Machine learning Vasicek model calibration with Gaussian processes

J. Beleza Sousa[*,a,b], M. L. Esquível[b], R. M. Gaspar[c]

[a] M2A/DEETC, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal
[b] CMA/DM, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal
[c] Advance Research Center, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Rua Miguel Lupi 20, 1249-078 Lisboa, Portugal

Abstract

In this paper we calibrate the Vasicek interest rate model under the risk neutral measure by learning the model parameters using Gaussian processes for machine learning regression. The calibration is done by maximizing the likelihood of zero coupon bond log prices, using zero coupon bond log prices mean and covariance functions computed analytically, as well as likelihood derivatives with respect to the parameters. The maximization method used is the conjugate gradients. The only prices needed for calibration are zero coupon bond prices and the parameters are directly obtained in the arbitrage free risk neutral measure.

Key words: Vasicek interest rate model, Arbitrage free risk neutral measure, Calibration, Gaussian processes for machine learning, Zero coupon bond prices

1. Introduction

Calibration of interest rate models typically entails the availability of some derivatives such as swaps, caps or swaptions.

In this paper we present an alternative method for calibrating Gaussian models, namely, the Vasicek interest rate model (Vasicek, 1977), which requires zero coupon bond prices only.

The presented method has the following features:

- The only prices needed for calibration are zero coupon bond prices.

- All the model parameters are directly obtained in the risk neutral measure.

---

[*]Corresponding author.

Email addresses: `jsousa@deetc.isel.ipl.pt` (J. Beleza Sousa), `mle@fct.unl.pt` (M. L. Esquível), `rmgaspar@iseg.utl.pt` (R. M. Gaspar)

- The calibration method does not require a discrete model approximation nor the establishment of an objective measure dynamics.

The method is based on Gaussian processes for Machine Learning, and its main drawback is his applicability to Gaussian models only.

One key issue in using Gaussian processes for machine learning is to have enough prior information on the data, in order to specify mean and covariance functions. Under the Vasicek interest rate model, the risk neutral zero coupon bond prices follow a log normal distribution, which can easily be transformed into a Gaussian process by taking the logarithm of the zero coupon prices. The mean and covariance functions of this Gaussian process can be computed analytically making it suitable for Gaussian processes for machine learning regression.

2. Vasicek interest rate model

Under the Vasicek model, the interest rate follows an Ornstein-Uhlenbeck mean-reverting process defined by the stochastic differential equation

$$dr(t) = k(\theta - r(t))dt + \sigma dW(t) \tag{1}$$

where k is the mean reversion velocity, θ is the mean interest rate level, σ is the volatility and $W(t)$ the Wiener process, and k and σ are positive.

Let $s \leq t$. The solution of equation 1 is (Brigo & Mercurio, 2006)

$$r(t) = r(s)e^{-k(t-s)} + \theta \left(1 - e^{-k(t-s)}\right) + \sigma e^{-kt} \int_s^t e^{ku} dW(u). \tag{2}$$

The interest rate $r(t)$, conditioned on $\mathcal{F}_s$, is normally distributed with mean

$$E\{r(t)|\mathcal{F}_s\} = r(s)e^{-k(t-s)} + \theta \left(1 - e^{-k(t-s)}\right) \tag{3}$$

and variance

$$Var\{r(t)|\mathcal{F}_s\} = \frac{\sigma^2}{2k}\left(1 - e^{-2k(t-s)}\right).$$

The model as an affine term structure, which means that the zero coupon bond prices $p(t, T)$, with maturity T are given by (Björk, 2004)

$$p(t, T) = e^{A(t,T) - B(t,T)r(t)} \tag{4}$$

where

$$A(t, T) = \left(\theta - \frac{\sigma^2}{2k^2}\right)(B(t, T) - T + t) - \frac{\sigma^2}{4k}B(t, T)$$

and

$$B(t, T) = \frac{1}{k}\left(1 - e^{-k(T-t)}\right).$$

Equation 4 shows that the zero coupon bond prices $p(t, T)$ are log normal and consequently $\log(p(t, T))$ are normal.

2

2.1. Zero coupon bond log prices mean function

Since

$$\log(p(t,T)) = A(t,T) - B(t,T)r(t) \tag{5}$$

the mean function $\mu(t,T)$ of $\log(p(t,T))$ is given by

$$
\begin{aligned}
\mu(t,T) &= E\left\{\log(p(t,T))|\mathcal{F}_s\right\} \\
&= E\left\{A(t,T) - B(t,T)r(t)|\mathcal{F}_s\right\} \\
&= A(t,T) - B(t,T)E\left\{r(t)|\mathcal{F}_s\right\}
\end{aligned}
$$

Considering the initial instant $s = 0$, and using equation 3 for $E\left\{r(t)|\mathcal{F}_s\right\}$ we get

$$
\begin{aligned}
\mu(t,T) &= A(t,T) - B(t,T)\left(r_0 e^{-kt} + \theta\left(1 - e^{-kt}\right)\right) \\
&= \frac{1}{4k^3}e^{-kT}\left(-4\left(e^{k(-t+T)} - 1\right)k^2(r_0 - \theta)\right. \\
&\quad -4e^{kT}k^3(T-t)\theta + 2(e^{kt} - e^{kT})\sigma^2 \\
&\quad \left. +k\left(e^{kt} + e^{kT}(2T - 2t - 1)\right)\sigma^2\right)
\end{aligned} \tag{6}
$$

where r_0 stands for the initial interest rate value, the value of the interest rate $r(t)$, at $t = 0$.

2.2. Zero coupon bond log prices covariance function

The covariance function $cov(t_1, t_2, T)$ of $\log(p(t,T))$ is given by

$$
\begin{aligned}
cov(t_1, t_2, T) &= E\left\{(\log(p(t_1,T)) - \mu(t_1,T))\right. \\
&\quad \left.(\log(p(t_2,T)) - \mu(t_2,T))|\mathcal{F}_s\right\} \\
&= E\left\{\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\right\} - \mu(t_1,T)\mu(t_2,T) \quad (7)
\end{aligned}
$$

Using equation 5, the term $E\left\{\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\right\}$, is given by

$$
\begin{aligned}
E\{&\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\} \\
&= E\{(A(t_1,T) - B(t_1,T)r(t_1)) \\
&\quad (A(t_2,T) - B(t_2,T)r(t_2))|\mathcal{F}_s\} \\
&= A(t_1,T)A(t_2,T) \\
&\quad -A(t_1,T)B(t_2,T)E\{r(t_2)|\mathcal{F}_s\} \\
&\quad -B(t_1,T)A(t_2,T)E\{r(t_1)|\mathcal{F}_s\} \\
&\quad +B(t_1,T)B(t_2,T)E\{r(t_1)r(t_2)|\mathcal{F}_s\} \\
&= A(t_1,T)A(t_2,T) \\
&\quad -A(t_1,T)B(t_2,T)\mu(t_2,T) \\
&\quad -B(t_1,T)A(t_2,T)\mu(t_1,T) \\
&\quad +B(t_1,T)B(t_2,T)E\{r(t_1)r(t_2)|\mathcal{F}_s\} \quad (8)
\end{aligned}
$$

Using the Vasicek SDE solution equation 2, the term $E\{r(t_1)r(t_2)|\mathcal{F}_s\}$ is given by

$$E\{r(t_1)r(t_2)|\mathcal{F}_s\}$$

$$= E\left\{\left(r_0 e^{-kt_1} + \theta\left(1 - e^{-kt_1}\right) + \sigma e^{-kt_1}\int_0^{t_1} e^{ku}dW(u)\right)\right.$$

$$\left.\left(r_0 e^{-kt_2} + \theta\left(1 - e^{-kt_2}\right) + \sigma e^{-kt_2}\int_0^{t_2} e^{ku}dW(u)\right)|\mathcal{F}_s\right\}$$

$$= r_0^2 e^{-k(t1+t2)} + r_0 e^{-kt_1}\theta\left(1 - e^{-kt_2}\right)$$

$$+ \theta\left(1 - e^{-kt_1}\right)r_0 e^{-kt_2} + \theta^2\left(1 - e^{-kt_1}\right)\left(1 - e^{-kt_2}\right)$$

$$+ \sigma^2 e^{-k(t1+t2)}E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\} \qquad (9)$$

In order to compute $E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\}$, we first consider $t_1 < t_2$. In this case, we have

$$E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{t_1} e^{ku}dW(u)\right)\left(\int_0^{t_1} e^{ku}dW(u) + \int_{t_1}^{t_2} e^{ku}dW(u)\right)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{t_1} e^{ku}dW(u)\right)^2 |\mathcal{F}_s\right\}. \qquad (10)$$

In case $t_2 < t_1$, we have

$$E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{t_2} e^{ku}dW(u) + \int_{t_2}^{t_1} e^{ku}dW(u)\right)\left(\int_0^{t_2} e^{ku}dW(u)\right)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{t_2} e^{ku}dW(u)\right)^2 |\mathcal{F}_s\right\}. \qquad (11)$$

Given equations 10 and 11, we get

$$E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{min(t_1,t_2)} e^{ku}dW(u)\right)^2 |\mathcal{F}_s\right\}$$

Finally, using Itô isometry

$$E\left\{\int_0^{t_1} e^{ku}dW(u)\int_0^{t_2} e^{ku}dW(u)|\mathcal{F}_s\right\}$$

$$= E\left\{\left(\int_0^{min(t_1,t_2)} e^{ku}dW(u)\right)^2 |\mathcal{F}_s\right\}$$

$$= \int_0^{min(t_1,t_2)} E\left\{\left(e^{ku}\right)^2\right\}du$$

$$= \int_0^{min(t_1,t_2)} e^{2ku}du$$

$$= \frac{1}{2k}\left(e^{2kmin(t_1,t_2)}-1\right). \tag{12}$$

Using equations 6, 7, 8, 9 and 12, the covariance function $cov(t_1, t_2, T)$ of $\log(p(t, T))$ is given by

$$
\begin{aligned}
cov(t_1, t_2, T) &= \frac{1}{2k^3}e^{-k(2T+t1+t2)}\left(e^{2kmin(t1,t2)}-1\right)\\
&\quad \left(e^{kT}-e^{kt1}\right)\left(e^{kT}-e^{kt2}\right)\sigma^2
\end{aligned}
\tag{13}
$$

3. Gaussian processes for machine learning

The goal of Gaussian processes for machine learning is to find the non linear unknown mapping $y = f(\mathbf{x})$, from data $(\mathbf{X}, \mathbf{y})$, using Gaussian distributions over functions (Rasmussen & Williams, 2005)

$$\mathcal{GP} \sim \mathcal{N}(\mu(\mathbf{x}), cov(\mathbf{x_1}, \mathbf{x_2})).$$

The pair $(\mathbf{X}, \mathbf{y})$ is the training set. The matrix $\mathbf{X}$ collects a set of n vectors $\mathbf{x}$ where the value $y = f(\mathbf{x})$ was observed. The corresponding y values are collected in vector $\mathbf{y}$.

The set of vectors $\mathbf{x}^\star$ where the values $y^\star = f(\mathbf{x}^\star)$ were not observed, is collected in matrix $\mathbf{X}^\star$. The matrix $\mathbf{X}^\star$ is the test set.

Under the Vasicek interest rate model the zero coupon bonds log prices $\log(p(t, T))$ are normal

$$\mathcal{GP} \sim \mathcal{N}(\mu(t, T), cov(t_1, t_2, T))$$

where $\mu(t, T)$ is given by equation 6 and $cov(t_1, t_2, T)$ is given by equation 13.

Since T, the bond maturity, is a bond feature, in this case the mapping we are interested in is the scalar mapping

$$y = f(t)$$

where y stands for the zero coupon bonds log prices. This reduces the training set to the pair of vectors $(\mathbf{t}, \mathbf{y})$, and the test set to vector $\mathbf{t}^\star$.

5

Since the process is Gaussian (Rasmussen & Williams, 2005)

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^\star \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_\star \\ \mathbf{K}_\star^T & \mathbf{K}_{\star\star} \end{bmatrix}\right)$$

and

$$p(\mathbf{y}^\star|\mathbf{t}^\star, \mathbf{t}, \mathbf{y}) \sim \mathcal{N}\left(\boldsymbol{\mu}^\star + \mathbf{K}_\star^T \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \mathbf{K}_{\star\star} - \mathbf{K}_\star^T \mathbf{K}^{-1} \mathbf{K}_\star\right)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^\star$ are mean vectors of train and test sets, $\mathbf{K}$ is the train set covariance matrix, $\mathbf{K}_\star$ the train-test covariance matrix and $\mathbf{K}_{\star\star}$ the test set covariance matrix.

The conditional distribution

$$p(\mathbf{y}^\star|\mathbf{t}^\star, \mathbf{t}, \mathbf{y})$$

corresponds to the posterior process on the data

$$\mathcal{GP}_\mathcal{D} \sim \mathcal{N}(m_\mathcal{D}(t), cov_\mathcal{D}(t_1, t_2))$$

where

$$m_\mathcal{D}(t) = m(t) + \mathbf{K}_{\mathbf{t},t}^T \mathbf{K}^{-1}(\mathbf{y} - \mu) \tag{14}$$

and

$$cov_\mathcal{D}(t_1, t_2) = cov(t_1, t_2) - \mathbf{K}_{\mathbf{t},t_1}^T \mathbf{K}^{-1} \mathbf{K}_{\mathbf{t},t_2} \tag{15}$$

where $\mathbf{K}_{\mathbf{t},t}$ is a correlation vector between every training instant and t.

Equation 14 is the regression function while equation 15 is the regression confidence. Equations 14 and 15 are the central equations of Gaussian processes for machine learning regression.

In order to learn the model parameters $\Theta = \{r_0, k, \theta, \sigma\}$ from data, the likelihood of the training data given the parameters (closed form) (Rasmussen, 2004)

$$\begin{aligned} L &= \log p(\mathbf{y}|\mathbf{t}, \Theta) \\ &= -\frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{n}{2}\log(2\pi) \end{aligned}$$

is maximized, based on the derivatives of L with respect to each of the parameters (closed forms).

Note that, since we want to learn the parameters in the arbitrage free risk neutral measure, the initial interest rate value r_0, is considered a parameter, like k, θ and σ, to be learned from the zero coupon bond log prices.

Since

$$\frac{\partial}{\partial \Theta}\log|\mathbf{K}| = tr\left(\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta}\right)$$

and

$$\frac{\partial}{\partial \Theta}\mathbf{K}^{-1} = -\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta}\mathbf{K}^{-1}$$

the derivatives $\frac{\partial L}{\partial \Theta}$ are given by

$$
\begin{aligned}
\frac{\partial L}{\partial \Theta} \;=\; & -\frac{1}{2} tr\left(\mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta}\right) \\
& +\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}\right)^T \mathbf{K}^{-1}\frac{\partial \mathbf{K}}{\partial \Theta}\mathbf{K}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right) \\
& +\left(\mathbf{y}-\boldsymbol{\mu}\right)^T \mathbf{K}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \Theta}.
\end{aligned}
$$

In order to compute the vector of derivatives, $\frac{\partial \boldsymbol{\mu}}{\partial \Theta}$, and the matrix of derivatives $\frac{\partial \mathbf{K}}{\partial \Theta}$, the derivatives of the mean function $\mu(t,T)$ (equation 6), and the derivatives of the covariance function $cov(t_1,t_2,T)$ (equation 13) with respect to the parameters are used, namely:

$$
\begin{aligned}
\frac{\partial \mu(t,T)}{\partial r_0} \;=\; & \frac{e^{-kT}-e^{-kt}}{k}; \\
\frac{\partial \mu(t,T)}{\partial k} \;=\; & \frac{1}{4k^4}e^{-kT}\Big(4k^3(e^{k(T-t)}t-T)(r_0-\theta) \\
& -6(e^{kt}-e^{kT})\sigma^2 \\
& +2k(e^{kT}(2t-2T+1)+e^{kt}(t-T-1))\sigma^2 \\
& +k^2(-4r_0+4e^{k(T-t)}(r_0-\theta)+4\theta+e^{kt}t\sigma^2-e^{kt}T\sigma^2)\Big); \\
\frac{\partial \mu(t,T)}{\partial \theta} \;=\; & \frac{e^{-kt}-e^{-kT}+kt-kT}{k}; \\
\frac{\partial \mu(t,T)}{\partial \sigma} \;=\; & \frac{\left(e^{k(t-T)}(k+2)+2kT-2kt-k-2\right)\sigma}{2k^3}; \\
\frac{\partial cov(t_1,t_2,T)}{\partial r_0} \;=\; & 0; \\
\frac{\partial cov(t_1,t_2,T)}{\partial k} \;=\; & \frac{1}{2k^4}e^{-k(t1+t2+2T)}\Big(\\
& e^{k(t1+t2)}(3+2kT)+e^{2kT}(3+k(t1+t2)) \\
& -e^{k(t1+T)}(3+k(t2+T))-e^{k(t2+T)}(3+k(t1+T)) \\
& +e^{k(t1+2min(t1,t2)+T)}(3+k(t2-2min(t1,t2)+T)) \\
& +e^{k(t2+2min(t1,t2)+T)}(3+k(t1-2min(t1,t2)+T)) \\
& +e^{k(t1+t2+2min(t1,t2))}(-3+2k(min(t1,t2)-T)) \\
& +e^{2k(min(t1,t2)+T)}(-3-k(t1+t2-2min(t1,t2)))\Big)\sigma^2; \\
\frac{\partial cov(t_1,t_2,T)}{\partial \theta} \;=\; & 0; \\
\frac{\partial cov(t_1,t_2,T)}{\partial \sigma} \;=\; & \frac{1}{k^3}e^{-k(t1+t2+2T)}\left(e^{kT}-e^{kt1}\right)\left(e^{kT}-e^{kt2}\right) \\
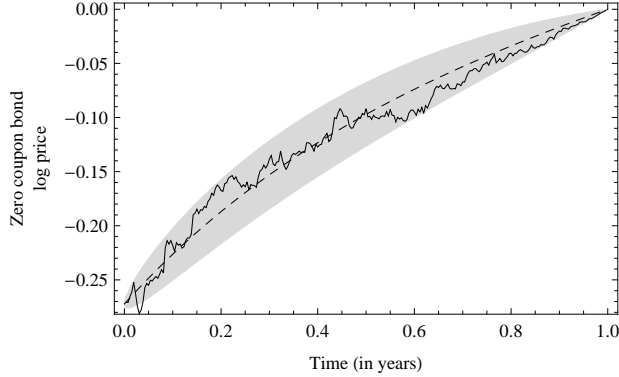& \left(e^{2kmin(t1,t2)}-1\right)\sigma.
\end{aligned}
$$

Figure 1: Zero coupon bond log prices simulated sequence (solid black), mean (dashed black) and two standard deviations interval (light gray).

Parameter	Value	Mean	StdDev.	95% CI		
r_0	0.5	0.527	0.482	-0.418	to	1.472
k	2.0	2.098	0.855	0.419	to	3.776
θ	0.1	0.083	0.443	-0.786	to	0.952
σ	0.2	0.203	0.039	0.126	to	0.280

Table 1: Parameters r_0, k, θ and σ, 1000 calibrations mean, standard deviation and 95% confidence interval.

4. Simulation results

In order to test the proposed calibration method we used equations 2 and 5, with fixed parameters values, to simulate 1000 sequences of zero coupon bond log prices.

The parameters values used were: initial interest rate $r_0 = 0.5$; mean interest rate level $\theta = 0.1$; mean reversion velocity $k = 2$; and volatility $\sigma = 0.2$.

We considered the zero coupon bond maturity of one year, $T = 1$, and simulated one year daily prices sequences by considering 260 prices per sequence (5 working days prices per week, 52 weeks per year).

Figure 1 illustrates a simulated sequence of zero coupon bond log prices, as well as the mean and variance functions.

We applied the calibration procedure by maximizing the likelihood of each one of the zero coupon bond log prices sequences, using Wolfram Mathematica 7 (Wolfram Research, 2009) conjugate gradients implementation with default configuration parameters.

Figure 2 illustrates the 50 bins parameters histograms obtained from the 1000 calibrations performed, and table 1 shows the corresponding mean, standard deviation and 95% confidence intervals.

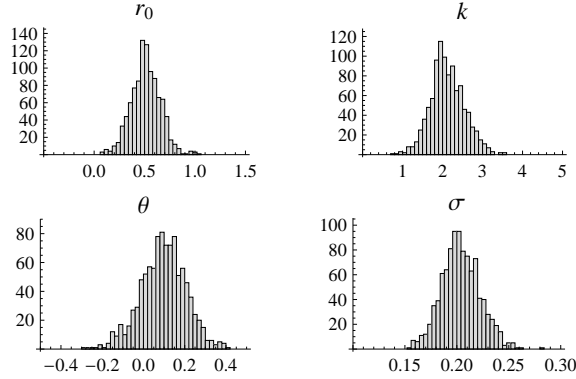As it can be observed, all parameters 95% confidence intervals contain the fixed parameter value used.

8

Figure 2: Learned parameters 50 bins histograms.

r_0	k	θ	σ
0.212	2.925	0.025	0.195

Table 2: Learned parameters r_0, k, θ and σ, for a real, two year maturity, zero coupon bond reference price, calibrated with approximately one year of available prices.

5. Calibration to real data

The great majority of the zero coupon bonds market is over the counter (OTC). This means that zero coupon bonds are traded between two market players instead of being traded in an exchange. This makes the access to zero coupon bond prices very difficult. To overcome this difficulty we used a theoretical price, supplied by one of the reference quote vendors, that stands as a reference price for the market.

At the time we got the data, the two year maturity zero coupon was live for approximately one year. Table 2 shows the parameters learned and figure 3 illustrates the corresponding mean and variance functions, along with the log prices sequence itself.

As it can be observed in figure 3, the mean and variance functions adjust quite well to the particular price sequence used.

6. Conclusions

In this paper we presented a calibration procedure of the Vasicek interest rate model under the risk neutral measure by learning the model parameters using Gaussian processes for machine learning regression with zero coupon bond log prices mean and covariance functions computed analytically.

Compared with other calibration procedures, in this one all the parameters are obtained in the arbitrage free risk neutral measure and the only prices needed for calibration are zero coupon bond prices. On the other hand, this
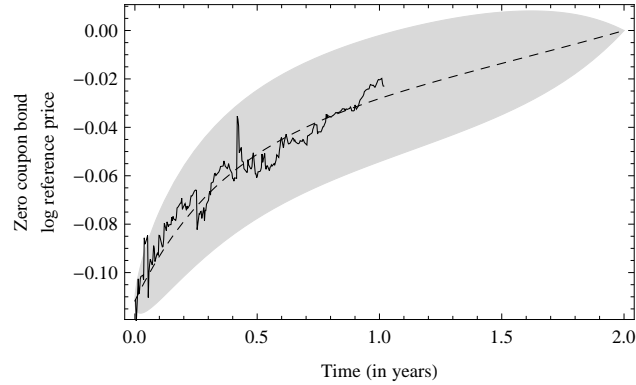
Figure 3: Real, two year maturity, zero coupon bond log reference prices sequence (solid black), learned mean (dashed black) and learned two standard deviations interval (light gray).

calibration procedure makes no discrete model approximation, makes no approximations that possibly allow arbitrage opportunities and does not require the establishment of an objective measure dynamics for the interest rate.

References

Björk, T. (2004). *Arbitrage Theory in Continuous Time*. (2nd ed.). Oxford University Press.

Brigo, D., & Mercurio, F. (2006). *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. (2nd ed.). Springer.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Revised Lectures* (pp. 63–71). Springer-Verlag volume 3176 of *Lecture Notes in Computer Science*.

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*, 177 – 188.

Wolfram Research, I. (2009). Mathematica edition: Version 7.01.0.