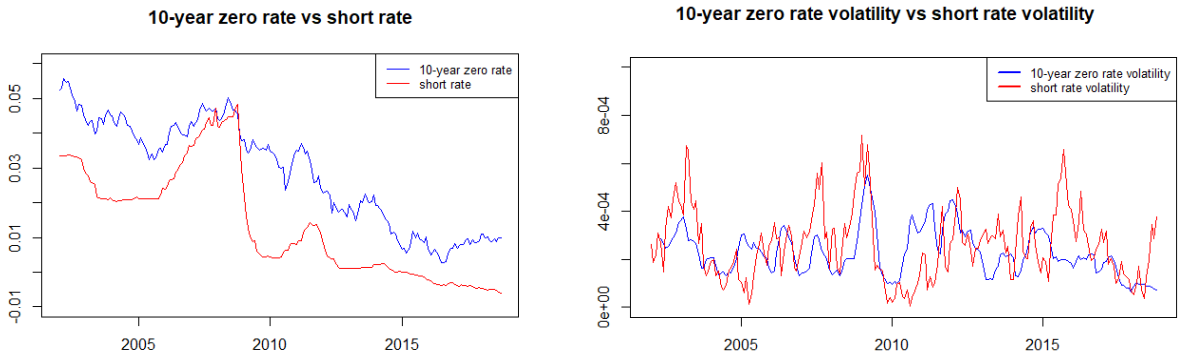
$$L = -T\ \ln\big(2\pi\big) - \frac{T}{2}\ \ln\big(\sigma^2_{\Delta Z_1}\sigma^2_{\Delta Z_2}\big) - \frac{1}{2}\left(\sum_{t=1}^{T}\frac{\big(\Delta Z_1(t)\big)^2}{\sigma^2_{\Delta Z_1}} + \sum_{t=1}^{T}\frac{\big(\Delta Z_2(t)\big)^2}{\sigma^2_{\Delta Z_2}}\right) \tag{46}$$

After, this log-likelihood function is maximized with respect to the model parameters. The maximum likelihood parameter estimates are shown in table 3.

parameter	**coefficient**	**standard error**
κ_1	0.07829	0.04473
θ_1	0.06818	0.01853
κ_2	0.05167	0.04732
θ_2	0.000314	0.000129
v	0.02052	0.01273
ρ	0.165583	0.1112
λ_1	-10.989	2.0452
λ_2	-5.1057	0.9432
Log Likelihood	3832.47	

Table 3: maximum likelihood parameter estimates including standard errors

Figure 5a depicts the observed 10-year zero rate curve versus the fitted short rate curve $\hat{r}(t)$ for $t = 1, ..., 202$. Figure 5b shows the 10-year zero rate volatility curve versus the fitted short rate volatility curve $\hat{y}(t)$ for $t = 1, ..., 202$. The 10-year zero rate volatility curve is derived by computing the moving standard deviation of the 10-year zero rate. It can be concluded that the observed 10-year zero rate curve and the fitted short rate curve roughly follow each other's pattern. Moreover, it becomes clear that the 10-year zero rate volatility curve and the fitted short rate volatility curve approximately follow each other's pattern as well. This is a positive sign with regards to the accuracy of the model parameter estimates.

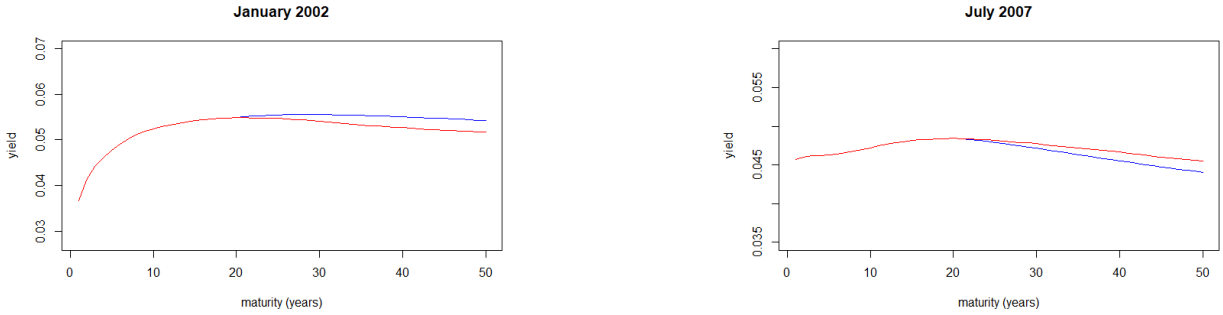


(a) (b)

Figure 5: plot (a) shows the 10-year zero rate curve (blue) versus the fitted short rate curve (red). Plot (b) shows the 10-year zero rate volatility curve (blue) versus the fitted short rate volatility curve (red)

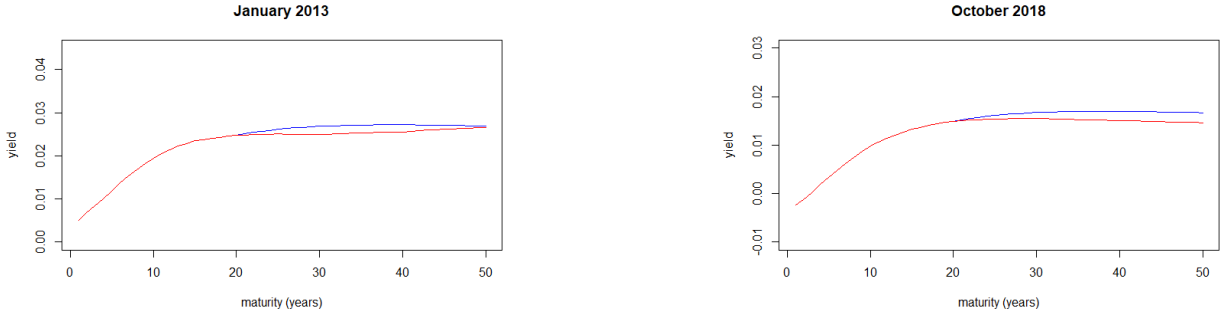
3.3 Extrapolation of the yield curve beyond twenty years to maturity

The estimated parameter values can be used to plot the fitted yield curves beyond twenty years to maturity for each month i ($i = 1, ..., 202$) against the observations of the zero rates using (37). The yield curves are fitted for maturities up to 50 years. In figure 6 and figure 7, the fitted yield curves for some months, including the months used in section 2, are plotted against the real observations. In all plots in figure 6 and figure 7, it can be observed that the differences between the observed yield values and the fitted yield values are very small. Furthermore, the downward sloping convexity effect seems pretty good suited to explain the lower long-term yield rates for January 2002, July 2007 and October 2018. The observed yield curve for January 2013 behaves a little bit strange, because the level of yield rates drops a little bit after 25 years but increases again after 30 years. All in all, the fitted yield curves based on the Fong-Vasicek model are better suited to explain the downward sloping convexity effect, and are more accurately picking up the point where the downward sloping convexity effect starts, than the fitted yield curves based on the Vasicek model. Next, in figure 8 and figure 9, the fitted yield curves for the same months as before are extended to maturities very far ahead in the future up to 100 years to maturity. There, no actual data is available to compare the fitted values that have a maturity between 50 and 100 years to observed data. Convexity is again visible in all plots and causes downward sloping fitted yield curves.



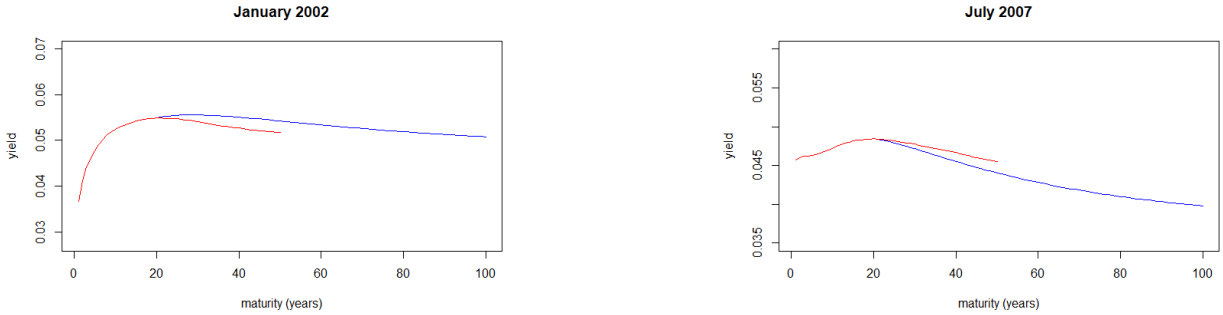
(a) (b)

Figure 6: fitted yield curves (blue) vs observed yield curves (red)



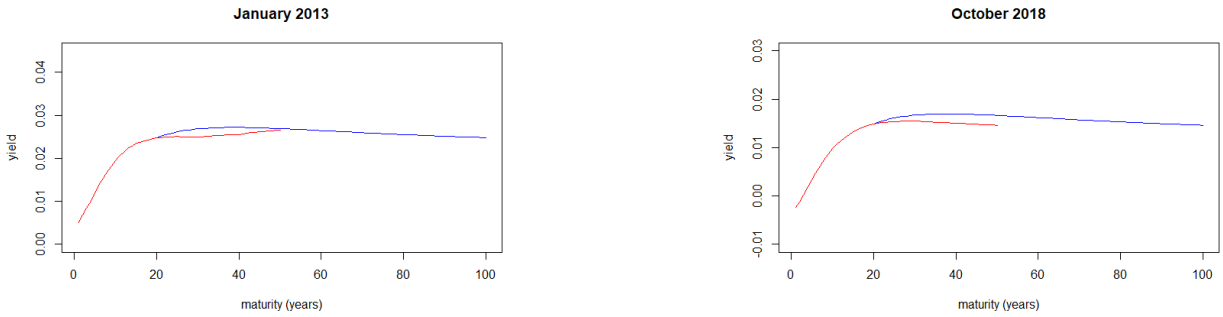
(a) (b)

Figure 7: fitted yield curves (blue) vs observed yield curves (red)



(a) (b)

Figure 8: fitted yield curves extrapolated up to 100 years (blue) vs observed yield curves (red)



(a) (b)

Figure 9: fitted yield curves extrapolated up to 100 years (blue) vs observed yield curves (red)

In addition to comparing the plots of the fitted yield curves based on the Vasicek and Fong-Vasicek model, it will be formally tested whether the Fong-Vasicek model indeed has a better goodness of fit than the Vasicek model. The likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods. This test requires that the two models are nested, which would mean in this case that the Fong-Vasicek model can be transformed into the Vasicek model by imposing constraints on the former's parameters. This can be accomplished by setting κ_2 and v equal to zero. In that case, $dy = 0$ in (28). This implies that y is constant just like in the Vasicek model. Then, (27) is left, which is the same stochastic process as the one in (7). Substituting (6) and (8) into (7) results in (4), which is the Vasicek model. Therefore, it is justified to use the likelihood ratio test. However, v will not be set exactly equal to zero, because $\Delta Z_2(t)$ in (45) is not defined for $v = 0$. A problem arises regarding the log-likelihood function of two normally distributed random variables. This can be explained by the limiting behavior of the bivariate normal distribution. In order for the bivariate normal distribution to be defined, the short rate as well as the volatility need to be stochastic. In that case, $\Delta Z_1(t)$ and $\Delta Z_2(t)$, two normally distributed random variables, are used to model the randomness for both the short rate and the volatility. The term $\left(\rho \Delta Z_1(t) + \sqrt{1-\rho^2}\Delta Z_2(t)\right)$ in (28) is used to model the randomness of dy. The volatility of the variance parameter v is related to this term. If v would reach zero, this randomness term would no longer be necessary. Consequently, a stochastic volatility process would not be necessary anymore and only $\Delta Z_1(t)$ would be needed to model the randomness. In that case, only a univariate normal distribution would be sufficient. All in all, this shows in this case that the log-likelihood function of two normally distributed random variables collapses if v would reach zero. Instead, v will be equated to 0.001. The statistic tests $H_0 : \theta \in \Theta_0$, against $H_1 : \theta \in \Theta_0^c$. Let $\Theta = \Theta_0 \cup \Theta_0^c$. The likelihood

ratio test statistic for the null hypothesis $H_0 : \theta \in \Theta_0$ is defined as

$$LRT = -2 \ \ln\left(\frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})}\right) = -2 \ \ln\left(\frac{\sup_{\theta\in\Theta_0} \ L(\theta)}{\sup_{\theta\in\Theta} \ L(\theta)}\right) \tag{47}$$

$L(\hat{\Theta}_0)$ represents the best explanation for the observed data for all $\theta \in \Theta_0$. Similarly, $L(\hat{\Theta})$ represents the best explanation for the observed data for all $\theta \in \Theta$. If $L(\hat{\Theta}_0) = L(\hat{\Theta})$, then a best explanation for the observed data can be found inside Θ_0 and we should not reject H_0. In that case, there is no evidence for the Fong-Vasicek model to have a significantly better goodness of fit than the Vasicek model. However, if $L(\hat{\Theta}_0) < L(\hat{\Theta})$, then the best explanation for the observed data could be found inside Θ_0^c, and we should consider rejecting H_0 in favor of H_1. In that case, there is evidence for the Fong-Vasicek model to have a significantly better goodness of fit than the Vasicek model. The test statistic is distributed chi-squared, with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0. Hence, the null hypothesis is rejected for values of the test statistic that exceed the $(1 - \alpha)$ quantile of the chi-squared distribution with 2 degrees of freedom. It was chosen to set α equal to 0.05. As a consequence, the corresponding critical value equals 5.991. Table 4 shows the maximum likelihood parameter estimates based on the restricted Fong-Vasicek model with $\kappa_2 = 0$ and $v = 0.001$ as well as with $\kappa_2 = 0$ and $v = 0.0001$. The restricted Fong-Vasicek model with $\kappa_2 = 0$ and $v = 0.0001$ is used as a robustness check for v. The test statistic is equal to

$$LRT = -2 \left(\ln\left(L(\hat{\Theta}_0)\right) - \ln\left(L(\hat{\Theta})\right)\right) = -2 \left(3819.26 - 3832.47\right) = 26.42$$

The value of the test statistic $26.42 > 5.991$. It can be noted that the value of the test statistic based on the restricted Fong-Vasicek model with $\kappa_2 = 0$ and $v = 0.0001$ also exceeds 5.991. Therefore, there is evidence for the Fong-Vasicek model to have a significantly better goodness of fit than the Vasicek model.

parameter	**restricted Fong-Vasicek v=0.001**	**restricted Fong-Vasicek v=0.0001**
κ_1	0.07931	0.07994
θ_1	0.06647	0.06618
κ_2	0	0
θ_2	0.000326	0.000318
v	0.001	0.0001
ρ	0.125358	0.101124
λ_1	-8.561	-8.782
λ_2	-4.6676	-4.9557
Log Likelihood	3819.26	3816.17

Table 4: parameter estimates based on the Fong-Vasicek model with $v = 0.001$ as well as $v = 0.0001$

3.4 Asymptotic approximation of the Fong-Vasicek model

In this subsection, the concept of fast-scaling volatility will be introduced. Based on this concept, an asymptotic approximation of the two factor Fong-Vasicek model will be constructed. Then, yield curves will be determined based on this asymptotic approximation of the Fong-Vasicek model. These yield curves will be compared to the ones obtained earlier in this section to see how close the yield curves based on the asymptotic approximation of the Fong-Vasicek model are to the yield curves based on the Fong-Vasicek model.

Fouque et al. (2000) have observed that the short rate process r and the variance process y operate on a different time scale. The scale of the variance process y is a lot faster. Fast scale volatility can be integrated into the model with stochastic volatility. Fouque et al. (2000) argue that the fast scale suggests a large value for κ_2 in (28). The approximation should be more accurate for large values of κ_2. Moreover, in (28) a Bessel process has been used to model volatility. Cox et al. (1985) determined the limiting distribution of y under the condition that $\frac{2\kappa_2\theta_2}{v^2} > 1$. This limiting distribution is defined as

$$f(y) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} & \text{if} \quad \text{y} > 0 \\ 0 & \text{if} \quad \text{y} \leq 0, \end{cases} \tag{48}$$

with shape parameter $\alpha = \frac{2\kappa_2\theta_2}{v^2}$, scale parameter $\beta = \frac{2\kappa_2}{v^2}$, $\mathbb{E}(\text{y}) = \frac{\alpha}{\beta} = \theta_2$ and $\text{var}(y) = \frac{\alpha}{\beta^2} = \frac{\theta_2 v^2}{2\kappa_2}$. After, Fouque et al. (2000) characterize the unit of fast scale $\epsilon = \frac{1}{\kappa_2}$. Furthermore, they suggest to utilize the fast scale of volatility to approximate two factor short rate models by a one factor short rate model plus a correction term. They prove that the price of a zero coupon bond P can be approximated by a function P^ϵ, which does not depend on the volatility y. Applying the substitution $\epsilon = \frac{1}{\kappa_2}$ to (28), while taking into account the variance of the limiting distribution of y, results in the stochastic volatility term structure specification that is defined by

$$dr = \kappa_1(\theta_1 - r)dt + \sqrt{y} dZ_1^{\mathbb{P}} \tag{49}$$

$$dy = \frac{1}{\epsilon}(\theta_2 - y)dt + \frac{1}{\sqrt{\epsilon}} v\sqrt{y} \left(\rho dZ_1^{\mathbb{P}} + \sqrt{1-\rho^2} dZ_2^{\mathbb{P}}\right), \tag{50}$$

with $\alpha = \frac{2\theta_2}{v^2}$, $\beta = \frac{2}{v^2}$ and everything else defined in the same way as in (27) and (28).
If $\lambda_1\sqrt{y}$ and $\lambda_2\sqrt{y}$ are the market prices of risk and the no-arbitrage argument result is used, then the expression for P^ϵ can be obtained by applying Ito's lemma:

$$\begin{aligned} &\frac{\partial P^\epsilon}{\partial t} + \left(\kappa_1(\theta_1 - r) - \lambda_1 y\right) \frac{\partial P^\epsilon}{\partial r} + \left(\frac{1}{\epsilon}(\theta_2 - y) - \frac{1}{\sqrt{\epsilon}}\lambda_2 v y\right) \frac{\partial P^\epsilon}{\partial y} \\ &\quad + \frac{y}{2}\frac{\partial^2 P^\epsilon}{\partial r^2} + \frac{1}{\epsilon}\frac{v^2 y}{2}\frac{\partial^2 P^\epsilon}{\partial y^2} + \frac{1}{\sqrt{\epsilon}}\rho v y \frac{\partial^2 P^\epsilon}{\partial r \partial y} - r P^\epsilon = 0, \end{aligned} \tag{51}$$

where $P^\epsilon(T,r,y) = 1$ is the boundary condition. This partial differential equation can be rewritten as

$$\Big(\frac{1}{\epsilon}\Pi_0 + \frac{1}{\sqrt{\epsilon}}\Pi_1 + \Pi_2\Big)P^\epsilon = 0, \tag{52}$$

with

$$\Pi_0 = (\theta_2 - y)\frac{\partial}{\partial y} + \frac{v^2 y}{2}\frac{\partial^2}{\partial y^2} \tag{53}$$

$$\Pi_1 = -\lambda_2 v y \frac{\partial}{\partial y} + \rho v y \frac{\partial^2}{\partial r \partial y} \tag{54}$$

$$\Pi_2 = \frac{\partial}{\partial t} + \big(\kappa_1(\theta_1 - r) - \lambda_1 y\big)\frac{\partial}{\partial r} + \frac{y}{2}\frac{\partial^2}{\partial r^2} - r \tag{55}$$

Π_0 only contains partial derivatives with respect to y, while Π_1 includes partial derivatives with respect to y and both r and y. Then, Fouque et al. (2000) rewrite P^ϵ in terms of an infinite sum, an asymptotic series based on powers of $\sqrt{\epsilon}$:

$$P^\epsilon = P_0 + \sqrt{\epsilon}P_1 + \epsilon P_2 + \epsilon\sqrt{\epsilon}P_3 + \epsilon^2 P_4 + ..., \tag{56}$$

where $P_0, P_1, ...$ are functions of t, r and y. Besides, $P_0(T,r,y) = 1$ and $P_i(T,r,y) = 0$ for all $i \geq 1$. Consequently, (52) can be rewritten as

$$\begin{aligned}\frac{1}{\epsilon}\Pi_0 P_0 + \frac{1}{\sqrt{\epsilon}}(\Pi_0 P_1 + \Pi_1 P_0) + (\Pi_0 P_2 + \Pi_1 P_1 + \Pi_2 P_0) \\ +\sqrt{\epsilon}(\Pi_0 P_3 + \Pi_1 P_2 + \Pi_2 P_1) + ... = 0\end{aligned} \tag{57}$$

Since this is an approximation, the solution procedure consists of gradually approximating solutions. More precisely, in the first step, the part of (57) belonging to $\frac{1}{\epsilon}$ will be set equal to zero. In the second step, the part of (57) belonging to $\frac{1}{\sqrt{\epsilon}}$ will be set equal to zero. Hence, in each step a partial differential equation is obtained from which as much information as possible about the functions P_i will be gathered. Finally, these functions will be expressed explicitly. The more parts of (57) will be included, the more difficult the calculation of the approximations will be. However, if too few parts of (57) are taken into account, it may be that important characteristics, like hump-shaped yield curves, will not be included in the model. The detailed solution procedure to derive (58) can be found in appendix D.

In the end, Fouque et al. (2000) approximate the price of a zero-coupon bond by

$$P^\epsilon(\tau,r,y) \approx P_0(\tau,r) + \sqrt{\epsilon}P_1(\tau,r) = \big(1 + \sqrt{\epsilon}D(\tau)\big)A(\tau)\ e^{-rB(\tau)}, \tag{58}$$

where $A(\tau), B(\tau)$ and $D(\tau)$ are defined in (134), (133) and (152).

Next, $\langle y\Phi' \rangle$, which appears in (145), (146) and (147), will be discussed in more detail, because the parameters V_1, V_2 and V_3 still depend on y. Equations (141) and (53) can be used to construct a second order linear inhomogeneous differential equation for Φ.

$$\Phi'' + \frac{2(\theta_2 - y)}{v^2 y}\Phi' + \frac{2(\theta_2 - y)}{v^2 y} = 0 \tag{59}$$

A first order inhomogeneous linear differential equation can easily be created by substituting $\Phi' = \Psi$.

$$\Psi' + \frac{2(\theta_2 - y)}{v^2 y}\Psi + \frac{2(\theta_2 - y)}{v^2 y} = 0 \tag{60}$$

This differential equation can be solved using the method variation of constants, which results in

$$\Psi(y) = \Phi'(y) = \frac{2}{v^2 y f(y)} \int_0^y \big(\xi - \langle \xi \rangle\big) f(\xi) d\xi \tag{61}$$

It has been checked that (61) is correct by substituting the derivative of (61) and (61) itself into (60). Then, by using the definition of the expected value of a random variable, $\langle y\Phi' \rangle$ can be expressed as

$$\langle y\Phi' \rangle = \Big\langle \frac{2}{v^2 f(y)} \int_0^y \big(\xi - \langle \xi \rangle\big) f(\xi) d\xi \Big\rangle = \frac{2}{v^2} \int_0^\infty \int_0^y \big(\xi - \langle \xi \rangle\big) f(\xi) d\xi dy \tag{62}$$

Three more lemmas are required in order to proceed.

Lemma 1. $\Gamma(a, z) = \int_z^\infty t^{a-1} e^{-t} dt$ *is the upper incomplete gamma function*

Lemma 2. $\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt$ *is the lower incomplete gamma function*

Consequently, this implies the following relationship:

$$\Gamma(a, z) + \gamma(a, z) = \Gamma(a) \tag{63}$$

The final lemma represents an additional border condition concerning the upper incomplete gamma function.

Lemma 3. *For an integral from the upper incomplete gamma function:*

$$\int \Gamma(a, z) dz = z\Gamma(a, z) - \Gamma(a + 1, z) \tag{64}$$

Next, the inner integral of (62) can be calculated in the following way:

$$\int_0^y \big(\xi - \langle\xi\rangle\big) f(\xi) d\xi = \int_0^y (\xi - \theta_2) \frac{\beta^\alpha}{\Gamma(\alpha)} \xi^{\alpha-1} e^{-\beta\xi} d\xi$$

$$= \int_0^{\beta y} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{t^\alpha}{\beta^{\alpha+1}} e^{-t} dt - \int_0^{\beta y} \theta_2 \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{t^{\alpha+1}}{\beta^\alpha} e^{-t} dt = \frac{1}{\beta\Gamma(\alpha)} \int_0^{\beta y} t^\alpha e^{-t} dt - \frac{\alpha}{\beta\Gamma(\alpha)} \int_0^{\beta y} t^{\alpha-1} e^{-t} dt$$

$$= \frac{\gamma(\alpha+1, \beta y) - \alpha\gamma(\alpha, \beta y)}{\beta\Gamma(\alpha)} = \frac{\Gamma(\alpha+1) - \Gamma(\alpha+1, \beta y) - \alpha\Gamma(\alpha) + \alpha\Gamma(\alpha, \beta y)}{\beta\Gamma(\alpha)}$$

$$= \frac{\alpha\Gamma(\alpha, \beta y) - \Gamma(\alpha+1, \beta y)}{\beta\Gamma(\alpha)}$$

In the second step, the substitutions $t = \beta\xi$ and $\theta_2 = \frac{\alpha}{\beta}$ are applied, while the definition of the lower incomplete gamma function is used in the fourth step. Furthermore, (63) is used in the fifth step and the well known property $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ is applied in the sixth step. Finally, $\langle y\Phi'\rangle$ can be calculated as

$$\langle y\Phi'\rangle = \frac{2}{v^2} \int_0^\infty \frac{\alpha\Gamma(\alpha, \beta y) - \Gamma(\alpha+1, \beta y)}{\beta\Gamma(\alpha)} dy = \frac{2}{v^2\beta^2\Gamma(\alpha)} \int_0^\infty \alpha\Gamma(\alpha, x) - \Gamma(\alpha+1, x) dx$$

$$= \frac{2}{v^2\beta^2\Gamma(\alpha)} \Big[\alpha\big(x\Gamma(\alpha, x) - \Gamma(\alpha+1, x)\big) - \big(x\Gamma(\alpha+1, x) - \Gamma(\alpha+2, x)\big)\Big]_0^\infty$$

$$= \frac{2}{v^2\beta^2\Gamma(\alpha)} \Big[\alpha\Gamma(\alpha+1) - \Gamma(\alpha+2)\Big] = \frac{2}{v^2\beta^2\Gamma(\alpha)} \Big[\alpha\Gamma(\alpha+1) - (\alpha+1)\Gamma(\alpha+1)\Big]$$

$$= -\frac{2\Gamma(\alpha+1)}{v^2\beta^2\Gamma(\alpha)} = -\frac{2\alpha}{v^2\beta^2} = -\theta_2$$

In the second step, the substitution $x = \beta y$ is applied, while lemma 3 is used in the third step. Moreover, the following definitions of the limits are used in the fourth step:

$$\lim_{x\to 0}\Gamma(a, x) = \Gamma(a), \quad \lim_{x\to\infty}\Gamma(a, x) = 0, \quad \lim_{x\to 0} x\Gamma(a, x) = 0, \quad \lim_{x\to\infty} x\Gamma(a, x) = \lim_{x\to\infty} x^{a+1}e^{-x} = 0$$

In the final steps, it is used that $\alpha = \frac{2\theta_2}{v^2}$ and $\beta = \frac{2}{v^2}$.

In the end, the group parameters V_1, V_2 and V_3, which have been defined in (145), (146) and (147), do not depend on y anymore and can be indicated in terms of the model parameters as

$$V_1 = -\lambda_1 \lambda_2 v \theta_2 \tag{65}$$

$$V_2 = \frac{1}{2}\lambda_2 v \theta_2 + \lambda_1 \rho v \theta_2 \tag{66}$$

$$V_3 = -\frac{1}{2}\rho v \theta_2 \tag{67}$$

All in all, it has been shown that the approximated bond price does not depend on y, but only on r and τ.

Moreover, (58) indicates that C should vanish for very small values of ϵ (very large values of κ_2). This is confirmed by figure 10. The figure shows that C gets closer to zero the larger κ_2 becomes. Furthermore, the fundamental difference between the initial Fong-Vasicek model with stochastic volatility and this approximation is that the functions $P_0(\tau, r)$ and $P_1(\tau, r)$ in the approximation do not depend on y. Fouque et al. (2000) also plot the functions P and P^ϵ for some values of r and ranges of y. Their results show that P^ϵ is a decent approximation of P. After the developments in this subsection, the approximated yield curve can be defined by

$$R^\epsilon(\tau, r) = -\frac{1}{\tau}\ \ln\left[P^\epsilon(\tau, r)\right] = -\frac{1}{\tau}\Big[\ \ln\left[A(\tau)\right] - rB(\tau) + \ln\left(1 + \sqrt{\epsilon} D(\tau)\right)\ \Big] \tag{68}$$

The approximation of the yield curve is the same for all the different values of y, because it does not depend on y.

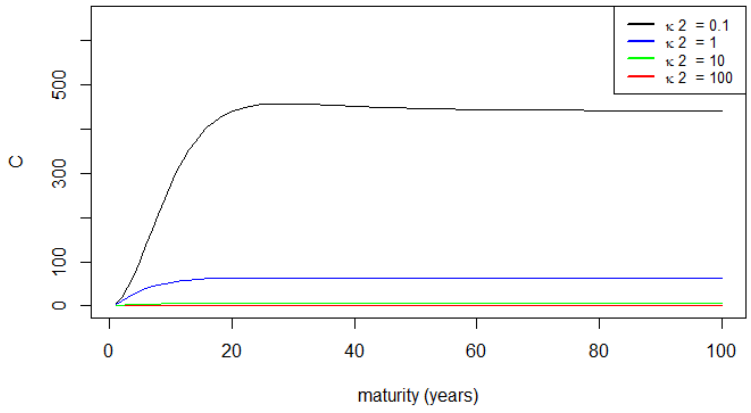
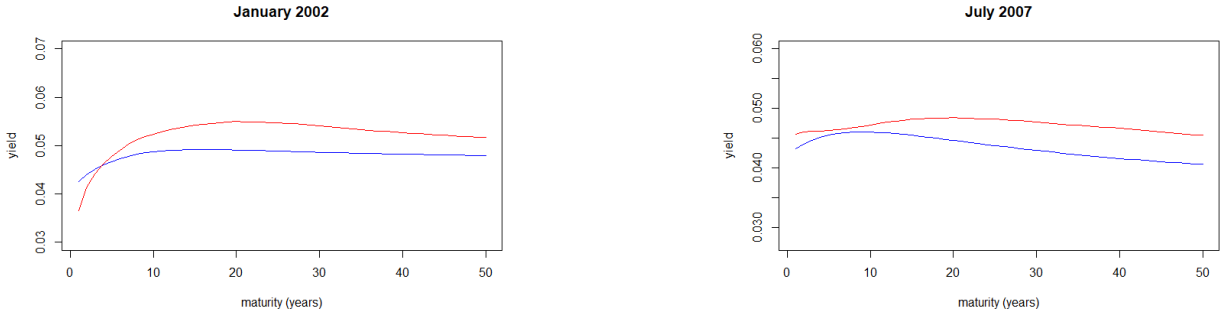
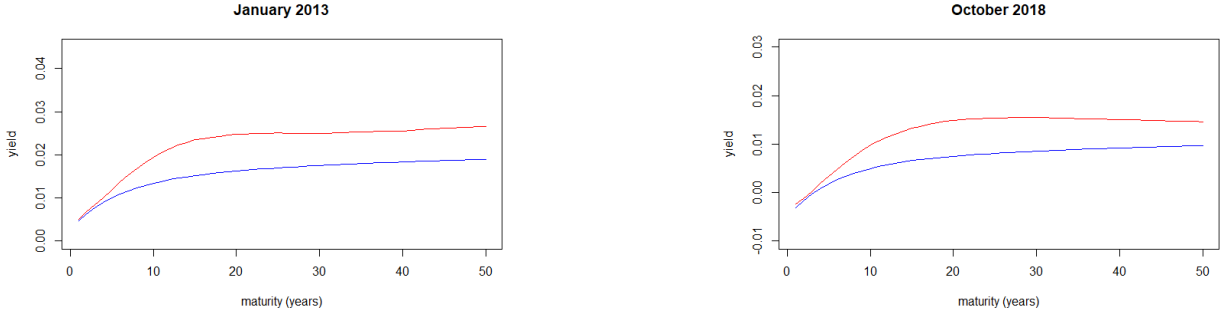


Figure 10: the plot shows the values of C while changing κ_2 and holding the other parameters fixed



(a) (b)

Figure 11: fitted yield curves (blue) vs observed yield curves (red)

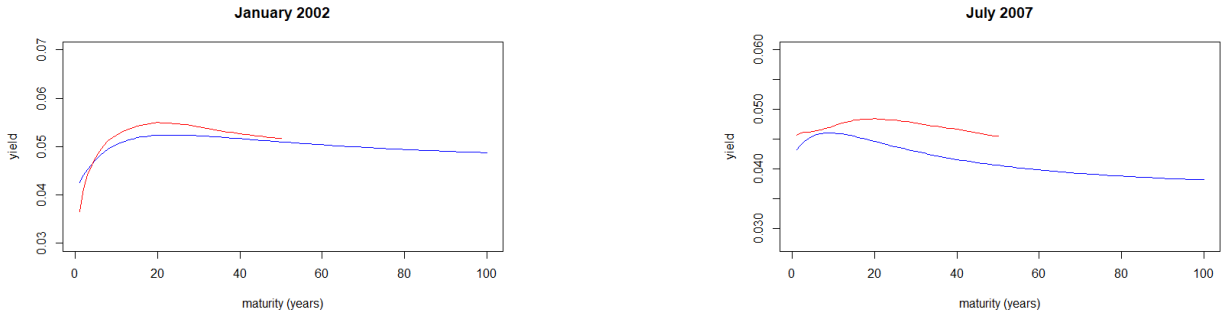


(a) (b)

Figure 12: fitted yield curves (blue) vs observed yield curves (red)

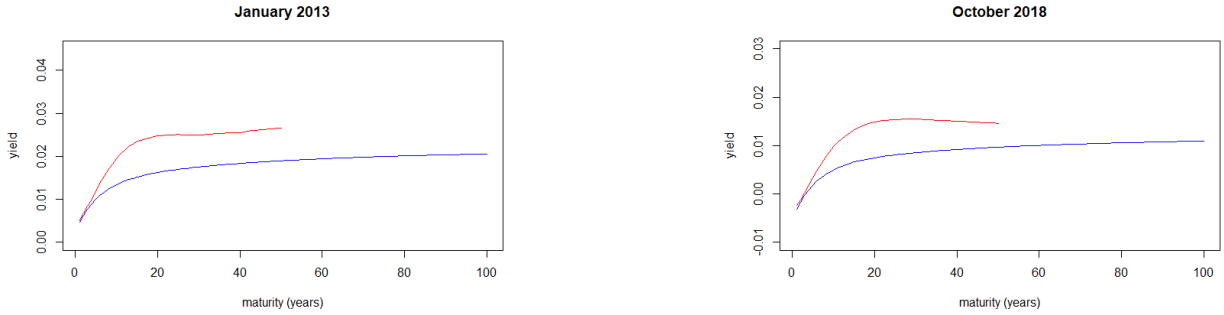
$\hat{r}(t)$ and the estimated parameter values in table 3 can be used to plot the fitted yield curves for each month i ($i = 1, ..., 202$) against the observations of the zero rates using (68). The maximum likelihood parameter estimate for κ_2 is pretty small. Hence, it is to be expected that the approximation will not be that accurate. However, it has been checked that the effect on the fitted yield curves, even after a change in the value of κ_2, is quite small. The yield curves are fitted for maturities up to 50 years. In figure 11 and figure 12, the fitted yield curves for the same months as before are plotted against the real observations. In all plots in figure 11 and figure 12, it can be noted that the differences between the observed yield values and the fitted yield values are no longer so small. These differences are substantially bigger than the differences in figure 8 and figure 9, which are based on the Fong-Vasicek model. This is not surprising, because the plots in figure 11 and figure 12 are based on an approximation and the parameter values of the Fong-Vasicek model have been used. Furthermore, the

downward sloping convexity effect seems reasonably good suited to explain the lower long-term yield rates for January 2002 and July 2007. However, this is certainly not the case for January 2013 and October 2018.



(a) (b)

Figure 13: fitted yield curves extrapolated up to 100 years (blue) vs observed yield curves (red)



(a) (b)

Figure 14: fitted yield curves extrapolated up to 100 years (blue) vs observed yield curves (red)

Next, in figure 13 and figure 14, the fitted yield curves are extended to maturities very far ahead in the future up to 100 years. Convexity is again discernible at the end of the fitted yield curves of January 2002 and July 2007. It causes downward sloping fitted yield curves in those months. All in all, the fitted yield curves based on the asymptotic approximation of the Fong-Vasicek model are reasonably good suited to explain the downward sloping convexity effect for some months. However, the differences between the observed yield curves and the fitted yield curves based on the asymptotic approximation of the Fong-Vasicek model are substantially bigger than the differences between the observed yield curves and the fitted

yield curves based on the Fong-Vasicek model. Therefore, the asymptotic approximation based on the Fong-Vasicek model will not be considered anymore in the further analysis of this paper.

4 Advice Committee Parameters

In section 2, the yield curve based on the Vasicek model has been compared to the EIOPA yield curve based on the Smith-Wilson extrapolation. It could be concluded that the Smith-Wilson extrapolation lets the forward rate converge to the UFR much faster than the case is for the extrapolated yield curves based on the Vasicek model. Eventually, both yield curves would diverge completely, because of the extreme convexity at the very long end of the yield curve, which is reflected in the extrapolated yield curves based on the Vasicek model. In this section, a derivation of the UFR method proposed by Committee Parameters will be provided. Moreover, the EIOPA yield curve as well as the yield curves based on the Vasicek and Fong-Vasicek model will be compared to the method used by Committee Parameters. This Dutch Committee focuses on the yield curve based on which pension liabilities are valued. Its main task is to provide advice on three components of the UFR method, i.e. the level of the UFR, the point on the yield curve from which the UFR method becomes operative and the extrapolation method itself (Dijsselbloem et al., 2019). Furthermore, the Committee investigates to what extent the proposals that are currently being discussed on a European level might be used as a reference. Consequently, it has been concluded that the nature and timing of the developments in Europe offer insufficient basis for its advice about the UFR method with respect to Dutch pension funds (Dijsselbloem et al., 2019). One main criticism is that the Smith-Wilson method does not show the convexity effect. Moreover, the level of the UFR would be too high and the rate of convergence to the UFR in the EIOPA-UFR method would be very high in comparison to convergence rates in standard academic yield term structure models like the Vasicek and Fong-Vasicek model. Consequently, pension funds run the risk of their obligations to be substantially undervalued. Therefore, the Committee decided to issue an independent subject-specific advice on the UFR curve for Dutch pension funds. In short, the committee recommends the first smoothing point to be 30 years, because the liquidity in the market up to this maturity is sufficient to fully rely on market information (Dijsselbloem et al., 2019). In this paper, the 'last liquid point' was set to 20 years and the yield curves based on the Vasicek and Fong-Vasicek model were extrapolated beyond 20 years to maturity in stead of 30 years to maturity. Moreover, the Committee advices the parameter that determines how fast the yield curve converges to the UFR to be 0.02 (Dijsselbloem et al., 2019). This value is larger than the estimate in this paper based on the Vasicek model, which indicates that the Committee considers a faster

rate of convergence to the UFR. This is expressed in a downward slope at the very long end of the yield curve based on the Vasicek and Fong-Vasicek model in this paper, whereas the yield curve in Dijsselbloem et al. (2019) does not represent this convexity effect. Finally, the Committee recommends to compute the level of the UFR as the 10-year unweighted moving average of the 30-year forward interest rate, which results in a highly predictable UFR level that changes only slightly from month to month, and a UFR level that is based on market information. (Dijsselbloem et al., 2019). In this paper, it has been established that a pretty low level of mean reversion is present at the very long end of the yield curve based on the Vasicek and Fong-Vasicek model. Consequently, it could be concluded that a constant UFR is impossible to estimate from long-term maturity data of yields. However, Dijsselbloem et al. (2019) notice that the impact of the UFR level on the yield curve, and thus on funding ratios and premiums, is limited, which reduces the importance of the UFR level.

Next, a more technical derivation of the UFR method proposed by the Committee will be provided. Up to and including the first smoothing point of 30 years, the zero rate term structure at time t is based on the euro swap curve for maturities 1 to 10, 12, 15, 20, 25 and 30 years. For the purpose of extrapolation, the obtained zero rates $z(t,h)$ at time t with term h are converted into continuously compounded zero rates $z_c(t,h)$ according to the following formula:

$$z_c(t,h) = \ln\big(1 + z(t,h)\big) \tag{69}$$

As of a maturity of 30 years, the zero rate term structure at time t is based on the following extrapolation method. As mentioned before, the UFR is equal to the 120-month unweighted moving average of the 30-year forward interest rates with a term of 1 year:

$$UFR(t) = \frac{1}{120}\sum_{i=0}^{119} f(t-i, 30, 31), \tag{70}$$

where $f(t,k,k+l)$ reflects the k-year forward rate with maturity l at time t:

$$f(t,k,k+l) = \left[\frac{\big(1+z(t,k+l)\big)^{k+l}}{\big(1+z(t,k)\big)^{k}}\right]^{1/l} - 1 \tag{71}$$

Then, the obtained $UFR(t)$ is transformed in the following way:

$$UFR_c(t) = \ln\big(1 + UFR(t)\big) \tag{72}$$

Next, based on the observed zero rates, the weighted forward rate $f_c^D(t,s)$ is determined for the last five trading days s of month t:

$$f_c^D(t,s) = \frac{2}{3}f_c^D(t,s,30,40) + \frac{1}{3}f_c^D(t,s,30,50), \tag{73}$$

where $f_c^D(t, s, k, k+l)$ reflects the natural logarithm of the k-year forward rate with maturity l on trading day s of month t:

$$f_c^D(t, s, k, k + l) = \ln\big(1 + f^D(t, s, k, k + l)\big) \tag{74}$$

After, the last liquid forward rate $f_c^*(t)$ at month t is calculated as the average $f_c^D(t, s)$ for the last 5 trading days s of month t. By including market information after the first smoothing point in this way, there is less interest rate sensitivity around the 30-year first smoothing point. The 30-year forward rate with a 10-year term is given greater weight than the 30-year forward rate with a 20-year term due to the higher liquidity.

Then, extrapolation of the forward rates after 30-years is done in the following way:

$$f_c(t, 30, 30 + h) = UFR_c(t) + \big(f_c^*(t) - UFR_c(t)\big)B(h), \tag{75}$$

where $B(h) = \frac{1-e^{-ah}}{ah}$ with convergence parameter $a = 0.02$.
Thereafter, extrapolation of the zero rates after the 30-year first smoothing point is done in the following way:

$$z_c(t, 30 + h) = \frac{30 z_c(t, 30) + h f_c(t, 30, 30 + h)}{30 + h} \tag{76}$$

Finally, the obtained continuously compounded zero rates are converted into annually compounded zero rates using the following formula:

$$z(t, 30 + h) = e^{z_c(t,30+h)} - 1 \tag{77}$$

In the next section, this methodology will be used to determine the zero curve and to discount pension obligations.

5 An application

Pension funds are promising very long-dated cash flows to their participants. They must use a discount rate to calculate the present value of their long-dated liabilities. In this way, pension funds can compute their funding ratio. The funding ratio is defined as

$$\text{funding ratio} = \frac{\text{current value of the investments}}{\text{present value of the pension obligations}} \tag{78}$$

In this section, the present value of the liabilities and the funding ratio of the Dutch pension fund ABP in the fourth quarter of 2018 will be calculated using the yield curves based on

the Vasicek and Fong-Vasicek model as well as based on the method proposed by Committee Parameters. The present value of the pension obligations is defined as the current and future pension benefits to be paid out by the pension fund. In this application, it is calculated in the following way:

$$\text{present value of the pension obligations } = \sum_{\tau=0}^{\tau=85} \frac{[CF]_\tau}{(1+R_\tau)^\tau} \tag{79}$$

The summation starts at $\tau = 0$, because the current pension benefits should be taken into account. The summation ends at $\tau = 85$, because there are no future cash flows promised to participants beyond 85 years in the future. Moreover, $[CF]_\tau$ denotes the cash flow that is promised to participants τ years in the future. R_τ is the zero rate belonging to a specific model with time to maturity τ.

ABP had approximately 2.9 million participants in the fourth quarter of 2018. Five main categories need to be considered to determine the value of the pension obligations: the sum of the accumulated pension entitlements of participants, retirement pension, partner's pension, disability pension and orphan's pension. Moreover, the retirement age and mortality rates are taken into account to approach the value of the pension obligations as close as possible.

	Vasicek	Fong-Vasicek	Committee Parameters
current value of the investments	399 billion euro	399 billion euro	399 billion euro
present value of the pension obligations	375.1 billion euro	400.2 billion euro	395.8 billion euro
funding ratio	106.4%	99.7%	100.8%

Table 5: the present value of the pension obligations and corresponding funding ratio of ABP in the fourth quarter of 2018 using the yield curves of December 2018 based on the Vasicek and Fong-Vasicek model as well as based on the method proposed by Committee Parameters

In table 5, the present value of the pension obligations and corresponding funding ratio in the fourth quarter of 2018 are shown, while using the yield curves of December 2018 based on the Vasicek and Fong-Vasicek model as well as based on the method proposed by Committee Parameters. ABP reported a current value of their investments of 399 billion euro in the fourth quarter of 2018. Moreover, they announced a present value of their pension obligations of 411 billion euro, while having used an UFR of 1.4%. Consequently, a funding ratio of 97% came about in the fourth quarter of 2018. It can be observed that the present value of the pension obligations based on the yield curves of all three models is lower than reported by ABP, while the corresponding funding ratios are higher. This can be explained by a yield curve of December 2018 based on the Vasicek model that is upward sloping until 86

years to maturity. Furthermore, it has a yield level that is above the UFR of 1.4% for most maturities. Besides, the yield curve of December 2018 based on the Fong-Vasicek model has a yield level that is just above 1.4% between 18 and 88 years to maturity. Hence, it also makes sense that the present value of the pension obligations based on this model is a bit lower than the one reported by ABP. Finally, the yield curve of December 2018 based on the method proposed by Committee Parameters has a yield level that is just above the UFR of 1.4% as of 18 years to maturity until the end of the yield curve. Furthermore, the UFR of 2.1%, which is computed using (72), is higher than 1.4%. Therefore, the present value of the pension obligations based on this model is also a bit lower than the one reported by ABP. The funding ratio based on the Fong-Vasicek model is closest to the 97% mentioned by ABP. This is no surprise, because this yield curve approaches the yield curve used by ABP. The long-term yield rate level of the Fong-Vasicek model is close to 1.4%. Figure 15 contains an overview of the the observed yield curve as well as the yield curves based on the models proposed by Vasicek, Fong-Vasicek and Committee Parameters. All in all, it can be concluded that the use of different yield curves might lead to significantly different funding ratios, which is clearly visible in this case. The yield curve based on the Fong-Vasicek model fits the observed data closest and leads to a slightly more favorable funding ratio than the one reported by ABP.

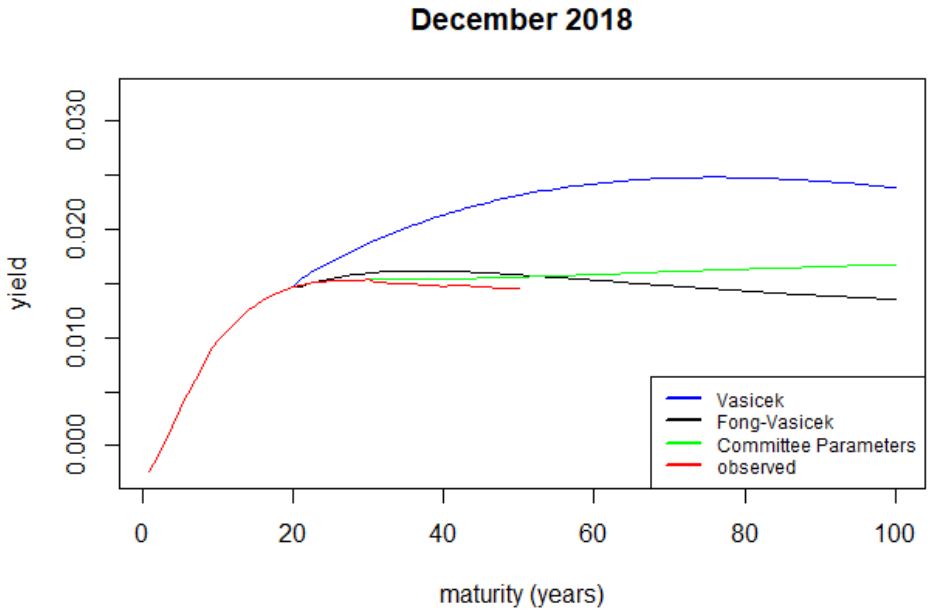


Figure 15: the plot shows the observed yield curve as well as the yield curves based on the models proposed by Vasicek, Fong-Vasicek and Committee Parameters

6 Conclusion

The main purpose of this paper has been to investigate what would be appropriate model-based yield rates to discount very long-dated liabilities. In the first part of this paper, maximum likelihood estimation has been used to estimate the parameters of the one factor Vasicek model. It has been established that a pretty low level of mean reversion is present. Moreover, it has been noticed that the Vasicek model predicts that yield rates with a maturity between 20 and 50 years should be higher than the observed yield rates in this same period. Consequently, that model would not be very suitable to explain the relatively low observed yield rates up to at least 50 years to maturity. However, the downward pattern at the very long end of the fitted yield curve is better suited to explain the relatively low observed yield rates at the very long end of the yield curve. In the second part of this paper, the two factor Fong-Vasicek model with stochastic volatility has been defined. A two factor model, for which the short rate and volatility are stochastic, is better suited to describe the behavior of markets than a one factor model, because it allows for more diversity in terms of potential outcomes than a one factor model. Maximum likelihood estimation has again been used to estimate the model parameters. It could be concluded that the fitted yield curves based on the Fong-Vasicek model are better suited to explain the downward sloping convexity effect, and are more accurately picking up the point where the downward sloping convexity effect starts, than the fitted yield curves based on the Vasicek model. In addition, it has been formally tested whether the Fong-Vasicek model indeed has a better goodness of fit than the Vasicek model. There was found evidence for the Fong-Vasicek model to have a significantly better goodness of fit than the Vasicek model. Moreover, the concept of fast-scaling volatility has been introduced. Based on this concept, an asymptotic approximation of the two factor Fong-Vasicek model has been constructed. Furthermore, it has been shown that the approximated bond price does not depend on y, but only on r and τ. Finally, it has been established that the fitted yield curves based on the asymptotic approximation of the Fong-Vasicek model are reasonably good suited to explain the downward sloping convexity effect for some months. However, the differences between the observed yield curves and the fitted yield curves based on the asymptotic approximation of the Fong-Vasicek model are substantially bigger than the differences between the observed yield curves and the fitted yield curves based on the Fong-Vasicek model. Therefore, the asymptotic approximation based on the Fong-Vasicek model has not been considered anymore in the further analysis of this paper. In the next part of this paper, a derivation of the UFR method proposed by Committee Parameters has been provided. It could be concluded that the yield curve based on this method approaches the observed yield curve much closer than the EIOPA yield curve based on the Smith-Wilson extrapolation. In addition, the yield curve based on the method proposed by Committee Parameters has been compared to the yield curves based on the Vasicek and Fong-Vasicek model. The yield curve in Dijsselbloem et al. (2019) does not represent the convexity effect, whereas the convexity effect is incorporated at the very long

end of the yield curves based on the Vasicek and Fong-Vasicek model. In the final part of this paper, the present value of the pension obligations and corresponding funding ratio of the Dutch pension fund ABP in the fourth quarter of 2018 has been computed using the yield curves of December 2018 based on the model proposed by Vasicek, Fong-Vasicek and Committee Parameters. It could be concluded that the use of different yield curves might lead to significantly different funding ratios.
All in all, the Fong-Vasicek model would be suitable to predict the behavior of very long-term yield rates. These long-term yield rates can be used to discount very long-dated liabilities.

Appendix A The derivation of (1)

The variables are defined in the following way:
$P(t,T)$: the price at time t of a discount bond maturing at time T

$$P(t,T) = e^{-R(t,T)(T-t)} = e^{-R(t,T)\tau} \tag{80}$$

$R(t,T)$: the internal rate of return / spot rate at time t of a discount bond maturing at T
$r(t)$: instantaneous interest rate (short rate) at time t

The following relationships apply:

$$R(t,T) = -\frac{1}{\tau}\ \ln\big[P(t,T)\big], \tag{81}$$

where $\tau > 0$, and τ is time to maturity.

$$r(t) = \lim_{\tau\to 0} R(t,t+\tau) = \lim_{\tau\to 0}\ -\frac{1}{\tau}\Big(\ \ln\big[P(t,t+\tau)\big] - \ln\big[P(t,t)\big]\ \Big) = -\frac{\partial}{\partial T}\ \ln\big[P(t,t)\big] \tag{82}$$

The assumption is made that $r(t)$ follows a stochastic process. It is defined by the stochastic differential equation

$$dr = v(r,t)dt + s(r,t)dz, \tag{83}$$

where $v(r,t)$ is the drift of r and $s(r,t)$ is the volatility of r. z is a wiener process under probability measure $\mathbb{Q}$.

Furthermore, it is assumed that the bond price may be written as a function of the current short rate

$$P(t,T) = P\big(r(t),t,T\big) \tag{84}$$

Moreover, the market is assumed to be efficient.

Then, Ito's Lemma is applied to the bond price and (83) is used to derive the stochastic differential equation for the bond price:

$$dP = \Big[v(r,t)\frac{\delta P}{\delta r} + \frac{\delta P}{\delta t} + \frac{1}{2}s^2(r,t)\frac{\delta^2 P}{\delta r^2}\Big]dt + s(r,t)\frac{\delta P}{\delta r}dz \tag{85}$$

When the dynamics of $P(t,T)$ are expressed in the following log-normal form

$$\frac{dP}{P} = \mu(t,T)dt + \sigma(t,T)dz, \tag{86}$$

the drift rate $\mu(t,T)$ and volatility $\sigma(t,T)$ of the bond price process are defined as

$$\mu(t,T) = \frac{1}{P}\Big[v(r,t)\frac{\delta P}{\delta r} + \frac{\delta P}{\delta t} + \frac{1}{2}s^2(r,t)\frac{\delta^2 P}{\delta r^2}\Big] \tag{87}$$

$$\sigma(t,T) = -\frac{1}{P}s(r,t)\frac{\delta P}{\delta r} \tag{88}$$

Then, (85) can be rewritten as

$$dP = \big[\mu(t,T)P\big]dt - \big[\sigma(t,T)P\big]dz, \tag{89}$$

where $\mu(t,T)$ and $\sigma(t,T)$ are the mean and volatility of the instantaneous rate of return at time t of a zero coupon bond maturing at time T.

Besides, $q(r,t)$ is the market price of risk and given by

$$q(r,t) = \frac{\mu(t,T) - r(t)}{\sigma(t,T)} \tag{90}$$

When substituting (87) and (88) into (90), a partial differential equation (term structure equation) for the bond price is derived:

$$\frac{\delta P}{\delta t} + \big[v(r,t) + s(r,t)q(r,t)\big]\frac{\delta P}{\delta r} + \frac{1}{2}s^2(r,t)\frac{\delta^2 P}{\delta r^2} - r(t)P = 0 \tag{91}$$

From now on, it is assumed that $q(r,t) = q$ and that the short-term interest rate follows an Ornstein-Uhlenbeck process

$$dr = \alpha(\gamma - r)dt + sdz, \tag{92}$$

where α is the mean reversion parameter, γ the unconditional mean of the factor under the risk-neutral measure $\mathbb{Q}$, and s the volatility. Furthermore, $\alpha > 0$.
$\alpha(\gamma - r)$ displays mean reversion, with r being pulled to its long-term mean.

When substituting (92) into (91), the term structure equation can be rewritten as

$$\frac{\delta P}{\delta t} + \big[\alpha(\gamma - r) + s(r,t)q(r,t)\big]\frac{\delta P}{\delta r} + \frac{1}{2}s^2(r,t)\frac{\delta^2 P}{\delta r^2} - r(t)P = 0 \tag{93}$$

Allowing the bond price to be of the form

$$P\big(r(t),t,T\big) = A(t,T)e^{-B(t,T)r(t)} \tag{94}$$

Then, (93) can be rewritten as

$$A_t e^{-Br} - rAB_t e^{-Br} - ABe^{-Br}\big[\alpha(\gamma - r) + sq\big] + \frac{1}{2}s^2AB^2e^{-Br} - rAe^{-Br} = 0$$

$$e^{-Br}\Big[A_t - rAB_t - AB\big[\alpha(\gamma - r) + sq\big] + \frac{1}{2}s^2AB^2 - rA\Big] = 0$$

$$A_t - (\alpha\gamma + sq)AB + \frac{1}{2}s^2AB^2 = rA + rAB_t - \alpha rAB$$

$$A_t - (\alpha\gamma + sq)AB + \frac{1}{2}s^2AB^2 = rA(1 + B_t - \alpha B),$$

where $A = A(t,T),\ B = B(t,T),\ r = r(t),\ s = s(r,t),\ q = q(r,t),\ A_t = \frac{\delta A(t,T)}{\delta t}$ and $B_t = \frac{\delta B(t,T)}{\delta t}$.

Since the right-hand side is a function of r and the left-hand side is a function of t and T only, the following must hold:

$$1 + B_t - \alpha B = 0 \tag{95}$$

$$A_t - (\alpha\gamma + sq)AB + \frac{1}{2}s^2AB^2 = 0 \tag{96}$$

Solving (95) with $B(T,T) = 0$ results in

$$B(t,T) = \frac{1}{\alpha}\big(1 + B_t\big) = \frac{1}{\alpha}\big(1 - e^{-\alpha(T-t)}\big) \tag{97}$$

Rearranging (96) with $A(T,T) = 0$ results in

$$A_t + AB\Big(\frac{1}{2}s^2B - \alpha\gamma - sq\Big) = 0$$

$$\frac{1}{A}dA + B\Big(\frac{1}{2}s^2B - \alpha\gamma - sq\Big)dt = 0$$

$$\int_t^T \frac{1}{A}dA + \int_t^T \Big(\frac{1}{2}s^2B^2(\mu,T) - (\alpha\gamma + sq)B(\mu,T)\Big)d\mu = 0$$

$$\ln\big(A(T,T)\big) - \ln\big(A(t,T)\big) + \Big[\ \frac{1}{2}s^2\Big(\frac{\mu}{\alpha^2} - \frac{2}{\alpha^3}e^{-\alpha T+\alpha\mu} + \frac{1}{2\alpha^3}e^{-2\alpha T+2\alpha\mu}\Big)\ \Big]_{\mu=t}^{\mu=T}$$

$$-\Big[\ \Big(\alpha\gamma + sq\Big)\Big(\frac{\mu}{\alpha} - \frac{1}{\alpha^2}e^{-\alpha T+\alpha\mu}\Big)\ \Big]_{\mu=t}^{\mu=T} = 0$$

$$\ln\big(A(t,T)\big) = \frac{s^2}{2\alpha^2}\Big[\ \mu - \frac{2}{\alpha}e^{-\alpha(T-\mu)} + \frac{1}{2\alpha}e^{-2\alpha(T-\mu)}\ \Big]_{\mu=t}^{\mu=T} - \ \Big(\gamma + \frac{sq}{\alpha}\Big)\Big[\ \mu - \frac{1}{\alpha}e^{-\alpha(T-\mu)}\ \Big]_{\mu=t}^{\mu=T} =$$

$$\frac{s^2}{2\alpha^2}\Big(T-t-\frac{2}{\alpha}\big(1-e^{-\alpha(T-t)}\big)+\frac{1}{2\alpha}\big(1-e^{-2\alpha(T-t)}\big)\Big) - \ \Big(\gamma+\frac{sq}{\alpha}\Big)\Big(T-t-\frac{1}{\alpha}\big(1-e^{-\alpha(T-t)}\big)\Big) =$$

$$\Big(\gamma+\frac{sq}{\alpha}-\frac{s^2}{2\alpha^2}\Big)\Big(-(T-t)+\frac{1}{\alpha}\big(1-e^{-\alpha(T-t)}\big)\Big) - \frac{s^2}{2\alpha^2}\Big(\frac{1}{\alpha}\big(1-e^{-\alpha(T-t)}\big)\Big) + \frac{s^2}{4\alpha^3}\Big(1-e^{-2\alpha(t-t)}\Big)$$

$$A(t,T) = \exp\Big[\ \Big(\gamma+\frac{sq}{\alpha}-\frac{s^2}{2\alpha^2}\Big)\ \Big(\frac{1}{\alpha}\big(1-e^{-\alpha(T-t)}\big)-\big(T-t\big)\Big) - \frac{s^2}{4\alpha^3}\Big(1-e^{-\alpha(T-t)}\Big)^2\ \Big] \quad (98)$$

Substituting the expression for $A(t,T)$ and $B(t,T)$ into (94) results in

$$\begin{aligned} P(r,t,T) = \exp\Big[\ &\Big(\gamma+\frac{sq}{\alpha}-\frac{s^2}{2\alpha^2}\Big)\Big(\frac{1}{\alpha}\big(1-e^{-\alpha(T-t)}\big)-\big(T-t\big)\Big) \\ &-\frac{s^2}{4\alpha^3}\Big(1-e^{-\alpha(T-t)}\Big)^2 - \frac{r}{\alpha}\Big(1-e^{-\alpha(T-t)}\Big)\ \Big] \end{aligned} \quad (99)$$

Next, the long-term interest rate can be defined as

$$R(\infty) = \gamma + \frac{sq}{\alpha} - \frac{s^2}{2\alpha^2} \quad (100)$$

Balter et al. (2019) leave out the part $\frac{sq}{\alpha}$. This is done, because this term is more important for determining the shape at intermediate maturities than at long maturities.

Next, the bond price can be rewritten as

$$P(r,t,T) = \exp\Big[\ \big(R(\infty)-r\big)\ \frac{1}{\alpha}\Big(1-e^{-\alpha(T-t)}\Big) - (T-t)R(\infty) - \frac{s^2}{4\alpha^3}\Big(1-e^{-\alpha(T-t)}\Big)^2\ \Big] \quad (101)$$

Then, $R(t,T)$ can be derived by substituting (101) into (81):

$$\begin{aligned} R(t,T) = -\frac{1}{\tau}\ \ln\Bigg[\exp\Big[\big(R(\infty)-r\big)\ \frac{1}{\alpha}\Big(1-e^{-\alpha(T-t)}\Big) - (T-t)R(\infty) - \frac{s^2}{4\alpha^3} \\ \Big(1-e^{-\alpha(T-t)}\Big)^2\Big]\ \Bigg] = R(\infty) + \big(r-R(\infty)\big)\ \frac{1}{\alpha\tau}\Big(1-e^{-\alpha\tau}\Big) + \frac{s^2}{4\alpha^3\tau}\Big(1-e^{-\alpha\tau}\Big)^2, \end{aligned} \quad (102)$$

where $\tau \geq 0$.

Equation (1) in Balter et al. (2019) corresponds to (92) and (2) in Balter et al. (2019) corresponds to (102), except for the small difference mentioned above.
In this paper, the notation of Balter et al. (2019) will be used:

$$y_t(\tau) = \theta + b(\kappa\tau)(r_t - \theta) + \frac{b(\kappa\tau)^2\tau\sigma^2}{4\kappa}, \tag{103}$$

with

$$b(z) = \frac{1 - e^{-z}}{z} \tag{104}$$

$$\theta = \mu - \frac{\sigma^2}{2\kappa^2} \tag{105}$$

$$dr = \kappa(\mu - r)dt + \sigma dW, \tag{106}$$

where κ is a mean reversion parameter, μ the unconditional mean of the level factor r under the risk-neutral measure $\mathbb{Q}$, σ the volatility, and dW a standard Brownian Motion.

Appendix B The derivation of (29)

The starting point is (27) and (28). Cholesky's decomposition has been applied to define the stochastic volatility term structure in terms of two uncorrelated Wiener processes $dZ_1^{\mathbb{P}}$ and $dZ_2^{\mathbb{P}}$. Next, a risk-neutral probability measure $\mathbb{Q}$ consistent with the no-arbitrage condition is needed to convert the discounted bond prices into martingales. In order to apply this change of probability measure, the parametric equations $\pi\sqrt{y_t}$ and $\gamma\sqrt{y_t}$ are introduced. Then, the Radon-Nikodym derivative can be expressed as

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\int_0^t(\pi\sqrt{y_s})dZ_{1,s}^{\mathbb{P}} - \int_0^t(\gamma\sqrt{y_s})dZ_{2,s}^{\mathbb{P}} - \frac{1}{2}\int_0^t(\pi\sqrt{y_s})^2ds - \frac{1}{2}\int_0^t(\gamma\sqrt{y_s})^2ds} \tag{107}$$

Thereafter, the Girsanov theorem is applied to express $Z_1^{\mathbb{Q}}$ and $Z_2^{\mathbb{Q}}$ as

$$Z_1^{\mathbb{Q}} = Z_1^{\mathbb{P}} + \int_0^t (\pi\sqrt{y_s})ds \tag{108}$$

$$Z_2^{\mathbb{Q}} = Z_2^{\mathbb{P}} + \int_0^t (\gamma\sqrt{y_s})ds \tag{109}$$

Consequently,

$$dZ_1^{\mathbb{P}} + (\pi\sqrt{y_t})dt = dZ_1^{\mathbb{Q}}$$
$$dZ_1^{\mathbb{P}} + (\pi\sqrt{y_t})dt - (\pi\sqrt{y_t})dt = dZ_1^{\mathbb{Q}} - (\pi\sqrt{y_t})dt$$

$$dZ_1^{\mathbb{P}} = dZ_1^{\mathbb{Q}} - (\pi\sqrt{y_t})dt \tag{110}$$

Similarly,

$$\begin{aligned} dZ_2^{\mathbb{P}} + (\gamma\sqrt{y_t})dt &= dZ_2^{\mathbb{Q}} \\ dZ_2^{\mathbb{P}} + (\gamma\sqrt{y_t})dt - (\gamma\sqrt{y_t})dt &= dZ_2^{\mathbb{Q}} - (\gamma\sqrt{y_t})dt \\ dZ_2^{\mathbb{P}} &= dZ_2^{\mathbb{Q}} - (\gamma\sqrt{y_t})dt \end{aligned} \tag{111}$$

Next, (110) and (111) are substituted into (27) and (28). As a consequence, (27) and (28) can be rewritten as

$$\begin{aligned} dr = \kappa_1(\theta_1 - r)dt + \sqrt{y}dZ_1^{\mathbb{P}} &= \kappa_1(\theta_1 - r)dt + \sqrt{y}\Big(dZ_1^{\mathbb{Q}} - (\pi\sqrt{y_t})dt\Big) \\ &= \Big(\kappa_1(\theta_1 - r) - (\pi\sqrt{y_t})\sqrt{y}\Big)dt + \sqrt{y}dZ_1^{\mathbb{Q}} \\ dr &= \Big(\kappa_1(\theta_1 - r) - \lambda_1 y\Big)dt + \sqrt{y}dZ_1^{\mathbb{Q}}, \quad \text{with } \lambda_1 = \pi \end{aligned} \tag{112}$$

$$\begin{aligned} dy = \kappa_2(\theta_2 - y)dt + v\sqrt{y}\ \Big(\rho\big(dZ_1^{\mathbb{Q}} - (\pi\sqrt{y_t})dt\big) + \sqrt{1-\rho^2}\big(dZ_2^{\mathbb{Q}} - (\gamma\sqrt{y_t})dt\big)\Big) \\ = \Big(\kappa_2(\theta_2 - y) - v\sqrt{y}\rho(\pi\sqrt{y_t}) - v\sqrt{y}\sqrt{1-\rho^2}(\gamma\sqrt{y_t})\Big)dt + v\sqrt{y}\rho dZ_1^{\mathbb{Q}} + v\sqrt{y}\sqrt{1-\rho^2}dZ_2^{\mathbb{Q}} \\ = \Big(\kappa_2(\theta_2 - y) - vy(\rho\pi + \sqrt{1-\rho^2}\gamma)\Big)dt + v\sqrt{y}\rho dZ_1^{\mathbb{Q}} + v\sqrt{y}\sqrt{1-\rho^2}dZ_2^{\mathbb{Q}} \\ dy = \Big(\kappa_2(\theta_2 - y) - \lambda_2 vy\Big)dt + v\sqrt{y}\Big(\rho dZ_1^{\mathbb{Q}} + \sqrt{1-\rho^2}dZ_2^{\mathbb{Q}}\Big), \quad \text{with } \lambda_2 = \rho\pi + \sqrt{1-\rho^2}\gamma \end{aligned} \tag{113}$$

In the next step, Ito's lemma is applied to find the differential of $P = P(\tau, r, y)$:

$$\begin{aligned} \Big(-\frac{\partial P}{\partial \tau} + \big(\kappa_1(\theta_1 - r) - \lambda_1 y\big)\ \frac{\partial P}{\partial r} + \big(\kappa_2(\theta_2 - y) - \lambda_2 vy\big)\ \frac{\partial P}{\partial y} + \frac{y}{2}\frac{\partial^2 P}{\partial r^2} + \frac{v^2 y}{2}\frac{\partial^2 P}{\partial y^2} \\ + \rho v y \frac{\partial^2 P}{\partial r \partial y}\ \Big)dt + \sqrt{y}\frac{\partial P}{\partial r}dZ_1^{\mathbb{Q}} + v\sqrt{y}\frac{\partial P}{\partial y}\Big(\rho dZ_1^{\mathbb{Q}} + \sqrt{1-\rho^2}dZ_2^{\mathbb{Q}}\Big) \end{aligned}$$

The discounted bond prices have to be martingales under the risk-neutral probability measure $\mathbb{Q}$. Hence, the Discounted Feynman-Kac Theorem can be applied. The drift term must vanish for the process to be a martingale. Then, the partial differential equation can be expressed as

$$\begin{aligned} -\frac{\partial P}{\partial \tau} + \big(\kappa_1(\theta_1 - r) - \lambda_1 y\big)\ \frac{\partial P}{\partial r} + \big(\kappa_2(\theta_2 - y) - \lambda_2 vy\big)\ \frac{\partial P}{\partial y} \\ + \frac{y}{2}\frac{\partial^2 P}{\partial r^2} + \frac{v^2 y}{2}\frac{\partial^2 P}{\partial y^2} + \rho v y \frac{\partial^2 P}{\partial r \partial y} - rP = 0 \end{aligned} \tag{114}$$

It can be noted that (114) corresponds to (29).

Appendix C The derivation of (31), (32) and (33)

The starting point is (29) and (30). Then, the following first-order and second-order partial derivatives need to be defined:

$$\frac{\partial P}{\partial \tau} = A'(\tau)\ e^{-rB(\tau)-yC(\tau)} + A(\tau)\big(-rB'(\tau) - yC'(\tau)\big)\ e^{-rB(\tau)-yC(\tau)} \tag{115}$$

$$\frac{\partial P}{\partial r} = -A(\tau)B(\tau)\ e^{-rB(\tau)-yC(\tau)} \tag{116}$$

$$\frac{\partial P}{\partial y} = -A(\tau)C(\tau)\ e^{-rB(\tau)-yC(\tau)} \tag{117}$$

$$\frac{\partial^2 P}{\partial r^2} = A(\tau)B^2(\tau)\ e^{-rB(\tau)-yC(\tau)} \tag{118}$$

$$\frac{\partial^2 P}{\partial y^2} = A(\tau)C^2(\tau)\ e^{-rB(\tau)-yC(\tau)} \tag{119}$$

$$\frac{\partial^2 P}{\partial r \partial y} = A(\tau)B(\tau)C(\tau)\ e^{-rB(\tau)-yC(\tau)} \tag{120}$$

Next, all these first-order and second-order partial derivatives are substituted into (29). Consequently, (29) can be rewritten as

$$\begin{aligned}\Big(&-A'(\tau) - A(\tau)\big(-rB'(\tau) - yC'(\tau)\big) - A(\tau)B(\tau)\big(\kappa_1(\theta_1 - r) - \lambda_1 y\big)\\ &-A(\tau)C(\tau)\big(\kappa_2(\theta_2 - y) - \lambda_2 v y\big) + \frac{y}{2}A(\tau)B^2(\tau) + \frac{v^2 y}{2}A(\tau)C^2(\tau)\\ &+\rho v y A(\tau)B(\tau)C(\tau) - rA(\tau)\Big)\ e^{-rB(\tau)-yC(\tau)} = 0\end{aligned}$$

$$\begin{aligned}&-A'(\tau) + A(\tau)C'(\tau)y - A(\tau)B(\tau)(\kappa_1\theta_1 - \lambda_1 y) - A(\tau)C(\tau)(\kappa_2\theta_2 - \kappa_2 y)\\ &+A(\tau)C(\tau)\lambda_2 v y + \frac{y}{2}A(\tau)B^2(\tau) + \frac{v^2 y}{2}A(\tau)C^2(\tau) + \rho v y A(\tau)B(\tau)C(\tau)\\ &= \Big(-B'(\tau) - B(\tau)\kappa_1 + 1\Big)rA(\tau)\end{aligned}$$

Setting the right-hand side equal to zero results in

$$B'(\tau) = 1 - \kappa_1 B(\tau) \tag{121}$$

Setting the left-hand side equal to zero leads to

$$\begin{aligned} &-A'(\tau) - A(\tau)B(\tau)\kappa_1\theta_1 - A(\tau)C(\tau)\kappa_2\theta_2 \\ &+\Big(C'(\tau) + B(\tau)\lambda_1 + C(\tau)\kappa_2 + C(\tau)\lambda_2 v \\ &+\frac{1}{2}B^2(\tau) + \frac{v^2}{2}C^2(\tau) + \rho v B(\tau)C(\tau)\Big)A(\tau)y = 0 \end{aligned}$$

Setting the first and second part of the left-hand side of this expression equal to zero results in

$$-A'(\tau) - A(\tau)B(\tau)\kappa_1\theta_1 - A(\tau)C(\tau)\kappa_2\theta_2 = 0$$

$$C'(\tau) + B(\tau)\lambda_1 + C(\tau)\kappa_2 + C(\tau)\lambda_2 v + \frac{1}{2}B^2(\tau) + \frac{v^2}{2}C^2(\tau) + \rho v B(\tau)C(\tau) = 0$$

$$A'(\tau) = -A(\tau)\Big(\kappa_1\theta_1 B(\tau) + \kappa_2\theta_2 C(\tau)\Big) \tag{122}$$

$$C'(\tau) = -\lambda_1 B(\tau) - \kappa_2 C(\tau) - \lambda_2 v C(\tau) - \frac{1}{2}B^2(\tau) - \frac{v^2}{2}C^2(\tau) - \rho v B(\tau)C(\tau) \tag{123}$$

It can be noted that (122), (121) and (123) correspond to (31), (32) and (33) respectively.

Appendix D The derivation of (58)

The starting point is (52), (53), (54), (55), (56) and (57). In the first step of the solution procedure, the part of (57) belonging to $\frac{1}{\epsilon}$ will be set equal to zero:

$$\Pi_0 P_0 = 0 \tag{124}$$

Π_0 contains partial derivatives with respect to y. Therefore, P_0 does not depend on y, which indicates that $P_0 = P_0(t, r)$. In the second step, the part of (57) corresponding to $\frac{1}{\sqrt{\epsilon}}$ will be set equal to zero:

$$\Pi_0 P_1 + \Pi_1 P_0 = 0 \tag{125}$$

Π_1 includes partial derivatives with respect to y and both r and y. P_0 does not depend on y. As a consequence, $\Pi_1 P_0 = 0$. Then, $\Pi_0 P_1 = 0$ as well, which indicates that $P_1 = P_1(t, r)$.

So far, it has been shown that $P_0 + \sqrt{\epsilon}P_1$ does not depend on y.

In the next step, the part of (57) belonging to 1 will be set equal to zero:

$$\Pi_0 P_2 + \Pi_1 P_1 + \Pi_2 P_0 = 0 \tag{126}$$

Π_1 includes partial derivatives with respect to y and both r and y. P_1 does not depend on y. Hence, $\Pi_1 P_1 = 0$. Then, (126) reduces to

$$\Pi_0 P_2 + \Pi_2 P_0 = 0 \tag{127}$$

When r is fixed and the focus is only on the dependence on y, Π_0 and y take the form of a Poisson equation. This is a partial differential equation of elliptic type. A border condition for solving Poisson equations is required in order to continue.

Lemma 4. *consider* $\Pi_0 P + g = 0$*, with* Π_0 *defined as in* (53)*. A necessary condition in order for* P *to have a solution is that the function* g *satisfies* $\langle g \rangle = 0$*, with* $\langle g \rangle = \int g(y) f(y) dy = 0$*, where* $f(y)$ *is the limiting distribution of* y*.*

The border condition for (127) requires $\langle \Pi_2 P_0 \rangle = 0$. P_0 does not depend on y. Hence, $\langle \Pi_2 \rangle P_0 = 0$ with

$$\langle \Pi_2 \rangle = \frac{\partial}{\partial t} + \big(\kappa_1(\theta_1 - r) - \lambda_1 \langle y \rangle\big)\frac{\partial}{\partial r} + \frac{\langle y \rangle}{2}\frac{\partial^2}{\partial r^2} - r \tag{128}$$

Then, $\langle y \rangle = \theta_2$ according to (48). Substituting this value into the border condition of the Poisson equation results in a partial differential equation that is defined by

$$\langle \Pi_2 \rangle P_0 = \frac{\partial P_0}{\partial t} + \Big(\kappa_1(\theta_1 - r) - (\lambda_1\sqrt{\theta_2})\sqrt{\theta_2}\Big)\frac{\partial P_0}{\partial r} + \frac{\theta_2}{2}\frac{\partial^2 P_0}{\partial r^2} - r P_0 \tag{129}$$

This partial differential equation corresponds to (93), which is the partial differential equation of the Vasicek model. $\sigma(r,t) = \sqrt{\theta_2}$ and the market price of risk $\lambda(r,t) = \lambda_1\sqrt{\theta_2}$. The results derived in (94), (95), (96), (97) and (98) will be applied here to (129), while taking into account that $\frac{\partial P_0}{\partial t} = -\frac{\partial P_0}{\partial \tau}$

$$P_0(\tau, r) = A(\tau)e^{-B(\tau)r} \tag{130}$$

$$B'(\tau) = 1 - \kappa_1 B(\tau) \tag{131}$$

$$A'(\tau) = A(\tau)B(\tau)(\lambda_1\theta_2 - \kappa_1\theta_1) + \frac{\theta_2}{2}A(\tau)B^2(\tau) \tag{132}$$

$$B(\tau) = \frac{1}{\kappa_1}\big(1 - e^{-\kappa_1\tau}\big) \tag{133}$$

$$A(\tau) = \exp\ \Big[\ \Big(\theta_1 - \frac{\lambda_1\theta_2}{\kappa_1} - \frac{\theta_2}{2\kappa_1^2}\Big)\ \big(B(\tau) - \tau\big) - \frac{\theta_2}{4\kappa_1}B^2(\tau)\ \Big] \tag{134}$$

$P_0(\tau, r)$ is the price approximated by the Vasicek model. Moreover, while returning back to the border condition $\langle \Pi_2 P_0\rangle = 0$, $\Pi_2 P_0$ can be rewritten as

$$\Pi_2 P_0 = \Pi_2 P_0 - \langle \Pi_2 P_0\rangle = \big(\Pi_2 - \langle \Pi_2\rangle\big)P_0, \tag{135}$$

with

$$\Pi_2 - \langle \Pi_2\rangle = \frac{1}{2}\big(y - \langle y\rangle\big)\frac{\partial^2}{\partial r^2} - \lambda_1\big(y - \langle y\rangle\big)\frac{\partial}{\partial r} \tag{136}$$

Next, P_2 in (127) can be rewritten as

$$P_2(t, r, y) = -\Pi_0^{-1}(\Pi_2 P_0) = \Pi_0^{-1}\big(\Pi_2 - \langle \Pi_2\rangle\big)P_0 + c(t, r), \tag{137}$$

with $c(t, r)$ a constant function with respect to y.

In the final step, the part of (57) belonging to $\sqrt{\epsilon}$ will be set equal to zero:

$$\Pi_0 P_3 + \Pi_1 P_2 + \Pi_2 P_1 = 0 \tag{138}$$

Equation (138) also takes the form of a Poisson equation for the function P_3 with respect to Π_0. The border condition for (138) requires $\langle \Pi_1 P_2 + \Pi_2 P_1\rangle = 0$. This expression can be rewritten using the linearity property of expectation:

$$\langle \Pi_2 P_1\rangle = -\langle \Pi_1 P_2\rangle \tag{139}$$

Then, by using (137), $\langle \Pi_2\rangle P_1$ can be expressed as

$$\begin{aligned}\langle \Pi_2\rangle P_1 = \langle \Pi_2 P_1\rangle = -\langle \Pi_1 P_2\rangle &= \Big\langle \Pi_1\big(\Pi_0^{-1}\big(\Pi_2 - \langle \Pi_2\rangle\big)P_0 + c(t, r)\big)\Big\rangle \\ &= \Big\langle \Pi_1\Pi_0^{-1}\big(\Pi_2 - \langle \Pi_2\rangle\big)\Big\rangle P_0 = \Upsilon P_0\end{aligned} \tag{140}$$

In the derivation it is used that $\Pi_1 c(t, r) = 0$, because Π_1 contains partial derivatives with respect to y in both parts.
At this point it is convenient to introduce the function $\Phi(y)$, which is the solution of the Poisson equation

$$\Pi_0\Phi = y - \langle y\rangle \tag{141}$$

The border condition is satisfied, since $\big\langle y - \langle y\rangle\big\rangle = \langle y\rangle - \langle y\rangle = 0$. Substituting (141) into (136) and (136) into Υ results in

$$\Upsilon = \Big\langle \Pi_1\Pi_0^{-1}\big(\Pi_2 - \langle \Pi_2\rangle\big)\Big\rangle = \Big\langle \Pi_1\Pi_0^{-1}\Pi_0\Phi\Big(\frac{1}{2}\frac{\partial^2}{\partial r^2} - \lambda_1\frac{\partial}{\partial r}\Big)\Big\rangle = \langle \Pi_1\Phi\rangle\Big(\frac{1}{2}\frac{\partial^2}{\partial r^2} - \lambda_1\frac{\partial}{\partial r}\Big) \tag{142}$$

Then, (54) is substituted into (142):

$$\begin{aligned}\Upsilon &= \Big\langle \Big(-\lambda_2 v y \frac{\partial}{\partial y} + \rho v y \frac{\partial^2}{\partial r \partial y}\Big)\Phi \Big\rangle \Big(\frac{1}{2}\frac{\partial^2}{\partial r^2} - \lambda_1 \frac{\partial}{\partial r}\Big) \\ &= \lambda_1\lambda_2 v\langle y\Phi'\rangle \frac{\partial}{\partial r} - \Big(\frac{1}{2}\lambda_2 v + \lambda_1 \rho v\Big)\langle y\Phi'\rangle \frac{\partial^2}{\partial r^2} + \frac{1}{2}\rho v \langle y\Phi'\rangle \frac{\partial^3}{\partial r^3}\end{aligned} \tag{143}$$

Υ can be rewritten as

$$\Upsilon = V_1 \frac{\partial}{\partial r} + V_2 \frac{\partial^2}{\partial r^2} + V_3 \frac{\partial^3}{\partial r^3}, \tag{144}$$

with

$$V_1 = \lambda_1\lambda_2 v\langle y\Phi'\rangle \tag{145}$$

$$V_2 = -(\frac{1}{2}\lambda_2 v + \lambda_1 \rho v)\langle y\Phi'\rangle \tag{146}$$

$$V_3 = \frac{1}{2}\rho v \langle y\Phi'\rangle \tag{147}$$

Remember that the aim is to calculate P_1, which satisfies (140). Hence, substituting (128), (130) and (144) into (140) results in a partial differential equation for $P_1(t,r)$ that is defined by

$$\begin{aligned}&-\frac{\partial P_1}{\partial \tau} + \big(\kappa_1(\theta_1 - r) - \lambda_1\theta_2\big)\frac{\partial P_1}{\partial r} + \frac{1}{2}\theta_2 \frac{\partial^2 P_1}{\partial r^2} - rP_1 \\ &-A(\tau)e^{-B(\tau)r}\big(-V_1 B(\tau) + V_2 B^2(\tau) - V_3 B^3(\tau)\big) = 0,\end{aligned} \tag{148}$$

where $A(\tau)$ and $B(\tau)$ satisfy (133) and (134). The solution of this partial differential equation can be expressed as

$$P_1(\tau, r) = D(\tau)A(\tau)\ e^{-rB(\tau)}, \tag{149}$$

where $D(0) = 0$, which corresponds to $P_1(0,r) = 0$. Equation (149) is used to substitute partial derivatives into (148). Besides, (131) and (132) are taken into account. Then, the partial differential equation of $D(\tau)$ can be defined as

$$D'(\tau) = V_1 B(\tau) - V_2 B^2(\tau) + V_3 B^3(\tau) \tag{150}$$

Integrating (150) results in

$$D(\tau) = \frac{V_1}{\kappa_1}\int_0^\tau \Big(1 - e^{-\kappa_1\tau}\Big)d\tau - \frac{V_2}{\kappa_1^2}\int_0^\tau \Big(1 - e^{-\kappa_1\tau}\Big)^2 d\tau + \frac{V_3}{\kappa_1^3}\int_0^\tau \Big(1 - e^{-\kappa_1\tau}\Big)^3 d\tau \tag{151}$$

Evaluating (151) results in

$$\begin{aligned}D(\tau) = &\frac{V_1}{\kappa_1}\Big(-B(\tau) + \tau\Big) - \frac{V_2}{\kappa_1^2}\Big(-B(\tau) - \frac{\kappa_1}{2}B^2(\tau) + \tau\Big) \\ &+\frac{V_3}{\kappa_1^3}\Big(-B(\tau) - \frac{\kappa_1}{2}B^2(\tau) - \frac{\kappa_1^2}{3}B^3(\tau) + \tau\Big)\end{aligned} \tag{152}$$

Therefore, the price of a zero-coupon bond could be approximated by

$$P^{\epsilon}(\tau, r, y) \approx P_0(\tau, r) + \sqrt{\epsilon} P_1(\tau, r) = \big(1 + \sqrt{\epsilon} D(\tau)\big) A(\tau)\ e^{-rB(\tau)}, \tag{153}$$

where $A(\tau), B(\tau)$ and $D(\tau)$ are defined in (134), (133) and (152). It can be noted that (153) corresponds to (58).

References

Andersen, T. G., Lund, J.: Estimating continuous-time stochastic volatility models of short-term interest rate, Journal of Econometrics, 77, pages 343-377, 1997.

Balter, A.G., Pelsser, A.A.J., & Schotman, P.C. (2019). What does a term structure model imply about very long-term interest rates?

Brown, R. and S. Schaefer (2000): Why Long Term Forward Interest Rates (Almost) Always Slope Downwards, Working Paper, London Business School.

Cox, J.C., Ingersoll, J.E., Ross, A. A Theory of the Term Structure of Interest Rates, Econometrica, Vol. 53, p.385-408, 1985.

Dijsselbloem, J.R.V.A., De Waegenaere, A.M.B., van Ewijk, C., van der Horst, A., Knoef, M.G., Steenbeek, O.W., & Werker, B.J.M. (2019). Advies Commissie Parameters.

Committee of European Insurance and Occupational Pensions Supervisors (2010): "QIS5: Risk Free Interest Rates - Extrapolation Method," eiopa.europa.eu.

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

Fong, H. G., Vasicek, O. A. Fixed-Income Volatility Management, Journal of Portfolio Management,17, pages 41-46, 1991.

Fouque, J.P, Papanicolaou, G., Sircar, R. Mean-Reverting Stochastic Volatility, International Journal of Theoretical and Applied Finance, 3, pages 101-142, 2000.

Fraser, D.A.S., Rekkas, M., & Wong, A. Highly Accurate Likelihood Analysis for the Seemingly Unrelated Regression Problem. Retrieved March 7, 2020, from `http://www.sfu.ca/~mrekkas/surv3.pdf`

Langejan, T.W., van Ewijk, C., Nijman, T.E., Pelsser, A.A.J., Sleijpen, O.H.C.M., & Steenbeek, O.W. (2013). Advisory Report of the UFR Committee.

Longstaff, F., Schwartz, E. Interest rate volatility and the term structure: a two factor general equilibrium model, Journal of Finance, pages 1259-1282, 47, 1992.

Svoboda, S. (2004). Interest rate modelling (Finance and Capital Market Series). United States of America: Palgrave Macmillan.