Machine learning Vasicek model calibration with Gaussian processes

J. Beleza Sousa*a,b, M. L. Esquivelb, R. M. Gasparc

aM2A/DEETC, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal
bCMA/DM, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal
cAdvance Research Center, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Rua Miguel Lups 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 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 under the risk neutral measure 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.
• The calibration method does not require a discrete model approximation nor the establishment of an objective measure dynamics.

*Corresponding author.
Email addresses: jsousa@deetc.isel.pt (J. Beleza Sousa), mle@fct.unl.pt (M. L. Esquivel), rmgaspar@iseg.ulis.pt (R. M. Gaspar)

Preprint submitted to Elsevier May 31, 2010

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

In the Vasicek model, the interest rate follows an Ornstein-Uhlenbeck mean-reverting process, under the risk neutral measure, defined by the stochastic differential equation

\[ dr(t) = k(\theta - r(t))dt + \sigma dW(t) \] (1)

where \( k \) is the mean reversion velocity, \( \theta \) is the mean interest rate level, \( \sigma \) is the volatility and \( W(t) \) the Wiener process, and \( k \) and \( \sigma \) are positive.

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

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

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

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

and variance

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

Zero coupon bonds are interest rate derivatives, therefore, their market prices are observed in the risk neutral measure. The Vasicek model has affine term structure, which means that the \( T \) maturity zero coupon bond prices \( p(t,T) \), observed in the risk neutral measure, are given by (Björk, 2004)

\[ p(t,T) = e^{A(t,T)-B(t,T)r(t)} \] (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.
2.1. Zero coupon bond log prices mean function

Since

\[ \log(p(t, T)) = A(t, T) - B(t, T)r(t) \]  \hspace{1cm} (5)

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

\[
\mu(t, T) = E \{ \log(p(t, T)) | \mathcal{F}_s \} \\
= E \{ A(t, T) - B(t, T)r(t) | \mathcal{F}_s \} \\
= A(t, T) - B(t, T)E \{ r(t) | \mathcal{F}_s \}
\]

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

\[
\mu(t, T) = A(t, T) - B(t, T) \left( r_0 e^{-kt} + \theta \left( 1 - e^{-kt} \right) \right) \\
= \frac{1}{4k^3} \sigma^2 k^4 \left( -4 \left( e^{k(T-t)} - 1 \right) k^2 (r_0 - \theta) \right. \\
\left. -4 e^{kt} k^3 (T - t) \theta + 2 (e^{kt} - e^{kT}) \sigma^2 \right. \\
\left. + k \left( e^{kt} + e^{kT} (2T - 2t - 1) \right) \sigma^2 \right) 
\]

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

\[
cov(t_1, t_2, T) = E \{ (\log(p(t_1, T)) - \mu(t_1, T)) \cdot (\log(p(t_2, T)) - \mu(t_2, T)) | \mathcal{F}_s \} \\
= E \{ \log(p(t_1, T)) \log(p(t_2, T)) | \mathcal{F}_s \} - \mu(t_1, T)\mu(t_2, T) \hspace{1cm} (7)
\]

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

\[
E \{ \log(p(t_1, T)) \log(p(t_2, T)) | \mathcal{F}_s \} \\
= E \{ \{ A(t_1, T) - B(t_1, T)r(t_1) \} \cdot \{ A(t_2, T) - B(t_2, T)r(t_2) \} | \mathcal{F}_s \} \\
= A(t_1, T)A(t_2, T) \\
- A(t_1, T)B(t_2, T)E \{ r(t_2) | \mathcal{F}_s \} \\
- B(t_1, T)A(t_2, T)E \{ r(t_1) | \mathcal{F}_s \} \\
+ B(t_1, T)B(t_2, T)E \{ r(t_1)r(t_2) | \mathcal{F}_s \} \\
= A(t_1, T)A(t_2, T) \\
- A(t_1, T)B(t_2, T)\mu(t_2, T) \\
- B(t_1, T)A(t_2, T)\mu(t_1, T) \\
+ B(t_1, T)B(t_2, T)E \{ r(t_1)r(t_2) | \mathcal{F}_s \} 
\]  \hspace{1cm} (8)
Using the Vasicek SDE solution equation 2, with \( s = 0 \), the term \( E \{ r(t_1)r(t_2) | F_s \} \) is given by

\[
E \{ r(t_1)r(t_2) | F_s \} = E \left\{ \left( r_0 e^{-kt_1} + \theta \left( 1 - e^{-kt_1} \right) + \sigma e^{-kt_1} \int_0^{t_1} e^{ku} dW(u) \right) \left( r_0 e^{-kt_2} + \theta \left( 1 - e^{-kt_2} \right) + \sigma e^{-kt_2} \int_0^{t_2} e^{ku} dW(u) \right) \right\} 
\]

\[
= r_0^2 e^{-k(t_1+t_2)} + r_0 e^{-kt_1} \theta \left( 1 - e^{-kt_2} \right) + \theta \left( 1 - e^{-kt_1} \right) r_0 e^{-kt_2} + \theta^2 \left( 1 - e^{-kt_1} \right) \left( 1 - e^{-kt_2} \right) + \sigma^2 e^{-k(t_1+t_2)} E \left\{ \int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u) \right\}. 
\]

(9)

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

\[
E \left\{ \int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u) \right\} = \left\{ \int_0^{t_1} e^{ku} dW(u) \right\} \left\{ \int_0^{t_2} e^{ku} dW(u) + \int_{t_1}^{t_2} e^{ku} dW(u) \right\} 
\]

\[
= E \left\{ \left( \int_0^{t_1} e^{ku} dW(u) \right)^2 \right\}. 
\]

(10)

In case \( t_2 < t_1 \), we have

\[
E \left\{ \int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u) \right\} = \left\{ \int_0^{t_2} e^{ku} dW(u) + \int_{t_2}^{t_1} e^{ku} dW(u) \right\} \left\{ \int_0^{t_2} e^{ku} dW(u) \right\} 
\]

\[
= E \left\{ \left( \int_0^{t_2} e^{ku} dW(u) \right)^2 \right\}. 
\]

(11)

Given equations 10 and 11, we get

\[
E \left\{ \int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u) \right\} = E \left\{ \left( \int_0^{\min(t_1,t_2)} e^{ku} dW(u) \right)^2 \right\}. 
\]
Finally, using Itô isometry

\[ E \left\{ \int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u) \right\} = E \left\{ \left( \int_0^{\min(t_1, t_2)} e^{ku} dW(u) \right)^2 \right\} \]

\[ = \int_0^{\min(t_1, t_2)} E \left\{ (e^{ku})^2 \right\} du = \int_0^{\min(t_1, t_2)} e^{2ku} du = \frac{1}{2k} \left( e^{2k \min(t_1, t_2)} - 1 \right). \]  

(12)

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

\[ \text{cov}(t_1, t_2, T) = \frac{1}{2k^3} e^{-k(2T+t_1+t_2)} \left( e^{2k \min(t_1, t_2)} - 1 \right) \frac{e^{kT} - e^{kt_1}}{e^{kT} - e^{kt_2}} \sigma^2 \]  

(13)

3. Gaussian processes for machine learning

The goal of Gaussian processes for machine learning is to find the non linear unknown mapping \( y = f(x) \), from data \((X, y)\), using Gaussian distributions over functions\(^1\) (Rasmussen & Williams, 2005)

\[ GP \sim N(\mu(x), \text{cov}(x_1, x_2)). \]

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

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

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

\[ GP \sim N(\mu(t, T), \text{cov}(t_1, t_2, T)) \]

where \( \mu(t, T) \) is given by equation 6 and \( \text{cov}(t_1, t_2, T) \) 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

\[ y = f(t) \]

\(^1\)See (Rasmussen, 2004) for a short introduction to the Gaussian distributions over functions framework.
where \( y \) stands for the zero coupon bonds log prices. This reduces the training set to the pair of vectors \((t, y)\), and the test set to vector \(t^*\).

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

\[
\begin{bmatrix} y \\ y^* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu \\ \mu^* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right)
\]

and

\[
p(y^*|t^*, t, y) \sim \mathcal{N} \left( \mu^* + K_T^T K^{-1} (y - \mu), K_{**} - K_T^T K^{-1} K_* \right)
\]

where \( \mu \) and \( \mu^* \) are mean vectors of train and test sets, \( K \) is the train set covariance matrix, \( K_* \) the train-test covariance matrix and \( K_{**} \) the test set covariance matrix.

The conditional distribution

\[
p(y^*|t^*, t, y)
\]

corresponds to the posterior process on the data

\[
\mathcal{GP}_D \sim \mathcal{N}(m_D(t), \text{cov}_D(t_1, t_2))
\]

where

\[
m_D(t) = m(t) + K_T^T K^{-1} (y - \mu)
\]

and

\[
\text{cov}_D(t_1, t_2) = \text{cov}(t_1, t_2) - K_T^T K^{-1} K_{t_1, t_2}
\]

where \( K_{t_1, t} \) 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)

\[
L = \log p(y|t, \Theta)
\]

\[
= -\frac{1}{2} \log |K| - \frac{1}{2} (y - \mu)^T K^{-1} (y - \mu) - \frac{n}{2} \log(2\pi)
\]

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 \( k \), \( \theta \) and \( \sigma \), to be learned from the zero coupon bond log prices.

Denoting each of the parameters in set \( \Theta \) by \( \Theta_i \), and since

\[
\frac{\partial}{\partial \Theta_i} \log |K| = tr \left( K^{-1} \frac{\partial K}{\partial \Theta_i} \right)
\]
and

$$\frac{\partial}{\partial \Theta_i} K^{-1} = -K^{-1} \frac{\partial K}{\partial \Theta_i} K^{-1}$$

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

$$\frac{\partial L}{\partial \Theta_i} = \frac{1}{2} tr \left( K^{-1} \frac{\partial K}{\partial \Theta_i} \right) + \frac{1}{2} (y - \mu)^T K^{-1} \frac{\partial K}{\partial \Theta_i} K^{-1} (y - \mu) + (y - \mu)^T K^{-1} \frac{\partial \mu}{\partial \Theta_i}.$$

In order to compute the vector of derivatives, $\frac{\partial \mu}{\partial \Theta_i}$, and the matrix of derivatives $\frac{\partial K}{\partial \Theta_i}$, 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:

$$\frac{\partial \mu(t, T)}{\partial \tau_0} = e^{-kt} - e^{-kT}.$$
\[
\frac{\partial \mu(t,T)}{\partial k} = \frac{1}{4k^4} e^{-kT} \left( 4k^3(e^{k(T-t)}(t-T)(r_0 - \theta) - 6(e^{kt} - e^{kT})\sigma^2 + 2k(e^{kT}(2t - 2T + 1) + e^{kt}(t - T - 1))\sigma^2 + k^2(-4r_0 + 4e^{k(T-t)}(r_0 - \theta) + 4\theta + e^{kt}\sigma^2 - e^{kT}\sigma^2) \right);
\]
\[
\frac{\partial \mu(t,T)}{\partial \theta} = \frac{e^{-kt} - e^{-kT} + kt - kT}{k};
\]
\[
\frac{\partial \mu(t,T)}{\partial \sigma} = \frac{(e^{k(T-t)}(k + 2) + 2kT - 2kt - k - 2)\sigma}{2k^3};
\]
\[
\frac{\partial \text{cov}(t_1,t_2,T)}{\partial r_0} = 0;
\]
\[
\frac{\partial \text{cov}(t_1,t_2,T)}{\partial k} = \frac{1}{2k^4} e^{-k(t_1+t_2+2T)} \left( e^{k(t_1+t_2)}(3 + 2kT) + e^{2kT}(3 + k(t_1 + t_2)) - e^{k(t_1+T)}(3 + k(t_2 + T)) - e^{k(t_2+T)}(3 + k(t_1 + T)) + e^{k(t_1+2\min(t_1,t_2)+T)}(3 + k(t_2 - 2\min(t_1,t_2) + T)) + e^{k(t_2+2\min(t_1,t_2)+T)}(3 + k(t_1 - 2\min(t_1,t_2) + T)) + e^{k(t_1+t_2+2\min(t_1,t_2))}(-3 - 2k(\min(t_1,t_1) - T)) + e^{2k(\min(t_1,t_2)+T)}(-3 - k(t_1 + t_2 - 2\min(t_1,t_2))) \right)\sigma^2;
\]
\[
\frac{\partial \text{cov}(t_1,t_2,T)}{\partial \theta} = 0;
\]
\[
\frac{\partial \text{cov}(t_1,t_2,T)}{\partial \sigma} = \frac{1}{k^2} e^{-k(t_1+t_2+2T)} \left( e^{kT} - e^{kt_1} \right) \left( e^{kT} - e^{kt_2} \right) \left( e^{2k\min(t_1,t_2)} - 1 \right) \sigma.
\]

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

Figure 2: Learned parameters 50 bins histograms.

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. Despite this fact, results in Table 1 also show that while the 1000 calibrations sample mean estimations, given by the parameters means showed in third column of Table 1, are quite accurate, the confidence intervals for each individual calibration are quite large. Therefore calibration with the proposed method should include as much zero coupon bond prices sequences as possible. Care should be taken when calibrating the model with just one price sequence, because the estimation errors can be quite large.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_0$</td>
<td>0.5</td>
<td>0.527</td>
<td>0.482</td>
<td>-0.418 to 1.472</td>
</tr>
<tr>
<td>$k$</td>
<td>2.0</td>
<td>2.098</td>
<td>0.855</td>
<td>0.419 to 3.776</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.1</td>
<td>0.083</td>
<td>0.443</td>
<td>-0.786 to 0.952</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2</td>
<td>0.203</td>
<td>0.039</td>
<td>0.126 to 0.280</td>
</tr>
</tbody>
</table>

Table 1: Parameters $r_0$, $k$, $\theta$ and $\sigma$, 1000 calibrations mean, standard deviation and 95% confidence interval.

<table>
<thead>
<tr>
<th>$r_0$</th>
<th>$k$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learned</td>
<td>0.212</td>
<td>2.925</td>
<td>0.025</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.110</td>
<td>1.897</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Table 2: Learned parameters $r_0$, $k$, $\theta$ and $\sigma$, for a real, two year maturity, zero coupon bond, calibrated with approximately one year of available prices.

5. Calibration to real data

In real life there are only a few zero coupon bonds and their great majority is traded over the counter (OTC). Those bonds are traded between two market players instead of openly being traded in an exchange and data is scarce. Zero coupon bond prices can also be implicitly extracted from the prices of other fixed income products, but in this case the prices would be theoretical and dependent on the assumptions underlying the term structure fitting.

We had, thus, two choices: either use actual market data on one of the few zero coupon bonds actually traded on the market and rely on OTC quotes, or use the price on other fixed income products and extract from those theoretical zero coupon bond prices. We chose the first approach. At the time we looked for data there was a two year maturity zero coupon in the OTC market live for approximately one year. We took this bond and used a quote service that delivers market prices aggregated from different dealers responsible for trading (market makers) this particular bond.

Table 2 shows the parameters learned and standard deviations obtained from the square root of the diagonal of the inverse Fisher information matrix, as a measure of the calibration errors. Figure 3 illustrates the corresponding mean and variance functions, along with the log prices sequence itself. It should be noted that the learned parameters along with equations 14 and 15 allow the plot in Figure 3 of the learned mean and the learned two standard deviations interval, not only in the time interval where there are available prices, but for all $t$ between the initial time $t = 0$ and maturity $t = T$.

As it can be observed in Figure 3, the mean and variance functions adjust quite well to the particular price sequence used. However, the large standard deviations obtained do confirm that the errors in the learned parameters can be quite large when calibrating with just one price sequence.
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 calibration procedure makes no discrete model approximation and makes no simplifications that possibly allow arbitrage opportunities, as it happens when calibrating using interest rate trees (Brigo & Mercurio, 2006). It also does not require the establishment of an objective measure dynamics for the interest rate, as it happens when applying the classical maximum likelihood estimation directly to the interest rate (Brigo & Mercurio, 2006).

References


