João Beleza Teixeira Seixas e Sousa

Mestre

Machine Learning Gaussian Short Rate

Dissertação para obtenção do Grau de Doutor em
Estatística e Gestão do Risco

Orientadores : Manuel Leote Tavares Inglês Esquível,
Professor Associado,
Universidade Nova de Lisboa
Raquel Maria Medeiros Gaspar,
Professora Associada,
Universidade de Lisboa

Júri:

Presidente: Professor Doutor Pedro Manuel Corrêa Calvente Barahona

Arguentes: Professor Doutor Paulo Eduardo Aragão Aleixo Neves de Oliveira
Professor Doutor João Pedro Vidal Nunes

Vogais: Professora Doutora Paula Manuela Lemos Pereira Milheiro de Oliveira
Professora Doutora Marta Cristina Vieira Faias Mateus
Professor Doutor Manuel Leote Tavares Inglês Esquível
Professora Doutora Raquel Maria Medeiros Gaspar

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2013

Machine Learning Gaussian Short Rate

To Edite, Miguel, Pedro and Afonso

Acknowledgements

First and above all I want to thank Prof. Esquível because this work would not have been possible with any other adviser. Prof. Esquível was sensible from the very first moment to my diverse and slightly chaotic background in signal processing, stochastic processes, computer science, computer graphics, both on academy and on industry, but also to my unconditional interest for financial markets, and took the enormous risk of advising an electrical engineer in mathematical finance. He gave me total freedom to develop my ideas but, at the same time, he was always there to break my deadlocks using the tough mathematical finance theory. I must say that the main results in this work were obtained working shoulder by shoulder with Prof. Esquível. I'm deeply grateful to him.

Secondly I want to thank Prof. Raquel Gaspar for her guidance and because this work would not yet been finished without her. I can't count the number of times she told me "you can continue writing papers after your PhD".

To Prof. Agra Coelho for working with me in an application of his research work that undoubtedly enlarges the scope of this thesis.

To Prof. Tiago Mexia the truly inspiration that working with him was.

Among my colleagues at ISEL I must thank Gonçalo Marques. There were many times when I asked my colleagues to ear me talking about some reasoning. Gonçalo Marques was by far the most selected listener and he helped me several times beyond my expectations in an area that is not his, frequently changing from listener to talker.

Among my discussion partners at ISEL I want also to thank Arnaldo Abrantes and Pedro Mendes Jorge.

Among my directors at ISEL I want to thank Arnaldo Abrantes and Manuel Barata for their unconditional support.

To João Hespanha for his surgical answers to my mathematics questions.

To Daniel Fryxell for his surgical answers to my financial markets questions.

To M2A/ISEL for financing my conference trips to Greensboro EUA, October 2007, Memphis EUA, May 2008 and Saint Petersburg Russia, June 2009.

To ISEL for financing my conference fees in EWGFM-XLI, Lisboa Portugal, SPE-XVII, Sesimbra Portugal, September 2009 and LinStat2010, Tomar Portugal, July 2010.

To CMA/FCT/UNL for financing my conference trips to Trieste Italy, June 2011, Sydney Australia, June 2012 and Greensboro USA, October 2012. Also for financing my conference fees in Optimization 2011, Caparica Portugal, July 2011 and yBIS-jSPE, Caparica Portugal, July 2012.

To ISEL/IPL the PROTEC scholarship that granted me 50% teaching hours reduction.

Finally I must thank the only teacher of English that could accept proofreading an English text she didn't understand at all. Thanks mum!

Abstract

The main theme of this thesis is the calibration of a short rate model under the risk neutral measure.

The problem of calibrating short rate models arises as most of the popular models have the drawback of not fitting prices observed in the market, in particular, those of the zero coupon bonds that define the current term structure of interest rates.

This thesis proposes a risk neutral Gaussian short rate model based on Gaussian processes for machine learning regression using the Vasicek short rate model as prior. The proposed model fits not only the prices that define the current term structure observed in the market but also all past prices. The calibration is done using market observed zero coupon bond prices, exclusively. No other sources of information are needed.

This thesis has two parts. The first part contains a set of self-contained finished papers, one already published, another accepted for publication and the others submitted for publication. The second part contains a set of self-contained unsubmitted papers. Although the fundamental work on papers in part two is finished as well, there are some extra work we want to include before submitting them for publication.

Part I:

- Machine learning Vasicek model calibration with Gaussian processes

  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. We stress that the only prices needed for calibration are market observed zero coupon bond prices and that the parameters are directly obtained in the arbitrage free risk neutral measure.

- One Factor Machine Learning Gaussian Short Rate

  In this paper we model the short rate, under the risk neutral measure, as a Gaussian process, conditioned on market observed zero coupon bonds log prices. The model is based on Gaussian processes for machine learning, using a single Vasicek factor as prior.

  All model parameters are learned directly under the risk neutral measure, using zero coupon bonds log prices only.

  The model supports observations of zero coupon bounds with distinct maturities limited to one observation per time instant. All the supported observations are automatically fitted.

- Brownian Bridge and other Path Dependent Gaussian Processes Vectorial Simulation

  The iterative simulation of the Brownian bridge is well known. In this paper we present a vectorial simulation alternative based on Gaussian processes for machine learning regression that is suitable for interpreted programming languages implementations.

  We extend the vectorial simulation of path dependent trajectories to other Gaussian processes, namely, sequences of Brownian bridges, geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process.

- Bonds Historical Simulation Value at Risk

  Bonds historical returns can not be used directly to compute Value at Risk (VaR) by historical simulation because the maturities of the yields implied by the historical prices are not the relevant maturities at time VaR is computed.

  In this paper we adjust bonds historical returns so that the adjusted returns can be used directly to compute VaR by historical simulation.

The adjustment is based on using implied historical yields to mark to model the bonds at the times to maturity relevant for the VaR computation.

We show that the obtained VaR values agree with the usual market trend of shorter times to maturity being traded with smaller yields, hence, carrying smaller risk and consequently having a smaller VaR.

Part II:

- Machine Learning Gaussian Short Rate

  In this paper we model the short rate, under the risk neutral measure, as a Gaussian process conditioned by the logarithm of market observed zero coupon bonds prices. The model is based on Gaussian processes for machine learning, using N addictive Vasicek factors as prior.

  The model automatically fits all observed zero coupon bond log prices, in particular those that define the current term structure of interest rates.

  The number of factors needed is equal to the maximum number of zero coupon bonds maturities observed in a single time instant.

  All model parameters are learned directly under the risk neutral measure, using zero coupon bonds log prices, exclusively.

- Interest Rate Market Changes Detection

  In this paper we check for interest rate market changes, using the distribution of the likelihood ratio criterion, to test if a covariance matrix, Σ , is equal to a given matrix, Σ_0.

  We start by transforming the original test into the equivalent test $\Sigma = I$. Then, the test $\Sigma = I$ is decomposed into two conditional independent tests, namely, the sphericity test $\Sigma = \sigma^2 I$, and the test $\sigma^2 = 1$, given that the data are spherical. The distribution moments and characteristic function are obtained. The characteristic function inversion is done numerically.

  We apply the covariance matrix test to check interest rate market changes using Euribor real data. We model the Euribor with a one factor machine learning Gaussian short rate model, using the Vasicek short rate model as prior, and assuming Vasicek short rate noise in the observations. In the beginning we calibrate the model to get a reference parameters set. Then, in the presence of newer data, we recalibrate the model and get a newer parameters set. We check the validity of the reference parameters set, using the statistical test applied to the model observations covariance matrix computed with both sets of parameters. Whenever the newer covariance matrix is not equal to the reference one, we say that the market conditions have changed.

Keywords: Short rate; Arbitrage free risk neutral measure; Gaussian processes for machine learning; Calibration; Zero coupon bond.

Resumo

O tema principal desta dissertação é a calibração de um modelo de taxa de juro infinitesimal *short rate*.

O problema da calibração de modelos de taxa de juro infinitesimal coloca-se, na medida em que a maioria dos modelos mais populares não se ajusta às curvas de taxas de juro observadas no mercado.

Nesta dissertação propõe-se um modelo Gaussiano de taxa de juro infinitesimal na mediada de risco neutral, baseado em processos Gaussianos para aprendizagem automática. O modelo proposto ajusta-se não só à curva actual de taxas de juro observada no mercado, como a todas as curvas observadas no passado. A calibração é efectuada recorrendo única e exclusivamente a preços de obrigações sem cupão, observados no mercado. Não são necessárias quaisquer outras fontes de informação.

Esta dissertação tem duas partes. A primeira parte contém um conjunto de artigos finalizados e auto-contidos, um deles já publicado, outro aceite para publicação e os restantes submetidos a publicação. A segunda parte contém um conjunto de artigos auto-contidos, não submetidos a publicação. Apesar do trabalho fundamental dos artigos da segunda parte estar também finalizado, pretende-se ainda incluir nesses artigos algum trabalho extra, antes de os submeter a publicação.

Parte I:

- Machine learning Vasicek model calibration with Gaussian processes

  Neste artigo calibra-se o modelo de taxa de juro de Vasicek, na medida de risco neutral, usando aprendizagem automática com processos Gaussianos para determinar os parâmetros do modelo. A calibração é efectuado por

maximização da verosimilhança do logaritmo de preços de obrigações sem cupões, usando funções média e covariância determinadas analiticamente, e usando também as derivadas da verosimilhança em ordem aos parâmetros. O método de maximização utilizado é o método dos gradientes conjugados. Os únicos preços necessários para efectuar a calibração são preços de obrigações sem cupões e os parâmetros são obtidos diretamente na medida de riso neutral.

- One Factor Machine Learning Gaussian Short Rate

  Neste artigo modela-se a taxa de juro infinitesimal, na medida de risco neutral, como um processo Gaussiano, condicionado ao logaritmo dos preços de obrigações sem cupões, observados no mercado. O modelo é baseado em aprendizagem automática com processos Gaussianos, usando como modelo à priori um único factor que segue o modelo de Vasicek.

  Todos os parâmetros do modelo são obtidos diretamente na medida de risco neutral, usando apenas o logaritmo dos preços de obrigações sem cupões.

  O modelo suporta obrigações sem cupões com diferentes maturidades, limitado a uma observação em cada instante de tempo. O modelo ajusta-se automaticamente a todas as observações suportadas.

- Brownian Bridge and other Path Dependent Gaussian Processes Vectorial Simulation

  A simulação iterativa da ponte Browniana é bem conhecida. Neste artigo apresenta-se uma alternativa vetorial, baseada em aprendizagem automática com processos Gaussianos, que é apropriada para implementações com linguagens de programação interpretadas.

  A simulação vectorial de trajectórias dependentes do caminho é extendida a outros processos Gaussianos, nomeadamente, sequências de pontes Brownianas, movimento Browniano geométrico, movimento Browniano fracionário e processo de reversão à média de Ornstein-Ulenbeck.

- Bonds Historical Simulation Value at Risk

  Os retornos históricos de obrigações não podem ser usados diretamente para calcular o Value at Risk (VaR) por simulação histórica, porque as maturidades das taxas implícitas nos preços históricos não são as maturidades relevantes à data do cálculo do VaR.

Neste artigo os retornos históricos são ajustados de forma a poderem ser usados directamente no cálculo do VaR, por simulação histórica.

O ajustamento é baseado na utilização das taxas implicítas nos preços históricos para calcular os preços das obrigações nos instantes correpondentes às maturidades relevantes para o cálculo do VaR.

Mostra-se que os valores de VaR obtidos estão de acordo com a tendência usual observada no mercado, caracterizada pelas maturidades mais curtas serem transacionadas com taxas menores, correspondendo a um menor risco e, consequentemente exibindo um VaR menor.

Parte II:

- Machine Learning Gaussian Short Rate

  Neste artigo modela-se a taxa de juro infinitesimal, na medida de risco neutral, como um processo Gaussiano, condicionado ao logaritmo dos preços de obrigações sem cupões, observados no mercado. O modelo é baseado em aprendizagem automática com processos Gaussianos, usando como modelo à priori a soma de N factores que seguem o modelo de Vasicek.

  O modelo ajusta-se automaticamente ao logaritmo de todos os preços de obrigações sem cupões observados, em particular àqueles que definem a estrutura de termo das taxas de juro atual.

  O número de fatores é igual ao número máximo de preços de obrigações sem cupões, observados num mesmo instante.

  Todos os parâmetros do modelo são obtidos diretamente na medida de risco neutral, usando apenas o logaritmo dos preços de obrigações sem cupões.

- Interest Rate Market Changes Detection

  Neste artigo detetam-se alterações no mercado de taxas de juro, usando a distribuição do critério da razão de verosimilhanças para testar se uma matriz de covariância, Σ, é igual a uma dada matriz, Σ_0.

  Começa-se por transformar o teste original no teste equivalente $\Sigma = I$. Em seguida o teste $\Sigma = I$ é decomposto em dois testes condicionalmente independentes, nomeadamente, o teste de esfericidade, $\Sigma = \sigma^2 I$, e o teste $\sigma^2 = 1$, assumindo dados esféricos. São obtidos os momentos e a função característica da distribuição. A inversão da função característica é efectuada numericamente.

O teste à matriz de covariância é aplicado na deteção de alterações do mercado de taxas de juro, usando dados reais da Euribor. O modelo usado para a Euribor é o modelo de taxa de juro infinitesimal de aprendizagem automática, com um factor Vasicek como modelo à priori, assumindo ruído Vasicek nas observações. Inicialmente, calibra-se o modelo e obtém-se um conjunto de parâmetros de referência. Em seguida, na presença de novos dados, recalibra-se o modelo e obtém-se um novo conjunto de parâmetros. A validade dos parâmetros de referência é aferida, usando o teste estatístico aplicado à matriz de covariância das observações, calculada com ambos os conjuntos de parâmetros. Sempre que a nova matriz de covariância não seja igual à de referência, dizemos que as condições do mercado mudaram.

Palavras-chave: Taxa de juro infinitesimal; Medida de risco neutral livre de arbitragem; Processos Gaussianos para aprendizagem automática; Calibração; Obrigação sem cupões.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Interest rate basics

The financial instrument used in this thesis is the zero coupon bond.

A T-maturity zero coupon bond is a contract that guarantees its holder the payment of one currency unit at time T. No intermediate payments prior to time T exist (no coupons). The price at time $t < T$ of the T-maturity zero coupon bond is denoted by $p(t, T)$.

Given three time instants $t < S < T$, consider the usual construction of contracting, at time t, a deterministic rate of return over the period S to T, using zero coupon bonds, namely (Björk 2004):

At time t Sell one S-maturity zero coupon bond by $p(t, S)$ currency units. With that amount buy $p(t, S)/p(t, T)$ T-maturity zero coupon bonds. The resulting net investment at time t equals zero.

At time S Pay one unit of currency to the S-maturity zero coupon bond holder.

At time T Receive $p(t, S)/p(t, T)$ currency units for holding the T-maturity zero coupon bonds.

The final result of this construction is that of contracting at time t, an investment of one currency unit at time S, that returns $p(t, S)/p(t, T)$ currency units at time T. In interest rate terms this contract locks at time t the continuously compounded interest rate R, over the future period S to T, which is the solution of:

$$1e^{R(T-S)} = \frac{p(t,S)}{p(t,T)}. \tag{1.1}$$

Such interest rate is called the continuously compounded forward rate and is given by

$$R(t,S,T) = -\frac{\log p(t,T) - \log p(t,S)}{T-S}. \tag{1.2}$$

Assuming that $p(t,T)$ is differentiable w.r.t. T, the instantaneous forward rate $f(t,T)$ is given by

$$f(t,T) = -\frac{\partial \log p(t,T)}{\partial T}. \tag{1.3}$$

The instantaneous short rate, or simply the short rate, is defined by

$$r(t) = f(t,t). \tag{1.4}$$

The importance of the short rate relies on the fact that on an arbitrage free market, the price, at time $t < T$, of any T-maturity contingent claim with payoff $\Phi(r(T))$, is given by

$$\pi(t,T) = E^Q\left[e^{-\int_t^T r(s)ds} \times \Phi(r(T))\right] \tag{1.5}$$

where the expectation is to be taken under the arbitrage free risk neutral measure Q.

In particular, the price of a T-maturity zero coupon bond, in which case $\Phi(r(T)) = 1$, is given by

$$p(t,T) = E^Q\left[e^{-\int_t^T r(s)ds}\right]. \tag{1.6}$$

1.2 Calibration problem

A large number of arbitrage free risk neutral models of the short rate exist, being the most popular ones (Björk 2004): the Vasicel model (Vasicek 1977); the Cox-Ingersoll-Ross model (Cox, Ingersoll Jr, and Ross 1985); the Dothan model (Dothan 1978); the Black-Derman-Toy model (Black, Derman, and Toy 1990); the Ho-Lee model (Ho and LEE 1986); and the Hull-White model (J. Hull and White 1990b).

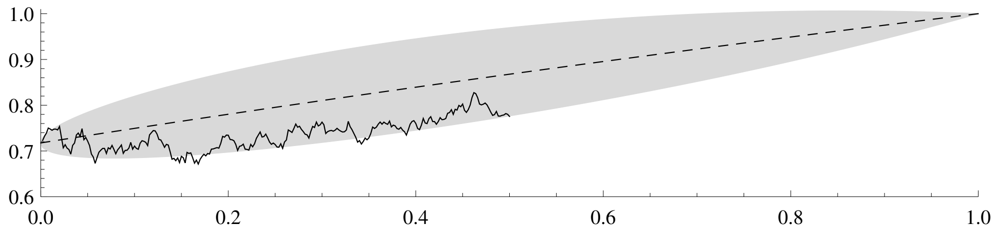

Figure 1.1: Vasicek zero coupon bond price mean (dashed) and two standard deviation band (light gray), along with a zero coupon bond prices sequence (solid) available until the current time.

The idea is to model the short rate dynamics directly under the risk neutral measure. This immediately raises the problem of obtaining the models parameters under the risk neutral measure while fitting the contingent claims prices observed in the market, in particular those of zero coupon bonds.

As noted by (Pang 1998) there are relatively few papers dedicated to this problem. Regarding the popular models mentioned above, the only one that explicitly deals with the calibration problem is the Hull-White model.

To illustrate this problem let's consider a zero coupon bond prices trajectory and the Vasicek short rate model under the risk neutral measure. Under this model zero coupon bond prices are given by

$$p(t,T) = e^{A(t,T)-B(t,T)r(t)} \tag{1.7}$$

where $A(t,T)$ and $B(t,T)$ are deterministic functions of the model's parameters. Zero coupon bond prices mean and covariance functions are given by closed forms.

Now suppose that the model parameters were obtained under the risk neutral measure from the zero coupon bond prices trajectory.

Figure 1.1 illustrates both the zero coupon bond prices trajectory and the model's zero coupon bond prices mean and covariance functions.

The calibration problem can be summarized by the following two issues:

1. As it can be observed in Figure 1.1, the current zero coupon bond model's price, given by the current model's mean, does not match the one observed in the market. This situation is quite uncomfortable.

2. Using Equation 1.7, one can solve for the current short rate value $r(t)$, using the current zero coupon bond price $p(t,T)$. Then, Equation 1.7 shows that the current price of all other zero coupon bonds with distinct maturities is given by a deterministic function of the obtained short rate value. This

means that the current term structure of interest rates obeys a fixed shape imposed by the model. This situation is highly unrealistic.

An additional description of this problem can be found in (Rainer 2009).

1.3 Gaussian processes for machine learning regression

Given a Gaussian process prior and a set of observations data, Gaussian processes for machine learning regression framework uses the conditional distribution of a Gaussian random vector (T. W. Anderson 2003) to construct the posterior process on data. The posterior process on data is the stochastic process whose trajectories are those of the prior process restricted to the ones that pass through the observed data. The posterior process is also Gaussian. Its mean and covariance functions are used as regression and regression confidence functions, respectively.

Consider the Gaussian process prior $y = g(\mathbf{x})$ defined by

$$g(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), cov(\mathbf{x}_i, \mathbf{x}_j)\right) \tag{1.8}$$

where the mean and covariance functions, $m(\mathbf{x})$ and $cov(\mathbf{x}_i, \mathbf{x}_j)$, are families of functions parametrized by the parameters vector Φ.

Consider also, data $\mathcal{D} = (\mathbf{X}, \mathbf{y})$, where matrix $\mathbf{X}$ collects a set of vectors $\{\mathbf{x}_1^\diamond, \mathbf{x}_2^\diamond, \cdots, \mathbf{x}_n^\diamond\}$ where the value $y^\diamond = g(\mathbf{x}^\diamond)$ was observed, and vector $\mathbf{y}$ collects the corresponding set of observed values $\{y_1^\diamond, y_2^\diamond, \cdots, y_n^\diamond\}$.

The posterior Gaussian process on data, $y = g_{\mathcal{D}}(\mathbf{x})$, is defined by

$$g_{\mathcal{D}}(\mathbf{x}) \sim \mathcal{GP}\left(m_{\mathcal{D}}(\mathbf{x}), cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)\right). \tag{1.9}$$

where mean and covariance function, $m_{\mathcal{D}}(\mathbf{x})$ and $cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)$, are given by (Rasmussen 2004)

$$m_{\mathcal{D}}(\mathbf{x}) = m(\mathbf{x}) + \mathbf{K}_{\mathbf{X},\mathbf{x}}^\top \mathbf{K}^{-1}(\mathbf{y} - \mathbf{m}) \tag{1.10}$$

and

$$cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j) = cov(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{X},\mathbf{x}_i}^\top \mathbf{K}^{-1} \mathbf{K}_{\mathbf{X},\mathbf{x}_j}. \tag{1.11}$$

Vector $\mathbf{m}$ is the prior training set mean vector, matrix $\mathbf{K}$ is the prior training set covariance matrix and vector $\mathbf{K}_{\mathbf{X},\mathbf{x}}$ is a prior covariance vector between every training vector and $\mathbf{x}$.

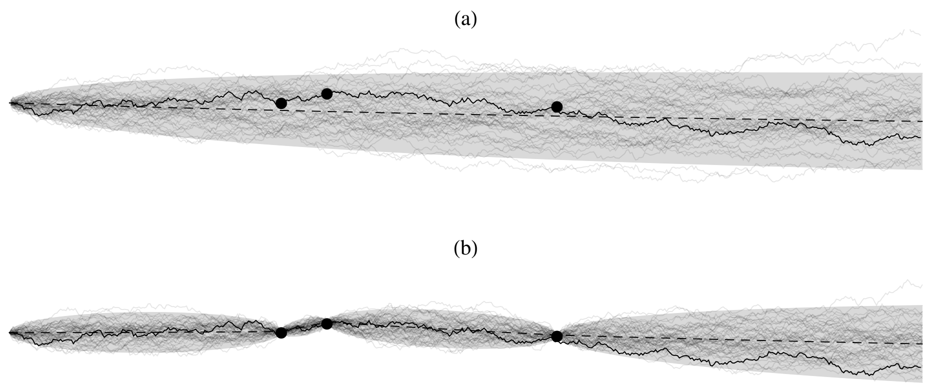

Figure 1.2: (a) A Gaussian process mean (dashed), two standard deviations band (gray) along with some simulated trajectories. (b) the same for the corresponding conditioned on data Gaussian. The conditioning data is a set of samples (dots) of the highlighted trajectory (solid).

Figure 1.2 illustrates both the prior process and the conditioned on data process.

Parameters Φ are obtained directly from data by maximizing the prior likelihood of the data given the parameters. Given that the process is Gaussian, closed forms of the derivatives of the likelihood w.r.t. each parameter are available and can be used by the maximization procedure.

A natural problem that arises under this framework is the selection of prior mean and covariance functions to use.

1.4 Thesis contribution

The main contribution of this thesis is the proposal of a risk neutral short rate model developed by merging arbitrage free interest rate theory with Gaussian processes for machine learning.

From arbitrage free interest rate theory we use the Vasicek model under the risk neutral measure as prior. Under this model, zero coupon bond log prices are Gaussian. Their mean and covariance functions are given by closed forms.

From Gaussian processes for machine learning we use the conditioned on data regression model. Under this framework the model parameters are learned directly from data and the resulting model automatically fits all the data.

Putting the two pieces together:

- we addressed the problem of obtaining the risk neutral short rate model

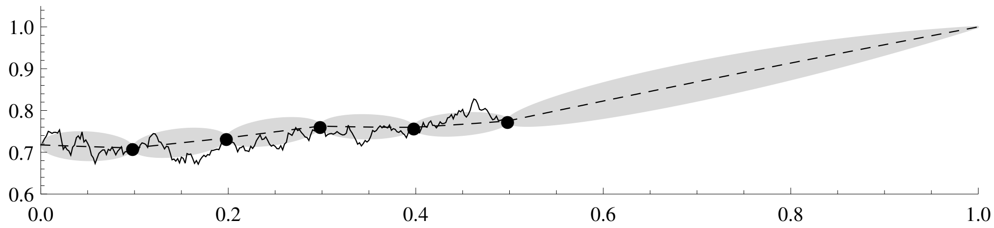

Figure 1.3: One factor machine learning Gaussian short rate model for maturity $T = 1$: zero coupon bond prices mean (dashed), two standard deviations band (gray), a zero coupon bond prices sequence (solid), and the conditioning data (dots).

parameters under the risk neutral measure by using the Gaussian processes for machine learning regression, learning parameters from data procedure, with market observed zero coupon log prices;

- we addressed the risk neutral short rate model problem of fitting zero coupon bond prices observed in the market problem by using the Gaussian processes for machine learning regression conditioned on data model;
- we addressed Gaussian processes for machine learning problem of choosing priors mean and covariance functions by using a log normal risk neutral short rate model with mean and covariance functions given by closed forms.

The proposed model is a risk neutral measure Vasicek short rate prior conditioned on zero coupon bond prices.

All the parameters of the proposed model are obtained directly under the risk neutral measure using market observed zero coupon bond log prices, exclusively. No other data sources are needed.

The model automatically fits, by its construction, all zero coupon bond prices observed in the market, in particular those that define the current term structure of interest rates.

Figure 1.3 illustrates the proposed model. It is the adaptation of Figure 1.2(b) to the setup in Figure 1.1 (to highlight the conditioning on data mechanism, only a subset of the available zero coupon bond prices is used).

As it can be observed in Figure 1.3, the proposed model solves the first calibration issue. The current zero coupon bond model's price, given by the current model's mean, exactly matches the one observed in the market. Regarding the second calibration issue, the addition of several short rate factors, while maintaining the conditioning on data procedure, solves it as well.

1.5 Thesis structure

This thesis has two parts. The first part contains a set of self-contained finished papers, one already published, another accepted for publication and the others submitted for publication. The second part contains a set of self-contained unsubmitted papers. Although the fundamental work on papers in part two is finished as well, there are some extra work we want to include before submitting them for publication.

All the work is supported by a large set of Wolfram Mathematica (Wolfram Research 2009) (Wolfram Research 2011) (Wolfram Research 2012) packages and notebooks. Facing the impossibility to present all those program in this thesis, we have chosen to include in Appendix A the public sections of two core packages, one symbolic and another numeric.

Part I

Machine learning Vasicek model calibration with Gaussian processes

This paper was the first step towards merging arbitrage free short rate theory with Gaussian processes for machine learning regression.

Using the zero coupon bond log prices of a single T-maturity bond, we have obtained the values of the Vasicek short rate model parameters directly under the risk neutral measure.

The main contributions of this paper are:

- Obtain the Vasicek zero coupon bond log prices mean and covariance functions closed forms;
- Show, by simulation, that the risk neutral model parameters are properly obtained from the zero coupon bond log prices, by maximizing the likelihood of the log prices given the parameters;
- Calibrate the model for a real zero coupon bond.

One Factor Machine Learning Gaussian Short Rate

In this paper we have recognized the conditioned on zero coupon bonds log prices short rate model, with the Vasicek short rate model as prior, as an alternative short rate model by itself.

The main contributions of this paper are:

- Obtain the model's deterministic time dependent stochastic differential equation parameters;
- Show, by simulation, that the risk neutral model parameters are properly obtained from several zero coupon bond log prices with distinct maturities, by maximizing the likelihood of the log prices given the parameters, as long as there is only one price observation in each time instant.

Brownian Bridge and other Path Dependent Gaussian Processes Vectorial Simulation

Both the iterative and the vectorial procedures for simulating the Wiener process are widely known, and described in reference books such as Glasserman 2003. However, regarding the Brownian bridge, only the iterative procedure is described.

In this paper we model the Brownian bridge using the Gaussian processes for machine learning regression framework, using the Wiener process, $W(t)$, as prior, and the single observation, $W(1) = 0$, the Brownian bridge condition, in the training set.

The main contributions of this paper are:

- Use the bridge mean vector and the covariance matrix, computed in a set of sampling instants, to simulate the bridge trajectories with the same vectorial procedure used to simulate any Gaussian vector;
- Extend the vectorial simulation procedure to other Gaussian processes priors, and for more than one conditions, by developing a general path dependent Gaussian process trajectories vectorial simulation framework;
- Show that the vectorial simulation procedure is relevant concerning the execution times of implementations with the interpreted programming languages widely used in today's research and development.

Bonds Historical Simulation Value at Risk

In several simulation situations spread across this thesis, in order to evaluate the existence of numerical problems, we have scaled zero coupon bond prices by using the implied yield at a certain time, to compute the bond price at another time, assuming the bond was held to maturity (mark to model).

This scaling procedure proved to be an important tool in the context of historical simulation value at risk (VaR) for portfolios with bonds.

In a joint work with Prof. Manuel Esquível and Prof. Pedro Corte Real, we have sold the authors wrights of an historical simulation value at risk implementation, for portfolios with bonds (among other securities), to a private bank, by 50.100,00 EUR. That implementation was based on this paper.

The main contributions of this paper are:

- Adjust bonds historical returns so that the adjusted returns can be used directly to compute VaR by historical simulation;
- Using real bond prices, to show that the developed method provides results consistent with the usual market observed trend, in which shorter times to maturity imply smaller yields, carrying smaller risk and consequently having smaller VaR;
- Using real bond prices, to show that the developed method strongly preserves the market implicit correlations between the instruments in the portfolio.

Part II

Machine Learning Gaussian Short Rate

Using a single Vasicek short rate factor, under the risk neutral measure, the machine learning Gaussian short rate model can't solve the term structure fitting issue mentioned in Section 1.2. In this paper a sum of Vasicek short rate factors is proposed in order to solve that problem.

The main contributions of this paper are:

- Propose a sum of Vasicek short rate factors, under the risk neutral measure, as a prior to Gaussian processes for machine learning regression;
- Obtain the zero coupon bond mean and covariance functions of the prior.

Interest Rate Market Changes Detection

A common problem that arises in mathematical finance when using models with parameters estimated from market data, available until a certain time, is that of checking the necessity of using new parameters as newer data become available.

In this paper we use a covariance matrix statistical test to evaluate the necessity of using new parameters of a machine learning Gaussian short rate model of the Euribor, as newer data become available. Whenever we detect such necessity, we say that the market conditions have changed.

The main contributions of this paper are:

- Obtain the likelihood ratio criterion to test if a covariance matrix, Σ , is equal to a given matrix, Σ_0, as a decomposition of simpler tests.
- Propose a machine learning Gaussian short rate model with Vasicek short rate noise in the observations.
- Using real data, model the Euribor with the proposed model and apply the changes detection procedure to the credit crisis years of 2007 and 2008.

Part I

Published and Submitted for Publication Papers

2

Machine learning Vasicek model calibration with Gaussian processes

2.1 Preamble

With the exception of this preamble and minor notation changes, this chapter contains the paper *Machine learning Vasicek model calibration with Gaussian processes*, joint work with Prof. Manuel Esquível and Prof. Raquel Gaspar, published in journal Communications in Statistics-Simulation and Computation, volume 41, number 6, pages 776 to 786, year 2012, by Taylor & Francis.

This paper was the first step towards merging arbitrage free short rate theory with Gaussian processes for machine learning regression.

Using the zero coupon bond log prices of a single T-maturity bond, we have obtained the values of the Vasicek short rate model parameters directly under the risk neutral measure.

The main contributions of this paper are:

- Obtain the Vasicek zero coupon bond log prices mean and covariance functions closed forms;
- Show, by simulation, that the risk neutral model parameters are properly obtained from the zero coupon bond log prices, by maximizing the likelihood of the log prices given the parameters;

- Calibrate the model for a real zero coupon bond.

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. We stress that the only prices needed for calibration are market observed zero coupon bond prices and that the parameters are directly obtained in the arbitrage free risk neutral measure.

Keywords: Vasicek interest rate model; Arbitrage free risk neutral measure; Calibration; Gaussian processes for machine learning; Zero coupon bond prices.

2.2 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.

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.

2.3 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

$$dr(t) = k(\theta - r(t))dt + \sigma dW(t) \tag{2.1}$$

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

Let $s \leq t$. The solution of equation 2.1 is (Brigo and Mercurio 2006)

$$r(t) = r(s)e^{-k(t-s)} + \theta\left(1 - e^{-k(t-s)}\right) + \sigma e^{-kt}\int_s^t e^{ku}dW(u). \tag{2.2}$$

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

$$E\left\{r(t)|\mathcal{F}_s\right\} = r(s)e^{-k(t-s)} + \theta\left(1 - e^{-k(t-s)}\right) \tag{2.3}$$

and variance

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

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)

$$p(t, T) = e^{A(t,T) - B(t,T)r(t)} \tag{2.4}$$

where

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

and

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

Equation 2.4 shows that the zero coupon bond prices $p(t,T)$ are log normal and consequently $\log(p(t,T))$ are normal.

2.3.1 Zero coupon bond log prices mean function

Since

$$\log(p(t,T)) = A(t,T) - B(t,T)r(t) \tag{2.5}$$

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 2.3 for $E\left\{r(t)|\mathcal{F}_s\right\}$ we get

$$\begin{aligned}\mu(t,T) &= A(t,T) - B(t,T)\left(r_0 e^{-kt} + \theta\left(1 - e^{-kt}\right)\right) \\ &= \left(\theta - \frac{\sigma^2}{2k^2}\right)\left(t - T - \frac{e^{k(t-T)} - 1}{k}\right) - \frac{\sigma^2\left(e^{k(t-T)} - 1\right)^2}{4k^3} \\ &\quad - \frac{e^{-kT}\left(e^{k(T-t)} - 1\right)\left(\theta\left(e^{kt} - 1\right) + r_0\right)}{k}\end{aligned} \tag{2.6}$$

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

2.3.2 Zero coupon bond log prices covariance function

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

$$\begin{aligned}cov(t_1,t_2,T) &= E\{(\log(p(t_1,T)) - \mu(t_1,T)) \\ &\quad (\log(p(t_2,T)) - \mu(t_2,T))\,|\mathcal{F}_s\} \\ &= E\left\{\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\right\} - \mu(t_1,T)\mu(t_2,T)\end{aligned} \tag{2.7}$$

Using equation 2.5, the term $E\left\{\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\right\}$, is given by

$$\begin{aligned}
&E\left\{\log(p(t_1,T))\log(p(t_2,T))|\mathcal{F}_s\right\}\\
&= E\left\{(A(t_1,T)-B(t_1,T)r(t_1))\right.\\
&\quad \left.(A(t_2,T)-B(t_2,T)r(t_2))|\mathcal{F}_s\right\}\\
&= A(t_1,T)A(t_2,T)\\
&\quad -A(t_1,T)B(t_2,T)E\left\{r(t_2)|\mathcal{F}_s\right\}\\
&\quad -B(t_1,T)A(t_2,T)E\left\{r(t_1)|\mathcal{F}_s\right\}\\
&\quad +B(t_1,T)B(t_2,T)E\left\{r(t_1)r(t_2)|\mathcal{F}_s\right\}
\end{aligned} \tag{2.8}$$

Using the Vasicek SDE solution equation 2.2, with $s=0$, the term $E\left\{r(t_1)r(t_2)|\mathcal{F}_s\right\}$ is given by

$$\begin{aligned}
&E\left\{r(t_1)r(t_2)|\mathcal{F}_s\right\}\\
&= E\left\{\left(r_0e^{-kt_1}+\theta\left(1-e^{-kt_1}\right)+\sigma e^{-kt_1}\int_0^{t_1}e^{ku}dW(u)\right)\right.\\
&\quad \left.\left(r_0e^{-kt_2}+\theta\left(1-e^{-kt_2}\right)+\sigma e^{-kt_2}\int_0^{t_2}e^{ku}dW(u)\right)\right\}\\
&= r_0^2e^{-k(t1+t2)}+r_0e^{-kt_1}\theta\left(1-e^{-kt_2}\right)\\
&\quad +\theta\left(1-e^{-kt_1}\right)r_0e^{-kt_2}+\theta^2\left(1-e^{-kt_1}\right)\left(1-e^{-kt_2}\right)\\
&\quad +\sigma^2e^{-k(t1+t2)}E\left\{\int_0^{t_1}e^{ku}dW(u)\int_0^{t_2}e^{ku}dW(u)\right\}.
\end{aligned} \tag{2.9}$$

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

$$\begin{aligned}
&E\left\{\int_0^{t_1}e^{ku}dW(u)\int_0^{t_2}e^{ku}dW(u)\right\}\\
&= E\left\{\left(\int_0^{t_1}e^{ku}dW(u)\right)\left(\int_0^{t_1}e^{ku}dW(u)+\int_{t_1}^{t_2}e^{ku}dW(u)\right)\right\}\\
&= E\left\{\left(\int_0^{t_1}e^{ku}dW(u)\right)^2\right\}.
\end{aligned} \tag{2.10}$$

In case $t_2 < t_1$, we have

$$
\begin{aligned}
& E\left\{\int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u)\right\} \\
&= E\left\{\left(\int_0^{t_2} e^{ku} dW(u) + \int_{t_2}^{t_1} e^{ku} dW(u)\right)\left(\int_0^{t_2} e^{ku} dW(u)\right)\right\} \\
&= E\left\{\left(\int_0^{t_2} e^{ku} dW(u)\right)^2\right\}. \qquad (2.11)
\end{aligned}
$$

Given equations 2.10 and 2.11, we get

$$
\begin{aligned}
& E\left\{\int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u)\right\} \\
&= E\left\{\left(\int_0^{min(t_1,t_2)} e^{ku} dW(u)\right)^2\right\}.
\end{aligned}
$$

Finally, using Itô isometry

$$
\begin{aligned}
& E\left\{\int_0^{t_1} e^{ku} dW(u) \int_0^{t_2} e^{ku} dW(u)\right\} \\
&= E\left\{\left(\int_0^{min(t_1,t_2)} e^{ku} dW(u)\right)^2\right\} \\
&= \int_0^{min(t_1,t_2)} E\left\{\left(e^{ku}\right)^2\right\} du \\
&= \int_0^{min(t_1,t_2)} e^{2ku} du \\
&= \frac{1}{2k}\left(e^{2k\, min(t_1,t_2)} - 1\right). \qquad (2.12)
\end{aligned}
$$

Using equations 2.6, 2.7, 2.8, 2.9 and 2.12, the covariance function $cov(t_1, t_2, T)$ of $\log(p(t, T))$ is given by

$$
\begin{aligned}
cov(t_1, t_2, T) &= \frac{1}{2k^3} e^{-k(2T+t1+t2)}\left(e^{2k\, min(t1,t2)} - 1\right) \\
& \left(e^{kT} - e^{kt1}\right)\left(e^{kT} - e^{kt2}\right)\sigma^2 \qquad (2.13)
\end{aligned}
$$

2.4 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[1] (Rasmussen and 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 2.6 and $cov(t_1, t_2, T)$ is given by equation 2.13.

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 $\mathbf{t}^\star$.

Since the process is Gaussian (Rasmussen and 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

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

[1] See (Rasmussen 2004) for a short introduction to the Gaussian distributions over functions framework.

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

$$m_{\mathcal{D}}(t) = m(t) + \mathbf{K}_{\mathbf{t},t}^T \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \tag{2.14}$$

and

$$cov_{\mathcal{D}}(t_1, t_2) = cov(t_1, t_2) - \mathbf{K}_{\mathbf{t},t_1}^T \mathbf{K}^{-1} \mathbf{K}_{\mathbf{t},t_2} \tag{2.15}$$

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

Equation 2.14 is the regression function while equation 2.15 is the regression confidence. Equations 2.14 and 2.15 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.

Denoting each of the parameters in set Θ by Θ_i, and since

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

and

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

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

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

In order to compute the vector of derivatives, $\frac{\partial \boldsymbol{\mu}}{\partial \Theta_i}$, and the matrix of derivatives $\frac{\partial \mathbf{K}}{\partial \Theta_i}$, the derivatives of the mean function $\mu(t,T)$ (equation 2.6), and the derivatives of the covariance function $cov(t_1, t_2, T)$ (equation 2.13) with respect to the parameters are used, namely:

$$\begin{aligned}\frac{\partial \mu(t,T)}{\partial r_0} &= \frac{e^{-kT} - e^{-kt}}{k}; \\ \frac{\partial \mu(t,T)}{\partial k} &= \frac{e^{-k(t+2T)}}{4k^4} \left(4k^2(kt+1)e^{2kT}(r_0 - \theta) \right. \\ &\quad -4k^2(kT+1)(r_0-\theta)e^{k(t+T)} + 4\sigma^2 e^{k(2t+T)}(k(t-T)-3) \\ &\quad \left. -\sigma^2 e^{3kt}(2k(t-T)-3) + \sigma^2 e^{k(t+2T)}(4k(t-T)+9) \right); \\ \frac{\partial \mu(t,T)}{\partial \theta} &= \frac{e^{-kt} - e^{-kT} + kt - kT}{k}; \\ \frac{\partial \mu(t,T)}{\partial \sigma} &= -\frac{\sigma \left(2k(t-T) - 4e^{k(t-T)} + e^{2k(t-T)} + 3 \right)}{2k^3}; \\ \frac{\partial cov(t_1, t_2, T)}{\partial r_0} &= 0;\end{aligned}$$

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

2.5 Simulation results

In order to test the proposed calibration method we used equations 2.2 and 2.5, 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 2.1 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 2.2 illustrates the 50 bins parameters histograms obtained from the 1000 calibrations performed, and Table 2.1 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 2.1 also show that while the 1000 calibrations sample mean estimations, given by the parameters

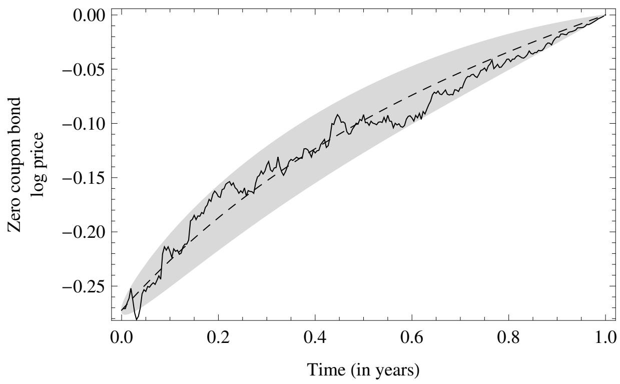

Figure 2.1: Zero coupon bond log prices simulated sequence (solid black), mean (dashed black) and two standard deviations interval (light gray).

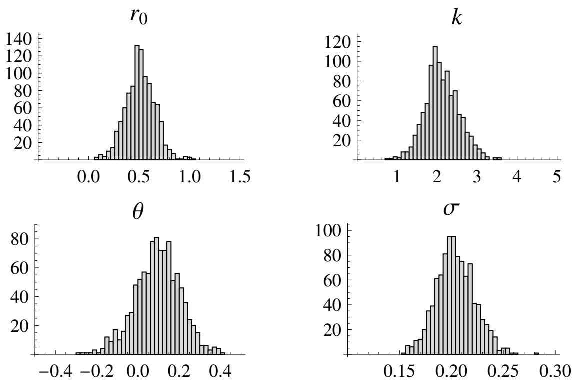

Figure 2.2: Learned parameters 50 bins histograms.

Parameter	Value	Mean	Std. Dev.	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 2.1: Parameters r_0, k, θ and σ, 1000 calibrations mean, standard deviation and 95% confidence interval.

	r_0	k	θ	σ
Learned	0.212	2.925	0.025	0.195
Std. Dev.	0.110	1.897	0.022	0.119

Table 2.2: Learned parameters r_0, k, θ and σ, for a real, two year maturity, zero coupon bond, calibrated with approximately one year of available prices.

means showed in third column of Table 2.1, 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.

2.6 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 2.2 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 2.3 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 2.14 and 2.15 allow the plot in Figure 2.3 of the learned mean and the learned two standard deviations interval, not only in the time interval where there are available prices, but for the all t between the initial time $t = 0$ and maturity $t = T$.

As it can be observed in Figure 2.3, the mean and variance functions adjust quite well to the particular price sequence used. However, the large standard

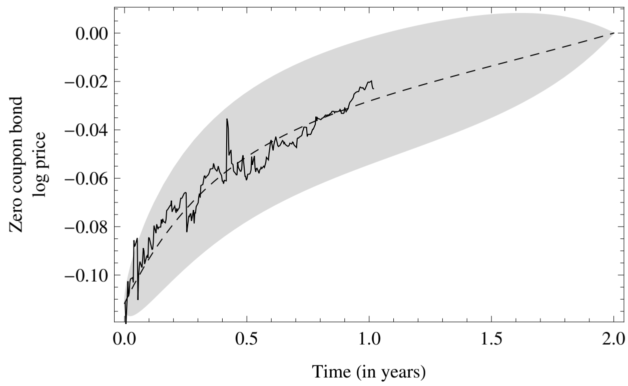

Figure 2.3: Real, two year maturity, zero coupon bond log prices sequence (solid black), learned mean (dashed black) and learned two standard deviations interval (light gray).

deviations obtained do confirm that the errors in the learned parameters can be quite large when calibrating with just one price sequence.

2.7 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 and 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 and Mercurio 2006).

3

One Factor Machine Learning Gaussian Short Rate

3.1 Preamble

With the exception of this preamble and minor notation changes, this chapter contains the paper *One Factor Machine Learning Gaussian Short Rate*, joint work with Prof. Manuel Esquível and Prof. Raquel Gaspar, which is submitted for publication in the journal Communications in Statistics-Simulation and Computation, by Taylor & Francis. The current submission status is "under revision".

In this paper we have recognized the conditioned on zero coupon bonds log prices short rate model, with the Vasicek short rate model as prior, as an alternative short rate model by itself.

The main contributions of this paper are:

- Obtain the model's deterministic time dependent stochastic differential equation parameters;
- Show, by simulation, that the risk neutral model parameters are properly obtained from several zero coupon bond log prices with distinct maturities, by maximizing the likelihood of the log prices given the parameters, as long as there is only one price observation in each time instant.

Abstract

In this paper we model the short rate, under the risk neutral measure, as a Gaussian process, conditioned on market observed zero coupon bonds log prices. The model is based on Gaussian processes for machine learning, using a single Vasicek factor as prior.

All model parameters are learned directly under the risk neutral measure, using zero coupon bonds log prices only.

The model supports observations of zero coupon bounds with distinct maturities limited to one observation per time instant. All the supported observations are automatically fitted.

We derive the model's SDE and model the Euribor using real data.

Keywords: Gaussian short rate, Gaussian processes for machine learning, risk neutral measure.

3.2 Introduction

In our previous work (Sousa, Esquível, and Gaspar 2012) we calibrated the Vasicek (Vasicek 1977) short rate model, for a single T maturity zero coupon bond, directly under the risk neutral measure, using zero coupon bond prices only. The method is based on conditioning the Vasicek zero coupon bond log prices Gaussian process, to market observed zero coupon bond log prices, using the Gaussian processes for Machine Learning framework (Rasmussen and Williams 2005). In this paper we recognize the conditioned on market observed zero coupon log prices underlying Gaussian process as an alternative one factor Gaussian short rate model by itself. We extend the calibration method to several T maturity zero coupon bonds, limited to one observation per time instant, and derive the model's stochastic differential equation.

Consider the Gaussian process $g(\mathbf{x})$, with mean function $m(\mathbf{x})$ and covariance function $cov(\mathbf{x}_i, \mathbf{x}_j)$,

$$g(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), cov(\mathbf{x}_i, \mathbf{x}_j)\right). \tag{3.1}$$

Consider also, data $\mathcal{D} = (\mathbf{X}, \mathbf{y})$, where the matrix $\mathbf{X}$ collects a set of vectors $\{\mathbf{x}_1^\diamond, \mathbf{x}_2^\diamond, \cdots, \mathbf{x}_n^\diamond\}$ where the value $y^\diamond = g(\mathbf{x}^\diamond)$ was observed, and vector $\mathbf{y}$ collects the corresponding set of observed values $\{y_1^\diamond, y_2^\diamond, \cdots, y_n^\diamond\}$.

The process, $g_{\mathcal{D}}(\mathbf{x})$, defined by all trajectories of $g(\mathbf{x})$ that pass through data $\mathcal{D}$ is also Gaussian (Rasmussen and Williams 2005), with mean function $m_{\mathcal{D}}(\mathbf{x})$ and covariance function $cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)$,

$$g_{\mathcal{D}}(\mathbf{x}) \sim \mathcal{GP}\left(m_{\mathcal{D}}(\mathbf{x}), cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)\right). \tag{3.2}$$

The process $g(\mathbf{x})$ is called the prior, $g_{\mathcal{D}}(\mathbf{x})$ is called the conditioned on data process, $\mathcal{D}$ is called the training set and $m_{\mathcal{D}}(\mathbf{x})$, $cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)$ are given by (Rasmussen 2004)

$$m_{\mathcal{D}}(\mathbf{x}) = m(\mathbf{x}) + \mathbf{K}_{\mathbf{X},\mathbf{x}}^{\top}\mathbf{K}^{-1}(\mathbf{y} - \mathbf{m}) \tag{3.3}$$

and

$$cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j) = cov(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{X},\mathbf{x}_i}^{\top}\mathbf{K}^{-1}\mathbf{K}_{\mathbf{X},\mathbf{x}_j} \tag{3.4}$$

where, $\mathbf{m}$ is the prior training set mean vector, $\mathbf{K}$ is the prior training set covariance matrix and $\mathbf{K}_{\mathbf{X},\mathbf{x}}$ is a prior covariance vector between every training vector and $\mathbf{x}$.

The "One Factor Machine Learning Gaussian Short Rate" is a single factor arbitrage free Vasicek short rate prior, conditioned on market observed zero coupon bonds log prices.

3.3 Short rate prior

The model's prior is that, under the arbitrage free, risk neutral measure, the short rate, $r(t)$, follows a Vasicek Ornstein-Uhlenbeck mean-reverting process, defined by the stochastic differential equation (SDE) (Vasicek 1977):

$$\mathrm{d}r(t) = k(\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{3.5}$$

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

Let s be the initial time, with $0 < s < t$. The solution of Equation 3.5 is given by (Brigo and Mercurio 2006)

$$r(t) = r(s)e^{-k(t-s)} + \theta(1 - e^{-k(t-s)}) + \sigma e^{-kt}\int_s^t e^{ku} dW(u). \tag{3.6}$$

The initial short rate value, $r(s)$, is considered as an extra model parameter since its value must be obtained under the risk neutral measure.

3.3.1 Short rate mean and covariance

Given Equation 3.6 the short rate prior mean and covariance functions, $m_r(t)$ and $cov_r(t_1, t_2)$, are given by

$$m_r(t) = r(s)e^{-k(t-s)} + \theta(1 - e^{-k(t-s)}) \tag{3.7}$$

and

$$cov_r(t_1, t_2) = \sigma^2 e^{-k(t_1+t_2)} \frac{1}{2k} \left(e^{2k\min(t_1,t_2)} - e^{2ks}\right). \tag{3.8}$$

3.3.2 Zero coupon bond log prices mean and covariance

Under the risk neutral measure, the price p, at time t, of a zero coupon bond that pays 1 at maturity T, is given by (Björk 2004)

$$p(t, T) = E\left[e^{-\int_t^T r(u)du}\right]. \tag{3.9}$$

Under the Vasicek model $p(t, T)$ is given by

$$p(t, T) = e^{A(t,T)-B(t,T)r(t)} \tag{3.10}$$

where

$$B(t, T) = \frac{1}{k}\left(1 - e^{-k(T-t)}\right) \tag{3.11}$$

and

$$A(t, T) \quad = \quad \left(\theta - \frac{\sigma^2}{2k^2}\right)(B(t, T) - T + t) - \frac{\sigma^2}{4k} B^2(t, T). \tag{3.12}$$

It is clear from Equation 3.10 that model has an affine term structure and that the logarithm of the zero coupon bonds prices is Gaussian

$$\log p(t, T) \sim \mathcal{GP}\left(m_p(t, T), cov_p(t_i, T_i, t_j, T_j)\right). \tag{3.13}$$

Since the logarithm of zero coupon bonds prices is given by

$$\log p(t,T) = A(t,T) - B(t,T)r(t) \tag{3.14}$$

the mean and covariance $m_p(t,T)$ and $cov_p(t_1,T_1,t_2,T_2)$ are given by

$$\begin{aligned} m_p(t,T) &= A(t,T) - B(t,T)m_r(t) \qquad (3.15)\\ &= \left(\theta - \frac{\sigma^2}{2k^2}\right)\left(\frac{1-e^{-k(T-t)}}{k} - T + t\right) \\ &\quad - \frac{\sigma^2\left(1-e^{-k(T-t)}\right)^2}{4k^3} \\ &\quad - \frac{\left(1-e^{-k(T-t)}\right)\left(r(s)e^{-k(t-s)} + \theta\left(1-e^{-k(t-s)}\right)\right)}{k} \qquad (3.16) \end{aligned}$$

and

$$\begin{aligned} &cov_p(t_1,T_1,t_2,T_2) \\ &= B(t_1,T_1)B(t_2,T_2)cov_r(t_1,t_2) \qquad (3.17)\\ &= \frac{\sigma^2 e^{-k(t_1+t_2)}\left(1-e^{-k(T_1-t_1)}\right)\left(1-e^{-k(T_2-t_2)}\right)\left(e^{2k\min(t_1,t_2)} - e^{2ks}\right)}{2k^3}. \qquad (3.18) \end{aligned}$$

3.4 One Factor Machine Learning Gaussian Short Rate

Let:

- $\mathbf{x} = [t \quad T]^\top$;
- $y = \log p(\mathbf{x}) = \log p(t,T)$;
- $m_p(\mathbf{x}) = m_p(t,T)$;
- $cov_p(\mathbf{x}_i,\mathbf{x}_j) = cov_p(t_i,T_i,t_j,T_j)$.

Following Section 3.2, let matrix $\mathbf{X}$ collect a set of vectors $\mathbf{x}^\diamond$ where the values of zero coupon log prices were observed, and let vector $\mathbf{y}$ collect the corresponding values $y^\diamond = \log p(\mathbf{x}^\diamond)$. Recall from Section 3.2 that $\mathcal{D} = (\mathbf{X},\mathbf{y})$ is the training set.

The "One Factor Machine Learning Gaussian Short Rate" is the Gaussian short rate process, $r_{\mathcal{D}}(t)$, underlying the zero coupon bond prices

$$p_{\mathcal{D}}(t,T) = E\left[e^{-\int_t^T r_{\mathcal{D}}(u)du}\right] \tag{3.19}$$

where $\log p_{\mathcal{D}}(t,T) = \log p_{\mathcal{D}}(\mathbf{x})$ is the conditioned on zero coupon bonds log prices Gaussian process

$$\log p_{\mathcal{D}}(\mathbf{x}) \sim \mathcal{GP}\left(m_{p_{\mathcal{D}}}(\mathbf{x}), cov_{p_{\mathcal{D}}}(\mathbf{x}_i, \mathbf{x}_j)\right) \tag{3.20}$$

using $\log p(\mathbf{x})$ as prior.
Given equations 3.3 and 3.4, $m_{p_{\mathcal{D}}}(\mathbf{x})$ and $cov_{p_{\mathcal{D}}}(\mathbf{x}_i, \mathbf{x}_i)$ are given by

$$m_{p_{\mathcal{D}}}(\mathbf{x}) = m_p(\mathbf{x}) + \mathbf{K}_{\mathbf{X},\mathbf{x}}^{\top}\mathbf{K}^{-1}(\mathbf{y} - \mathbf{m}) \tag{3.21}$$

and

$$cov_{p_{\mathcal{D}}}(\mathbf{x}_i, \mathbf{x}_j) = cov_p(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{X},\mathbf{x}_i}^{\top}\mathbf{K}^{-1}\mathbf{K}_{\mathbf{X},\mathbf{x}_j} \tag{3.22}$$

where

- $\mathbf{m}$ is the prior mean on the training set. It results from applying $m_p(\mathbf{x})$ function (Equation 3.16) on all $\mathbf{X}$ collected vectors;
- $\mathbf{K}$ is the prior covariance matrix on the training set. It results from applying $cov_p(\mathbf{x}_i, \mathbf{x}_j)$ function (Equation 3.18) on all pairs of $\mathbf{X}$ collected vectors;
- $\mathbf{K}_{\mathbf{X},\mathbf{x}}$ is the prior covariance between every vector in the training set and $\mathbf{x}$. It results from applying $cov_p(\mathbf{x}_i, \mathbf{x}_j)$ function (Equation 3.18) on all pairs composed by each $\mathbf{X}$ collected vector, and the $\mathbf{x}$ vector.

3.4.1 Properties

1. The model supports a single observation y on each vector $\mathbf{x} = [t \quad T]^{\top}$.

   Given Equation 3.18, the observation of more than one log price y, of a T maturity zero coupon bond, at time t, would result in two equal lines in matrix $\mathbf{K}$, which would not be invertible, as required by Equations 3.21 and 3.22.

2. The model only supports the observation of a single T maturity zero coupon bond log price on each time t.

As in the previous property, given Equation 3.17, the observation of more than one T maturity zero coupon bond log price, on each time t, would result in two equal lines in matrix $\mathbf{K}$.

3. Despite the parameters values, all supported observations are automatically fitted.

Given Equation 3.21, the model mean, $m_{p_{\mathcal{D}}}(\mathbf{x})$, on each training observation, $y^{\diamond} = \log p(\mathbf{x}^{\diamond})$, equals $y^{\diamond}$:

$$m_{p_{\mathcal{D}}}(\mathbf{x}^{\diamond}) = y^{\diamond}. \tag{3.23}$$

Furthermore, given Equation 3.22, the model variance on each training observation equals zero:

$$cov_{p_{\mathcal{D}}}(\mathbf{x}^{\diamond}, \mathbf{x}^{\diamond}) = 0. \tag{3.24}$$

3.4.2 SDE

In order to obtain the "One Factor Machine Learning Gaussian Short Rate" SDE we assume, at a first moment, that the conditioned on data short rate follows, under the arbitrage free, risk neutral measure, the deterministic time dependent parameters SDE of the generalized Hull and White[1] model.

Then, writing the generalized Hull and White model zero coupon bond log prices mean and covariance functions, making them equal to the corresponding "One Factor Machine Learning Gaussian Short Rate" functions, and solving in order to the deterministic time dependent parameters, we get the SDE parameters and confirm our initial assumption.

Let the "One Factor Machine Learning Gaussian Short Rate" short rate, $r_{\mathcal{D}}(t)$, follow, under the arbitrage free, risk neutral measure, the generalized Hull and White model SDE (J. Hull and White 1990a)

$$\mathrm{d}r_{\mathcal{D}}(t) = (\theta(t) - \alpha(t) r_{\mathcal{D}}(t))\mathrm{d}t + \sigma(t)\mathrm{d}W(t). \tag{3.25}$$

The solution of Equation 3.25 is

$$r_{\mathcal{D}}(t) = r_{\mathcal{D}}(s) e^{H(s)-H(t)} + e^{-H(t)} \int_s^t e^{H(u)} \theta(u) du + e^{-H(t)} \int_s^t e^{H(u)} \sigma(u) \mathrm{d}W(u) \tag{3.26}$$

[1] The generalized Hull and White model is also known as the extended Vasicek model, as it is a Vasicek model extended with time dependent parameters.

where

$$H(t) = \int_0^t \alpha(u)\mathrm{d}u. \tag{3.27}$$

Under this assumption:

1. The short rate mean, $m_{r_\mathcal{D}}(t)$, and covariance, $cov_{r_\mathcal{D}}(t_1, t_2)$, functions are given by

$$m_{r_\mathcal{D}}(t) = r_\mathcal{D}(s)e^{H(s)-H(t)} + e^{-H(t)} \int_s^t e^{H(u)}\theta(u)\mathrm{d}u \tag{3.28}$$

and

$$cov_{r_\mathcal{D}}(t_1, t_2) = e^{-H(t_1)-H(t_2)} \int_s^{min(t_1,t_2)} e^{2H(u)}\sigma^2(u)\mathrm{d}u; \tag{3.29}$$

2. The zero coupon bond prices are given by (Bingham and Kiesel 2004)

$$p_\mathcal{D}(t, T) = e^{A_\mathcal{D}(t,T)-B_\mathcal{D}(t,T)r_\mathcal{D}(t)} \tag{3.30}$$

where

$$B_\mathcal{D}(t, T) = e^{H(t)} \int_t^T e^{-H(u)}du \tag{3.31}$$

and

$$\begin{aligned} A_\mathcal{D}(t, T) \quad = \quad & \int_t^T \int_t^s e^{-H(u)-H(s)} \int_t^u e^{2H(v)}\sigma^2(v)dvduds \\ & - \int_t^T e^{-H(u)} \int_t^u e^{H(v)}\theta(v)dvdu; \end{aligned} \tag{3.32}$$

3. The model is affine and the zero coupon bond log prices, $\log p_\mathcal{D}(t, T)$, are Gaussian

$$\log p_\mathcal{D}(\mathbf{x}) \sim \mathcal{GP}\left(m_{p_\mathcal{D}}(\mathbf{x}), cov_{p_\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)\right) \tag{3.33}$$

4. The zero coupon bond log prices mean and covariance are given by

$$m_{p_\mathcal{D}}(t, T) = A_\mathcal{D}(t, T) - B_\mathcal{D}(t, T)m_{r_\mathcal{D}}(t) \tag{3.34}$$

and

$$cov_{p_{\mathcal{D}}}(t_1, T_1, t_2, T_2) = B_{\mathcal{D}}(t_1, T_1) B_{\mathcal{D}}(t_2, T_2) cov_{r\mathcal{D}}(t_1, t_2). \tag{3.35}$$

3.4.2.1 Parameter $\alpha(t)$

In order to get parameter $\alpha(t)$ we first note that, under the risk neutral measure, the initial value of the short rate conditioned on data, $r_{\mathcal{D}}(s)$, equals the prior initial value, $r(s)$, because the observations that distinguish the two processes occur, by definition of the initial time, at times greater than s.

Then, expanding the zero coupon bond log prices mean conditioned on data, $m_{p_{\mathcal{D}}}(t, T)$, in Equations 3.34 and 3.21, and making the term with $r_{\mathcal{D}}(s)$ equal to the term with $r(s)$ we get

$$B_{\mathcal{D}}(t, T) r_{\mathcal{D}}(s) e^{H(s)-H(t)} = B(t, T) r(s) e^{-k(t-s)} \tag{3.36}$$

$$B_{\mathcal{D}}(t, T) r(s) e^{H(s)-H(t)} = B(t, T) r(s) e^{-k(t-s)}. \tag{3.37}$$

Making the exponential functions of s equal

$$\begin{aligned} \Leftrightarrow \quad & e^{H(s)-H(t)} = e^{-k(t-s)} \\ \Leftrightarrow \quad & \int_s^t \alpha(u) \mathrm{d}u = k(t-s) \\ \Leftrightarrow \quad & \alpha(t) = k \end{aligned} \tag{3.38}$$

Therefore,

$$H(t) = \int_0^t \alpha(u) \mathrm{d}u = \int_0^t k \mathrm{d}u = kt. \tag{3.39}$$

Also, substituting Equation 3.39 in Equation 3.31 shows that $B_{\mathcal{D}}(t, T)$ becomes equal to $B(t, T)$

$$\begin{aligned} B_{\mathcal{D}}(t,T) &= e^{H(t)} \int_t^T e^{-H(u)} du \\ &= e^{kt} \int_t^T e^{-ku} du \\ &= \frac{1}{k}\left(1 - e^{-k(T-t)}\right) \\ &= B(t,T). \end{aligned} \tag{3.40}$$

3.4.2.2 Parameter $\sigma(t)$

Given Equations 3.35 and 3.40, and setting $cov_{p_{\mathcal{D}}}(t_1,T_1,t_2,T_2)$, in Equations 3.35 and 3.22, equal

$$\begin{aligned} & B(t_i,T_i)B(t_j,T_j)cov_{r\mathcal{D}}(t_i,t_j) = \\ & \quad cov_p(t_i,T_i,t_j,T_j) - \mathbf{K}^\top_{\mathbf{X},[t_iT_i]^\top}\mathbf{K}^{-1}\mathbf{K}_{\mathbf{X},[t_jT_j]^\top}. \end{aligned} \tag{3.41}$$

Inserting Equation 3.17,

$$\begin{aligned} B(t_i,T_i)B(t_j,T_j)cov_{r\mathcal{D}}(t_i,t_j) &= B(t_i,T_i)B(t_j,T_j)cov_r(t_i,t_j) \\ & \quad -\mathbf{K}^\top_{\mathbf{X},[t_iT_i]^\top}\mathbf{K}^{-1}\mathbf{K}_{\mathbf{X},[t_jT_j]^\top} \\ \Leftrightarrow cov_{r\mathcal{D}}(t_i,t_j) &= cov_r(t_i,t_j) - \frac{\mathbf{K}^\top_{\mathbf{X},[t_iT_i]^\top}}{B(t_i,T_i)}\mathbf{K}^{-1}\frac{\mathbf{K}_{\mathbf{X},[t_jT_j]^\top}}{B(t_j,T_j)} \end{aligned} \tag{3.42}$$

$\mathbf{K}_{\mathbf{X},[t\ T]^\top}$ is a vector of prior covariances, $cov_p(t_i,T_i,t_j,T_j)$, between every $T^\diamond$ maturity zero coupon bond log price at time $t^\diamond$ in the training set, and a T maturity zero coupon bond log price at time t. Given equation 3.17,

$$\begin{aligned} \frac{cov_p(t^\diamond,T^\diamond,t,T)}{B(t,T)} &= \frac{B(t^\diamond,T^\diamond)B(t,T)cov_r(t^\diamond,t)}{B(t,T)} \\ \Leftrightarrow \frac{cov_p(t^\diamond,T^\diamond,t,T)}{B(t,T)} &= B(t^\diamond,T^\diamond)cov_r(t^\diamond,t). \end{aligned} \tag{3.43}$$

Therefore

$$cov_{r\mathcal{D}}(t_i,t_j) \quad = \quad cov_r(t_i,t_j) - \mathbf{V}_{\mathbf{X},t_i}^\top \mathbf{K}^{-1}\mathbf{V}_{\mathbf{X},t_j} \tag{3.44}$$

where each element $\mathbf{v}_{[t^\diamond\ T^\diamond]^\top,t}$ of vector $\mathbf{V}_{\mathbf{X},t}$ is given by

$$\mathbf{v}_{[t^\diamond\ T^\diamond]^\top,t} = B(t^\diamond,T^\diamond)cov_r(t^\diamond,t). \tag{3.45}$$

Given Equations 3.29 and 3.44

$$\begin{aligned} e^{-k(t_i+t_j)}\int_s^{min(t_i,t_j)} e^{2ku}\sigma^2(u)\mathrm{d}u &= \frac{\sigma^2 e^{-k(t_i+t_j)}}{2k}\left(e^{2k\min(t_i,t_j)}-e^{2ks}\right) \\ &\quad -\mathbf{V}_{\mathbf{X},t_i}^\top\mathbf{K}^{-1}\mathbf{V}_{\mathbf{X},t_j}. \end{aligned} \tag{3.46}$$

Equation 3.46 applies to every t_i and t_j, in particular if $t_i = t_j = t$

$$e^{-2kt}\int_s^t e^{2ku}\sigma^2(u)\mathrm{d}u \quad = \quad \frac{\sigma^2 e^{-2kt}}{2k}\left(e^{2kt}-e^{2ks}\right) - \mathbf{V}_{\mathbf{X},t}^\top\mathbf{K}^{-1}\mathbf{V}_{\mathbf{X},t} \tag{3.47}$$

Therefore

$$\begin{aligned} \int_s^t e^{2ku}\sigma^2(u)\mathrm{d}u &= \frac{\sigma^2}{2k}\left(e^{2kt}-e^{2ks}\right) - \frac{\mathbf{V}_{\mathbf{X},t}^\top}{e^{-kt}}\mathbf{K}^{-1}\frac{\mathbf{V}_{\mathbf{X},t}}{e^{-kt}} \\ &= \frac{\sigma^2}{2k}\left(e^{2kt}-e^{2ks}\right) - \mathbf{U}_{\mathbf{X},t}^\top\mathbf{K}^{-1}\mathbf{U}_{\mathbf{X},t} \end{aligned} \tag{3.48}$$

where each element $\mathbf{u}_{[t^\diamond\ T^\diamond]^\top,t}$ of vector $\mathbf{U}_{\mathbf{X},t}$ is given by

$$\begin{aligned} \mathbf{u}_{[t^\diamond\ T^\diamond]^\top,t} &= \frac{B(t^\diamond,T^\diamond)}{e^{-kt}}cov_r(t^\diamond,t) \\ &= \frac{B(t^\diamond,T^\diamond)}{e^{-kt}}\sigma^2 e^{-k(t^\diamond+t)}\frac{1}{2k}\left(e^{2k\min(t^\diamond,t)}-e^{2ks}\right) \\ &= B(t^\diamond,T^\diamond)\sigma^2 e^{-kt^\diamond}\frac{1}{2k}\left(e^{2k\min(t^\diamond,t)}-e^{2ks}\right) \end{aligned} \tag{3.49}$$

Differentiating both sides of Equation 3.48 w.r.t. t

$$
\begin{aligned}
e^{2kt}\sigma^2(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\sigma^2}{2k}\left(e^{2kt}-e^{2ks}\right)-\mathbf{U}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}\mathbf{U}_{\mathbf{X},t}\right)\\
&= \sigma^2 e^{2kt}-2\mathbf{U}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_{\mathbf{X},t}\\
&= \sigma^2 e^{2kt}-2\mathbf{U}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}\mathbf{Q}_{\mathbf{X},t} \qquad (3.50)
\end{aligned}
$$

where each element $\mathbf{q}_{[t^\diamond\ T^\diamond]^\top,t}$ of vector $\mathbf{Q}_{\mathbf{X},t}$ is given by

$$
\mathbf{q}_{[t^\diamond\ T^\diamond]^\top,t} = \mathbb{1}_{\mathbb{R}^+}(t^\diamond - t)B(t^\diamond, T^\diamond)\sigma^2 e^{-kt^\diamond}e^{2kt}. \qquad (3.51)
$$

Therefore

$$
\sigma^2(t) = \sigma^2 - 2\mathbf{U}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}\mathbf{R}_{\mathbf{X},t} \qquad (3.52)
$$

where each element $\mathbf{r}_{[t^\diamond\ T^\diamond]^\top,t}$ of vector $\mathbf{R}_{\mathbf{X},t}$ is given by

$$
\mathbf{r}_{[t^\diamond\ T^\diamond]^\top,t} = \mathbb{1}_{\mathbb{R}^+}(t^\diamond - t)B(t^\diamond, T^\diamond)\sigma^2 e^{-kt^\diamond} \qquad (3.53)
$$

and

$$
\sigma(t) = \left(\sigma^2 - 2\mathbf{U}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}\mathbf{R}_{\mathbf{X},t}\right)^{\frac{1}{2}}. \qquad (3.54)
$$

3.4.2.3 Parameter $\Theta(t)$

Expanding $m_{p_{\mathcal{D}}}(t,T)$ in Equations 3.34 and 3.21

$$
A_{\mathcal{D}}(t,T) - B_{\mathcal{D}}(t,T)m_{r_{\mathcal{D}}}(t) = m_p(t,T) + \mathbf{K}_{\mathbf{X},[t\ T]^\top}^{\top}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}) \qquad (3.55)
$$

using Equations 3.34 ($A_{\mathcal{D}}(t,T)$), 3.40 ($B_{\mathcal{D}}(t,T)$ equal to $B(t,T)$), 3.28 ($m_{r_{\mathcal{D}}}(t)$), 3.38 ($\alpha(t) = k$) and 3.16 ($m_p(t,T)$), canceling the terms $B(t,T)r(s)e^{-k(t-s)}$ on both sides (Equation 3.36), keeping the terms with $\Theta(t)$ in the left hand side and moving all other terms to the right hand side, equality of Equation 3.55 becomes

$$
\begin{aligned}
& -\int_t^T e^{-ku}\int_t^u e^{kv}\theta(v)dvdu - \frac{1-e^{-k(T-t)}}{k}e^{-kt}\int_s^t e^{ku}\theta(u)\mathrm{d}u && (3.56)\\
= \; & -\int_t^T\int_t^s e^{-k(u+s)}\int_t^u e^{2kv}\sigma^2(v)dvduds && (3.57)\\
& +\left(\theta-\frac{\sigma^2}{2k^2}\right)\left(\frac{1-e^{-k(T-t)}}{k}-T+t\right) && (3.58)\\
& -\frac{\sigma^2\left(1-e^{-k(T-t)}\right)^2}{4k^3} && (3.59)\\
& -\frac{1-e^{-k(T-t)}}{k}\theta\left(1-e^{-k(t-s)}\right) && (3.60)\\
& +\mathbf{K}_{\mathbf{X},[t\;T]^\top}^\top\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}). && (3.61)
\end{aligned}
$$

Differentiating the left hand side in Expression 3.56 w.r.t. T

$$
\begin{aligned}
& \frac{\mathrm{d}}{\mathrm{d}T}\left(-\int_t^T e^{-ku}\int_t^u e^{kv}\theta(v)dvdu - \frac{1-e^{-k(T-t)}}{k}e^{-kt}\int_s^t e^{ku}\theta(u)\mathrm{d}u\right)\\
= \; & -e^{-kT}\int_t^T e^{kv}\theta(v)dv - e^{-kT}\int_s^t e^{ku}\theta(u)\mathrm{d}u && (3.62)
\end{aligned}
$$

Equation 3.62 applies to every $t \leq T$, in particular, if $t = T$ becomes

$$
\begin{aligned}
& \left.\left(\frac{\mathrm{d}}{\mathrm{d}T}\left(-\int_t^T e^{-ku}\int_t^u e^{kv}\theta(v)dvdu - \frac{1-e^{-k(T-t)}}{k}e^{-kt}\int_s^t e^{ku}\theta(u)\mathrm{d}u\right)\right)\right|_{t=T}\\
= \; & -e^{-kT}\int_s^T e^{ku}\theta(u)\mathrm{d}u. && (3.63)
\end{aligned}
$$

Regarding the right hand side part in Expression 3.57, differentiating w.r.t. T, and making $t = T$, cancel this part

$$
\begin{aligned}
& \left.\left(-\frac{\mathrm{d}}{\mathrm{d}T}\int_t^T\int_t^s e^{-k(u+s)}\int_t^u e^{2kv}\sigma^2(v)dvduds\right)\right|_{t=T}\\
= \; & \left.\left(-\int_t^T e^{-k(u+T)}\int_t^u e^{2kv}\sigma^2(v)dvdu\right)\right|_{t=T}\\
= \; & 0. && (3.64)
\end{aligned}
$$

Proceeding with the right hand side part in Expressions 3.58, 3.59 and 3.60, differentiating w.r.t. T, and making $t = T$,

$$\begin{aligned}
&\left(\frac{\mathrm{d}}{\mathrm{d}T}\left(\left(\theta-\frac{\sigma^2}{2k^2}\right)\left(\frac{1-e^{-k(T-t)}}{k}-T+t\right)\right.\right. \\
&-\frac{\sigma^2\left(1-e^{-k(T-t)}\right)^2}{4k^3} \\
&\left.\left.-\frac{1-e^{-k(T-t)}}{k}\theta\left(1-e^{-k(t-s)}\right)\right)\right)\Bigg|_{t=T} \\
=\; &\theta\left(e^{-k(T-s)}-1\right). \qquad (3.65)
\end{aligned}$$

Regarding the part in Expression 3.61, recall that each element $\mathbf{k}_{[t^\diamond\ T^\diamond]^\top,[t\ T]^\top}$ of vector $\mathbf{K}_{\mathbf{X},[t\ T]^\top}$ is given by the covariance $cov_p(\mathbf{x}_i,\mathbf{x}_j)$, in Equation 3.17, between each vector $[t^\diamond\ T^\diamond]^\top$ in the training set, and $[t\ T]^\top$

$$\mathbf{k}_{[t^\diamond\ T^\diamond]^\top,[t\ T]^\top}=B(t^\diamond,T^\diamond)B(t,T)cov_r(t^\diamond,t). \qquad (3.66)$$

Given that

$$\left(\frac{\mathrm{d}}{\mathrm{d}T}B(t,T)\right)\Bigg|_{t=T}=1, \qquad (3.67)$$

differentiating w.r.t. T, and making $t=T$, the part in Expression 3.61

$$\begin{aligned}
&\left(\frac{\mathrm{d}}{\mathrm{d}T}\left(\mathbf{K}^\top_{\mathbf{X},[t\ T]^\top}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m})\right)\right)\Bigg|_{t=T} \\
=\; &\mathbf{N}^\top_{\mathbf{X},T}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}) \qquad (3.68)
\end{aligned}$$

where each element, $\mathbf{n}_{[t^\diamond\ T^\diamond]^\top,T}$, of vector $\mathbf{N}_{\mathbf{X},T}$ is given by

$$\begin{aligned}
\mathbf{n}_{[t^\diamond\ T^\diamond]^\top,T} &= B(t^\diamond,T^\diamond)cov_r(t^\diamond,T) \\
&= \frac{\left(1-e^{-k(T^\diamond-t^\diamond)}\right)}{k}\sigma^2e^{-k(t^\diamond+T)}\frac{1}{2k}\left(e^{2k\min(t^\diamond,T)}-e^{2ks}\right). \qquad (3.69)
\end{aligned}$$

Grouping together the results in Equations 3.63, 3.65 and 3.68

$$\begin{aligned}
&-e^{-kT}\int_s^T e^{ku}\theta(u)\mathrm{d}u=\theta\left(e^{-k(T-s)}-1\right)+\mathbf{N}^\top_{\mathbf{X},T}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}) \\
\Leftrightarrow\quad &\int_s^T e^{ku}\theta(u)\mathrm{d}u=-\theta e^{ks}+\theta e^{kT}-\frac{\mathbf{N}^\top_{\mathbf{X},T}}{e^{-kT}}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}). \qquad (3.70)
\end{aligned}$$

Differentiating w.r.t. T, Equation 3.70

$$e^{kT}\theta(T) = k\theta e^{kT} - \frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{\mathbf{N}_{\mathbf{X},T}^{\top}}{e^{-kT}}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m})\right) \tag{3.71}$$

and given that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{\mathbf{n}_{[t^\diamond\ T^\diamond]^\top,T}}{e^{-kT}}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}T}\left(\frac{\left(1-e^{-k(T^\diamond-t^\diamond)}\right)}{k}\sigma^2 e^{-kt^\diamond}\frac{1}{2k}\left(e^{2k\min(t^\diamond,T)}-e^{2ks}\right)\right) \\ &= \mathbb{1}_{\mathbb{R}^+}(t^\diamond - T)\frac{\left(1-e^{-k(T^\diamond-t^\diamond)}\right)}{k}\sigma^2 e^{-kt^\diamond}e^{2kT}, \end{aligned} \tag{3.72}$$

$\theta(T)$ is given by

$$\theta(T) = k\theta - \mathbf{Z}_{\mathbf{X},T}^{\top}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}) \tag{3.73}$$

where each element $\mathbf{z}_{[t^\diamond\ T^\diamond]^\top,T}$ of vector $\mathbf{Z}_{\mathbf{X},T}$ is given by

$$\mathbf{z}_{[t^\diamond\ T^\diamond]^\top,T} = \mathbb{1}_{\mathbb{R}^+}(t^\diamond - T)\frac{\left(1-e^{-k(T^\diamond-t^\diamond)}\right)}{k}\sigma^2 e^{-kt^\diamond}e^{kT}. \tag{3.74}$$

Finally, since Equation 3.73 applies to all $T > s$, it can be rewritten as a function of t

$$\theta(t) = k\theta - \mathbf{Z}_{\mathbf{X},t}^{\top}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}). \tag{3.75}$$

3.4.3 Learning the parameters

The prior parameters, $r(s)$, θ, k and σ, are obtained, directly under the risk neutral measure, by maximizing the prior log likelihood, L, of the market observed zero coupon bonds log prices in the training set $\mathcal{D}$, given the parameters

$$L = -\frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m}) - \frac{n}{2}\log(2\pi) \tag{3.76}$$

using the closed forms of the log likelihood derivative w.r.t. to each of the

Parameter	Fixed Value	Mean	Std. Dev.	95% CI		
$r(0)$	0.035	0.035	0.0016	0.032	to	0.038
k	0.26	0.26	0.027	0.20	to	0.031
θ	0.08	0.081	0.006	0.070	to	0.093
σ	0.04	0.039	0.0014	0.037	to	0.042

Table 3.1: Prior parameters $r(0)$, k, θ and σ, mean, standard deviation and 95% confidence interval, learned from 1000 simulated data calibrations.

parameters (Sousa, Esquível, and Gaspar 2012).

3.5 Simulation

In order to evaluate the ability of learning parameters from data, we executed the following simulation procedure:

1. Fix the initial time to zero, and the time period to 1 year. Fix the time increment to $1/260$ (assuming quotes on 5 working days per week, 52 weeks per year);
2. Fix the set of possible maturities to 7, 14 and 21 days, and 1 to 12 months (assuming 30 days months);
3. Fix the set of prior parameters to $r(0) = 0.035$, $k = 0.26$, $\theta = 0.08$ and $\sigma = 0.04$ (the approximate values obtained in the real data case described in Section 3.6);
4. Use Equations 3.7 and 3.8 simulate one prior short rate trajectory;
5. Use Equation 3.14 to compute one, randomly selected maturity, zero coupon bond log price for each day;
6. Learn the prior parameters using the simulated data, the method described in Section 3.4.3 and the Conjugate Gradient method, available in Wolfram Mathematica 9 (Research 2013);
7. Repeat the previous steps 4 to 6, for 1000 trajectories.

Figure 3.1 shows the learned parameters histograms and Table 3.1 the corresponding mean, standard deviation, and 95% confidence intervals.

As it can be observed in Table 3.1, all the confidence intervals include the corresponding fixed value in step 3.

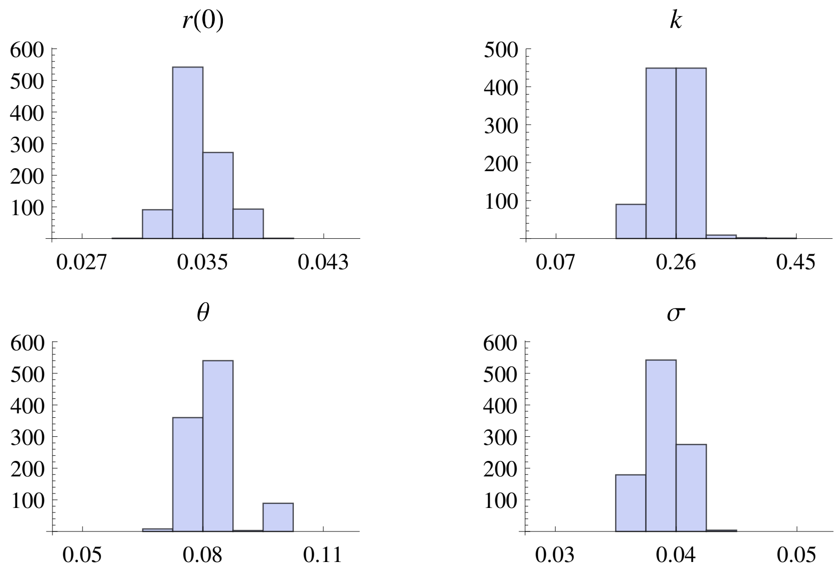

Figure 3.1: Prior parameters histograms, learned from simulated data.

For purposes of illustration, Figure 3.2 shows the short rate SDE deterministic time dependent parameters, $\alpha(t)$, $\theta(t)$ and $\sigma(t)$, of Equations 3.38, 3.75, and 3.54, respectively, of one of the simulated trajectories, computed from the learned prior parameters and the simulated data. These are the deterministic time dependent parameters of the short rate SDE in Equation 3.25 that exactly fit the simulated zero coupon bond log prices observations in the training set (the additional version of $\sigma(t)$, in the vicinity of 1.0, shows the detailed evolution of $\sigma(t)$, which can not be observed with the original time scale).

3.6 Real data

In this section we model the Euribor rates, quoted by the Portuguese bank Caixa Geral de Depósitos (CGD), which belongs to the Euribor panel banks. All the data used is publicly available at the Euribor Internet site[2].

We choose the crisis years of 2007 and 2008 as the period to model.

Given the model limitation in Property 2, Section 3.4.1, we randomly selected one of the 15 Euribor maturities to get one observation for each day in the chosen period. The selected Euribor rates were converted to the equivalent zero coupon bond log prices and used as the training set.

Table 3.2 shows the prior parameters learned from the real data used as the training set.

[2] `http://www.euribor-ebf.eu/`

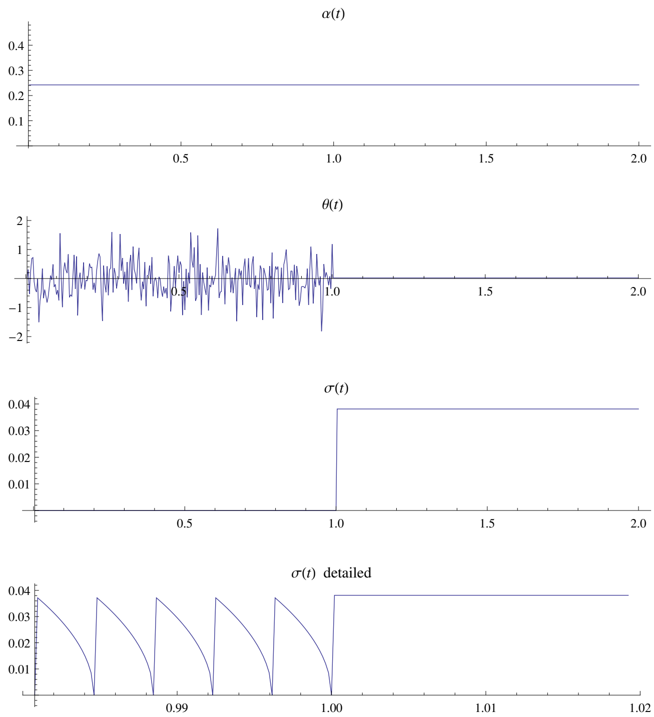

Figure 3.2: Short rate SDE deterministic time dependent parameters $\alpha(t)$, $\theta(t)$ and $\sigma(t)$, for one of the simulated trajectories.

Prior Parameter	$r(0)$	k	θ	σ
Learned Value	0.0349	0.2661	0.0826	0.0381

Table 3.2: Euribor model prior parameters, learned from one randomly selected maturity quote per day, from CGD bank, during 2007 and 2008.

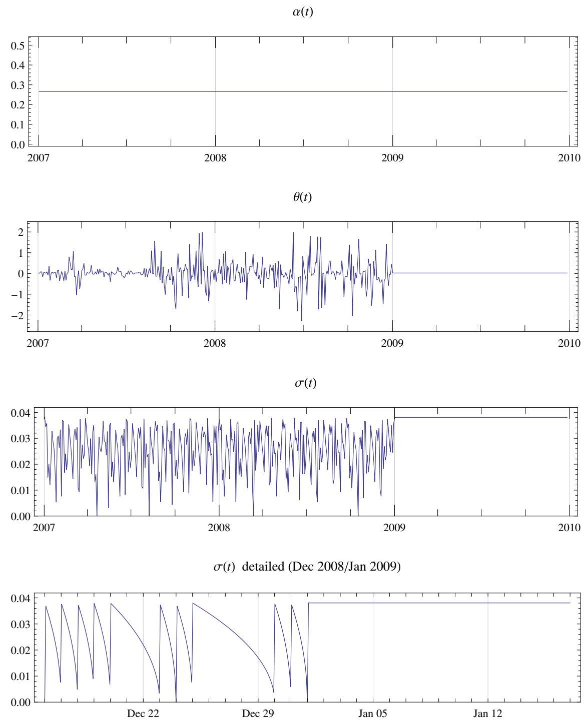

Figure 3.3: Short rate SDE deterministic time dependent parameters $\alpha(t)$, $\theta(t)$ and $\sigma(t)$, for the Euribor, quoted by CGD during 2007 and 2008, using a single randomly selected maturity per day.

Figure 3.3 illustrates the short rate SDE deterministic time dependent parameters, $\alpha(t)$, $\theta(t)$ and $\sigma(t)$, in Equation 3.25, computed from the learned prior parameters and data, using Equations 3.38, 3.75, and 3.54, respectively.

3.7 Conclusions

In this paper we propose to model the short rate, under the arbitrage free risk neutral measure, as a conditioned on zero coupon bonds log prices Gaussian process.

All model parameters are learned directly under the risk neutral measure, using zero coupon bonds log prices only.

The model supports observations of zero coupon bonds with distinct maturities limited to one observation per time instant. All the supported observations are automatically fitted.

4

Brownian Bridge and other Path Dependent Gaussian Processes Vectorial Simulation

4.1 Preamble

With the exception of this preamble and minor notation changes, this chapter contains the paper *Brownian Bridge and other Path Dependent Gaussian Processes Vectorial Simulation*, joint work with Prof. Manuel Esquível and Prof. Raquel Gaspar, which is submitted for publication in the journal Communications in Statistics-Simulation and Computation, by Taylor & Francis. The current submission status is "minor changes".

Both the iterative and the vectorial procedures for simulating the Wiener process are widely known, and described in reference books such as Glasserman 2003. However, regarding the Brownian bridge, only the iterative procedure is described.

In this paper we model the Brownian bridge using the Gaussian processes for machine learning regression framework, using the Wiener process, $W(t)$, as prior, and the single observation, $W(1) = 0$, the Brownian bridge condition, in the training set.

The main contributions of this paper are:

- Use the bridge mean vector and the covariance matrix, computed in a set of sampling instants, to simulate the bridge trajectories with the same vectorial procedure used to simulate any Gaussian vector;
- Extend the vectorial simulation procedure to other Gaussian processes priors, and for more than one conditions, by developing a general path dependent Gaussian process trajectories vectorial simulation framework;
- Show that the vectorial simulation procedure is relevant concerning the execution times of implementations with the interpreted programming languages widely used in today's research and development.

Abstract

The iterative simulation of the Brownian bridge is well known. In this paper we present a vectorial simulation alternative, based on Gaussian processes for machine learning regression, that is suitable for interpreted programming languages implementations.

We extend the vectorial simulation of path dependent trajectories to other Gaussian processes, namely, sequences of Brownian bridges, geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process.

4.2 Introduction

Interpreted programming languages like Sage, Octave, Mathematica and Matlab are currently important frameworks in research and development.

In these programming languages it is crucial to use vectorial algorithms instead of iterative ones, in order to achieve the execution speeds of compiled languages. This is because vectorial operations are typically supported by built-in functions which are implemented by optimized machine code.

The iterative simulation of the Brownian bridge is well known (Glasserman 2003) (Group 2012). In this paper we present a vectorial simulation alternative based on Gaussian processes for machine learning regression that, is suitable for interpreted programming languages implementations.

We extend the vectorial simulation of path dependent trajectories to other Gaussian processes, namely, sequences of Brownian bridges, geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process, by developing a Gaussian path dependent trajectories simulation vectorial framework.

We illustrate the flexibility of the path dependent vectorial simulation procedure, by creating a 2D Wiener process representation of a Norbert Wiener photograph.

Simulation of Gaussian processes immediately spread with computers availability and became an important tool in many science areas, such as, mathematics (Kloeden and Platen 1992), finance (Glasserman 2003), engineering (Kasdin 1995), hydrology (Mandelbrot 1971) and geology (Alabert 1987), among many others.

In particular, the simulation of Brownian bridges, geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process, play an important role in Monte Carlo methods (Moskowitz and Caflisch 1996), securities pricing (Broadie and Glasserman 1997), communication networks (Paxson 1997) and particles motion (Gillespie 1996), respectively.

A wide range of path dependent Gaussian trajectories simulation methods exist, spanning from the earlier, based on sampling the unconditional distribution (Hoffman and Ribak 1991), Cholesky factorization (Davis 1987) or FFT (Dietrich and Newsam 1996), to the more recent efforts of implementing the old methods in the emerging parallel architectures (Garland, Le Grand, Nickolls, J. Anderson, Hardwick, Morton, Phillips, Zhang, and Volkov 2008) (Ltaief, Tomov, Nath, and Dongarra 2010) (Volkov and Demmel 2008).

In this paper we use the Cholesky method. Despite having known limitations (Jean-François 2000), it allows the extension to the path dependent case and we show that it is relevant concerning the execution speed of interpreted programming languages implementations.

4.3 Brownian bridge iterative simulation

A Brownian bridge is a standard Brownian motion W conditioned to $W(1) = 0$. The Brownian bridge condition $W(1) = 0$ can be generalized to other time instants greater than zero and to other values besides zero.

The standard Brownian motion W, defined in $\mathbb{R}_0^+$, is also called a Wiener process (Björk 2004) and has the following properties:

1. $W(0) = 0$;

2. W has independent increments, i.e. if $r < s \leq t < u$ then $W(u) - W(t)$ and $W(s) - W(r)$ are independent random variables;

3. For $s < t$ the random variable $W(t) - W(s)$ has the Gaussian distribution $\mathcal{N}(0, \sqrt{t-s})$;

4. W has almost surely continuous trajectories.

In addition, the Wiener process is a Gaussian process with mean function $m(t)$ and covariance function $cov(s,t)$:

$$m(t) = 0, \tag{4.1}$$

$$cov(s,t) = \min(s,t). \tag{4.2}$$

In order to illustrate the iterative simulation of a Brownian bridge trajectory B, consider that at some step we have $0 < u < s < t < 1$, $B(u) = \alpha$, $B(t) = \beta$ and we want to simulate the value $B(s)$ (Glasserman 2003). Given the Wiener process properties, the random vector $[B(u)B(s)B(t)]^T$ is Gaussian with mean vector and covariance matrix:

$$\begin{bmatrix} B(u) \\ B(s) \\ B(t) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} u & u & u \\ u & s & s \\ u & s & t \end{bmatrix} \right) \tag{4.3}$$

Therefore the conditional distribution $B(s)|B(u), B(t)$ is given by:

$$B(s)|B(u), B(t) \sim \mathcal{N}\left(\frac{(t-s)\alpha + (s-u)\beta}{t-u}, \frac{(s-u)(t-s)}{t-u} \right) \tag{4.4}$$

Thus, the value $B(s)$ is simulated by:

$$B(s) = \frac{(t-s)\alpha + (s-u)\beta}{t-u} + \sqrt{\frac{(s-u)(t-s)}{t-u}} Z, \tag{4.5}$$

where $Z \sim \mathcal{N}(0,1)$ is an increment independent of all Z values previously used in the simulation.

Finally, the iterative simulation of a Brownian bridge trajectory consists of starting with $u = 0$, $t = 1$, $B(0) = 0$, $B(1) = 0$ and iteratively filling a trajectory sample at time s (between u and t) with Equation 4.5, then moving one of the end points to the simulated sample and repeating the process until all trajectory samples are filled.

Figure 4.1 shows a sequence of 500 simulated independent Gaussian $\mathcal{N}(0,1)$ increments (white noise), and the corresponding simulated Wiener process and Brownian bridge trajectories (sampled uniformly 500 times between zero and

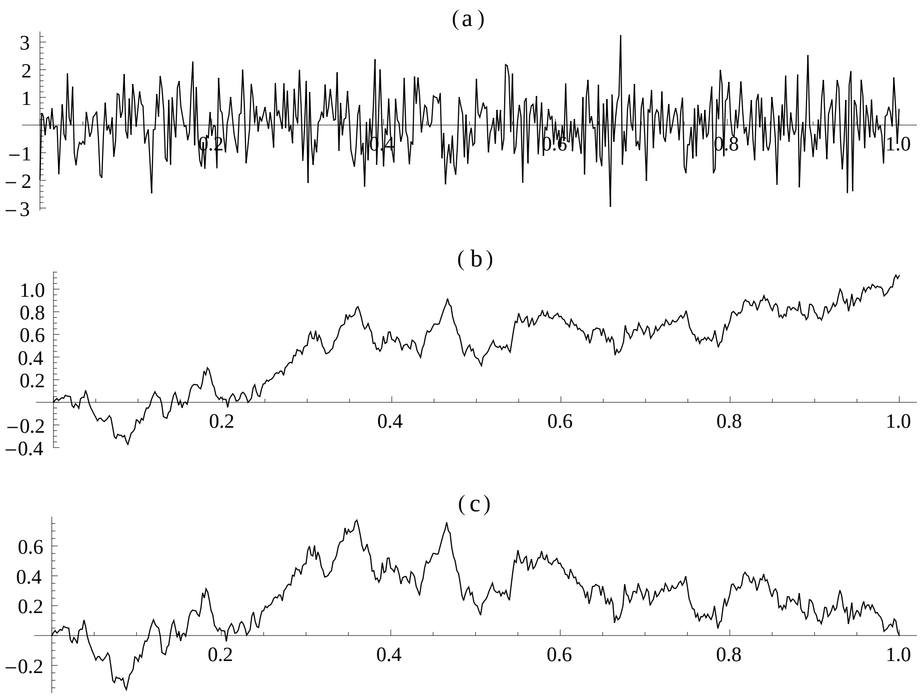

Figure 4.1: Simulated trajectories of: (a) white noise; (b) the corresponding Wiener process; (c) the corresponding Brownian bridge.

one). The Brownian bridge trajectory was simulated by the iterative procedure above.

4.4 Gaussian processes for machine learning

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

$$\mathcal{GP} \sim \mathcal{N}(m(\mathbf{x}), cov(\mathbf{x_1}, \mathbf{x_2})). \tag{4.6}$$

The Gaussian process defined by $m(\mathbf{x})$ and $cov(\mathbf{x_1}, \mathbf{x_2})$, in Equation 4.6, is the prior process.

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.

The regression function is the mean function of the process defined by all the trajectories of the prior process that passes through the training set. The regression confidence is the corresponding covariance function. The regression mean and the regression confidence define the posterior process on data.

As presented in the previous section, the Wiener process is a scalar Gaussian process

$$W \sim \mathcal{N}(m(t), cov(s,t)) \tag{4.7}$$

where $m(t)$ is given by Equation 4.1 and $cov(s,t)$ by Equation 4.2.

In this case the mapping f is the scalar mapping $y = f(t)$, where y is the value of W at time t. This reduces the training set to the pair of vectors $(\mathbf{t}, \mathbf{y})$, and the test set to vector $\mathbf{t}^\star$.

Since the process is Gaussian (Rasmussen and Williams 2005)

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m}^\star \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_\star \\ \mathbf{K}_\star^T & \mathbf{K}_{\star\star} \end{bmatrix} \right) \tag{4.8}$$

and

$$p(\mathbf{y}^\star | \mathbf{t}^\star, \mathbf{t}, \mathbf{y}) \sim \mathcal{N}\left(\mathbf{m}^\star + \mathbf{K}_\star^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}), \mathbf{K}_{\star\star} - \mathbf{K}_\star^T \mathbf{K}^{-1} \mathbf{K}_\star \right) \tag{4.9}$$

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

The conditional distribution

$$p(\mathbf{y}^\star | \mathbf{t}^\star, \mathbf{t}, \mathbf{y}) \tag{4.10}$$

corresponds to the posterior process on the data

$$\mathcal{GP}_\mathcal{D} \sim \mathcal{N}(m_\mathcal{D}(t), cov_\mathcal{D}(s,t)) \tag{4.11}$$

where

$$m_\mathcal{D}(t) = m(t) + \mathbf{K}_{\mathbf{t},t}^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \tag{4.12}$$

and

$$cov_\mathcal{D}(s,t) = cov(s,t) - \mathbf{K}_{\mathbf{t},s}^T \mathbf{K}^{-1} \mathbf{K}_{\mathbf{t},t} \tag{4.13}$$

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

Equation 4.12 is the regression function while Equation 4.13 is the regression confidence. Equations 4.12 and 4.13 are the central equations of Gaussian processes for machine learning regression.

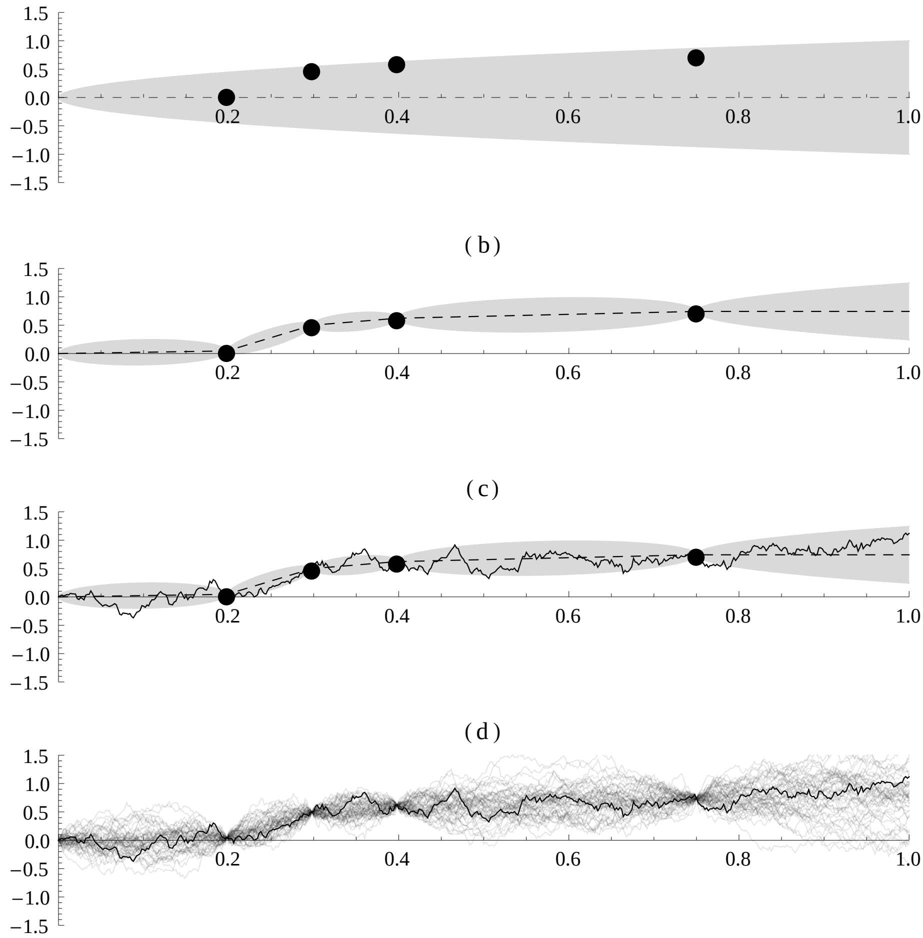

Figure 4.2: Gaussian processes for machine learning regression with the Wiener process as prior:(a) prior process mean (dashed), prior process two standard deviations band (gray) and the training set (circles); (b) regression function (dashed) and two standard deviations regression confidence band (gray); (c) training set simulated trajectory; (d) simulated Wiener process trajectories passing through the training set.

Figure 4.2 shows an example of Gaussian processes for machine learning regression using the Wiener process as the prior process and the set of 500 time instants, uniformly distributed between zero and one, as the test set.

4.5 Browning bridge vectorial simulation

The vectorial simulation of Brownian bridge trajectories is as achieved by joining Sections 4.3 and 4.4 .

Considering:

1. the Wiener process W with mean and covariance functions given by Equations 4.1 and 4.2;

2. the training set with the single pair $(t, y) = (1, 0)$ corresponding to the Brownian bridge condition $W(1) = 0$;
3. the test vector $\mathbf{t}^\star = [t_1, t_2, \ldots, t_n]^T$ where $t_1, t_2, \ldots, t_n$ are the time instants where to sample the Brownian bridge trajectory.

The Brownian bridge process B is Gaussian and the random vector $\mathbf{B} = B(\mathbf{t}^\star)$ is also Gaussian

$$\mathbf{B} \sim \mathcal{N}(\mathbf{m}_\mathcal{D}, \mathbf{cov}_\mathcal{D}) \tag{4.14}$$

where $\mathbf{m}_\mathcal{D}$ is $\mathbf{B}$ mean vector and $\mathbf{cov}_\mathcal{D}$ is $\mathbf{B}$ covariance matrix. The element i of vector $\mathbf{m}_\mathcal{D}$ is given by

$$\mathbf{m}_{\mathcal{D}i} = m_\mathcal{D}(t_i) \tag{4.15}$$

and the element i, j of matrix $\mathbf{cov}_\mathcal{D}$ is given by

$$\mathbf{cov}_{\mathcal{D}i,j} = cov_\mathcal{D}(t_i, t_j). \tag{4.16}$$

Functions $m_\mathcal{D}(t_i)$ and $cov_\mathcal{D}(t_i, t_j)$ are those of Equations 4.1 and 4.2.

Therefore a Brownian bridge trajectory can be simulated as any other Gaussian vector (Glasserman 2003), using:

$$\mathbf{B} = \mathbf{m}_\mathcal{D} + \mathbf{C}\mathbf{Z} \tag{4.17}$$

where $\mathbf{C}$ is the Cholesky decomposition of $\mathbf{cov}_\mathcal{D}$ and $\mathbf{Z}$ is a sample of the Gaussian random vector $\mathcal{N}(0, I)$.

Figure 4.3 shows some Brownian bridge trajectories simulated with the vectorial Equation 4.17. The solid black one was simulated with the Gaussian increments of Figure 4.1. Since Equation 4.17 is just a vectorial alternative to the iterative procedure of section 4.3, the solid black trajectory is, as it would be expected, equal to the Brownian bridge trajectory of Figure 4.1.

4.6 Execution time comparison

In order to compare the execution times of the iterative Brownian bridge simulation procedure of Section 4.3 and the vectorial procedure of Section 4.5, under an interpreted language framework, we implemented both using the Mathematica 8 language (Wolfram Research 2011) and tested the two alternatives on two different stages:

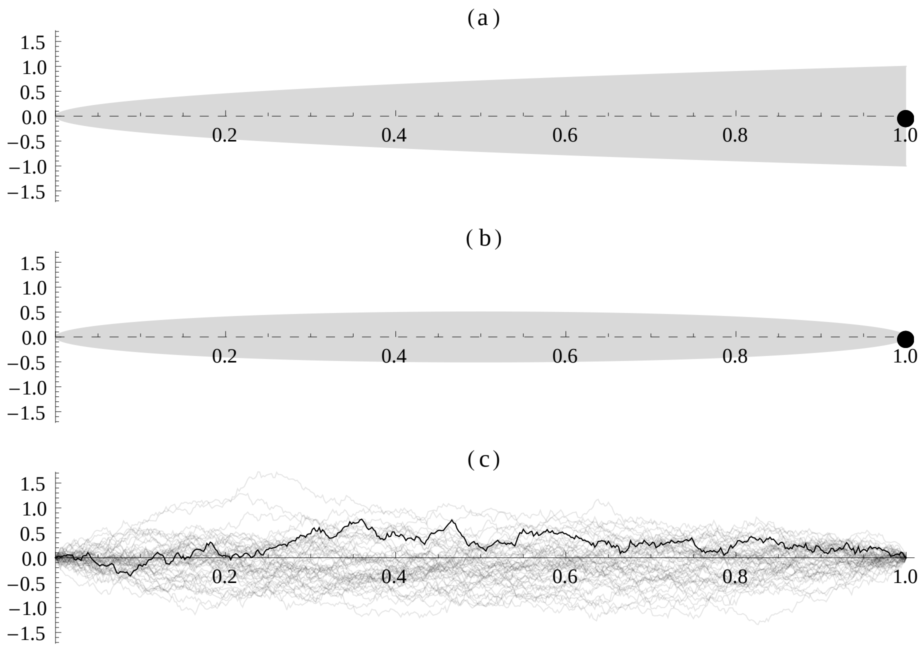

Figure 4.3: Brownian bridge trajectories simulated with the vectorial Equation 4.17: (a) prior process mean (dashed), prior process two standard deviations band (gray) and the training set (circle); (b) regression function (dashed) and two standard deviation regression confidence band (gray); (c) Brownian bridge simulated trajectories.

Stage 1 Inspired by the order of magnitude of typical setups found in financial markets, such as 250 daily prices per year, and stock indices with up to 500 stocks, we defined the reference task of generating 1000 Brownian bridges sampled uniformly 1000 times. This would correspond to simulate 4 years, of daily prices, of an index as bigger as twice the S&P500.

Stage 2 In order to evaluate the performance sensitivity to the task specification, we varied both the number of samples per trajectory and the number of trajectories.

Table 4.1 describes the execution system, the reference task and the execution times obtained for both alternatives on Stage 1.

As it would be expected, the Brownian bridge vectorial simulation with the Mathematica 8 language is faster than the iterative alternative. In the particular case of the reference task, approximately 10 times faster. As mentioned before, this is because vectorial operations are supported by built-in functions, which are implemented by optimized machine code.

Tables 4.2 and 4.3 describe the execution times obtained for both alternatives on Stage 2.

(a) System

CPU	Intel Core2 CPU 6300 1.86GHz
Memory	4GB
OS	Linux x86 (32bit)
Language	Mathematica 8.0.4.0

(b) Task

Simulation	Brownian bridge trajectories
Number of trajectories	1000
Number of samples per trajectory	1000 (uniformly)

(c) Execution Times (in seconds)

Iterative	32.31
Vectorial	3.49

Table 4.1: Iterative and vectorial execution times comparison for the reference task.

Number of samples	Execution time (s)		Improvement (times faster)
	Iterative	Vectorial	
10	0.26	0.003	86.67
100	3.39	0.04	84.75
1000	32.42	3.93	8.25
10000	305.43	Out of Memory	–
100000	2989.15	Out of Memory	–

Table 4.2: 1000 trajectories execution time sensitivity to the number of samples.

Number of trajectories	Execution time (s)		Improvement (times faster)
	Iterative	Vectorial	
10	0.29	3.32	11.45 (slower)
100	3.40	3.36	1.01
1000	32.24	3.91	8.24
10000	300.28	8.48	35.41
100000	2981.13	Out of Memory	–

Table 4.3: 1000 samples per trajectory execution time sensitivity to the number of trajectories.

Table 4.2 shows, in a clear way, the main limitation of the vectorial simulation procedure, which is memory space. Simulation of all trajectories at a time, using Equation 4.17, requires memory space for the number-of-samples-by-number-of-samples square matrix $\mathbf{C}$ and for the number-of-samples-by-number-of-trajectories rectangular matrix $\mathbf{Z}$. As the number of trajectories and the number of samples per trajectory grow, the memory space becomes a severe limitation of the vectorial simulation procedure.

Table 4.3 shows that, for a small number of trajectories (up to 100) with a reasonable number of samples (1000), the vectorial simulation procedure is useless, due to the overhead execution time for computing matrix $\mathbf{C}$.

4.7 Extensions

It is clear by the Brownian bridge vectorial simulation construction that the simulation procedure can be naturally extended in the following 3 ways:

1. considering a condition different from $W(1) = 0$ (either in the time instant and its value);

2. considering more than one condition (sequences of bridges);

3. considering other Gaussian processes besides the Wiener process (considering mean and covariance functions different from the Wiener process ones).

The first two ways were already illustrated by Figure 4.2(d), where there were a total of four conditions, different from the Brownian bridge condition.

Regarding the third way, Figures 4.4, 4.5 and 4.6 illustrate the same example of Figure 4.2, but now for geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process. We chose these processes for their importance in modeling stock prices and interest rates. The simulation procedure is the same as the one in the example of Figure 4.2, except that the appropriate mean and covariance functions are used.

In the geometric Brownian motion case, the simulation was done for the underlying log normal process, which is a Gaussian process with mean and covariance functions given by:

$$m(t) = \left(\mu - \frac{\sigma^2}{2}\right) t \tag{4.18}$$

and

$$cov(s,t) = \sigma^2 \min(s,t) \tag{4.19}$$

where μ is the drift and σ the volatility. The geometric Brownian motion trajectories were obtained by taking the exponential of the log normal ones and multiplying by the process initial value x_0. The values used in the simulation examples of Figure 4.4 were: $x_0 = 0.5$; $\mu = 1.0$ and $\sigma = 1.0$.

In the fractional Brownian motion case the mean and covariance functions are given by:

$$m(t) = 0 \tag{4.20}$$

and

$$cov(s,t) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right) \tag{4.21}$$

where H is the Hurst index. The value used in the simulation examples of Figure 4.5 was: $H = 0.3$.

In the Ornstein-Ulenbeck mean reversion process case, the mean and covariance functions are given by:

$$m(t) = x_0 e^{-kt} + \theta(1 - e^{-kt}) \tag{4.22}$$

and

$$cov(s,t) = \frac{\sigma^2}{2k} e^{-k(s+t)} \left(e^{2k\min(s,t)} - 1\right) \tag{4.23}$$

where x_0 is the process initial value, k is the mean reversion velocity, θ is the mean reversion level and σ the volatility. The values used in the simulation examples of Figure 4.6 were: $x_0 = 0.5$; $k = 2.0$; $\theta = 0.1$ and $\sigma = 0.5$.

4.8 Illustration

In this section we illustrate the great flexibility of the path dependent vectorial simulation procedure by constructing a 2D Wiener process single path representation of a Norbert Wiener photo. The steps taken to construct the representation are the following:

1. Choose a white background photo.

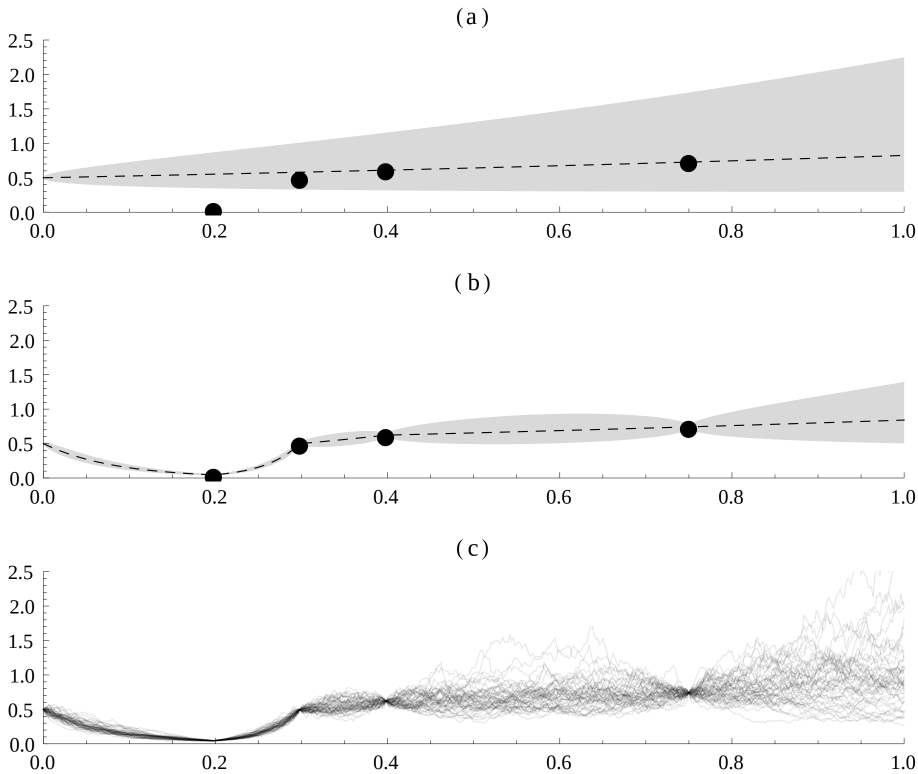

Figure 4.4: Gaussian processes for machine learning regression with geometric Brownian motion as prior: (a) prior process mean (dashed), prior process two standard deviations band (gray) and the training set (circle); (b) regression function (dashed) and two standard deviation regression confidence band (gray); (c) path dependent simulated trajectories (passing through the training set).

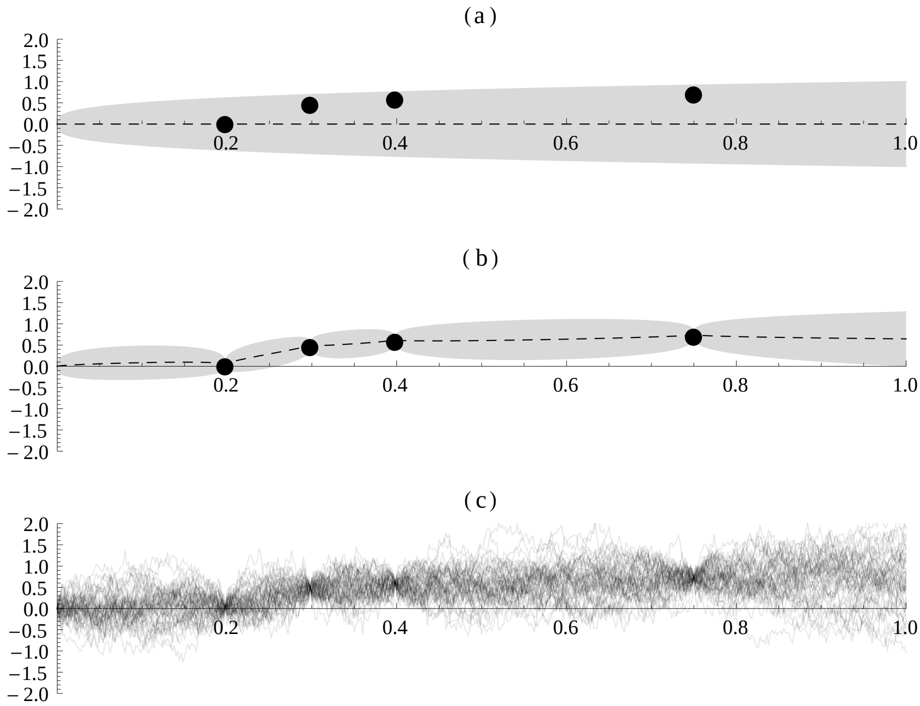

Figure 4.5: Gaussian processes for machine learning regression with fractional Brownian motion as prior: (a) prior process mean (dashed), prior process two standard deviations band (gray) and the training set (circles); (b) regression function (dashed) and two standard deviation regression confidence band (gray); (c) path dependent simulated trajectories (passing through the training set).

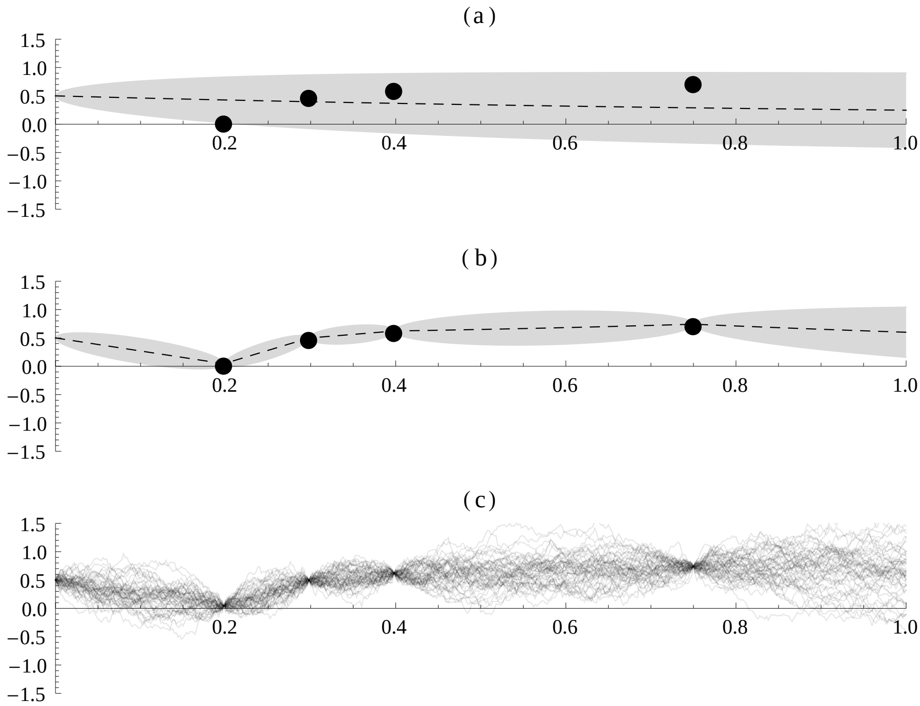

Figure 4.6: Gaussian processes for machine learning regression with the Ornstein-Ulenbeck mean reversion process as prior: (a) prior process mean (dashed), prior process two standard deviations band (gray) and the training set (circles); (b) regression function (dashed) and two standard deviation regression confidence band (gray); (c) path dependent simulated trajectories (passing through the training set).

2. Obtain a binarized with dithering version of the photo.

3. Obtain a possible sequence of nearest black pixels: starting at a random black pixel, find its nearest black pixel neighbor; repeat the procedure from the found neighbor, not considering the pixels already processed, until reaching the last unprocessed pixel.

4. Consider the black pixels coordinates, x and y, as the conditioning constraints.

5. For the sequence of the x coordinate constraints, simulate a Wiener process trajectory, by sampling uniformly 50 times each successive pair of constraints.

6. Repeat the previous step for the y coordinate (using increments independent from those used for the x coordinate).

7. Plot the y coordinate trajectory as a function of the x coordinate trajectory.

Figure 4.7 shows the resulting image.

4.9 Conclusions

The contribution of the present paper is twofold:

1. It presents a vectorial alternative to the iterative simulation of Brownian bridge trajectories, which is based on Gaussian processes for machine learning regression, and is relevant regarding the execution speed of interpreted programming languages implementations. The main limitation of the presented alternative is memory space;

2. It extends in a natural way the vectorial simulation of path dependent trajectories to other Gaussian processes, such as sequences of Brownian bridges, geometric Brownian motion, fractional Brownian motion and Ornstein-Ulenbeck mean reversion process.

Figure 4.7: 2D Wiener process single path representation of a Norbert Wiener photo.

5

Bonds Historical Simulation Value at Risk

5.1 Preamble

With the exception of this preamble and minor notation changes, this chapter contains the paper *Bonds Historical Simulation Value at Risk*, joint work with Prof. Manuel Esquível, Prof. Raquel Gaspar and Prof. P. Corte Real, which is submitted for publication in the Journal of Banking and Finance, by Elsevier. The current submission status is "under revision".

In several simulation situations spread across this thesis, in order to evaluate the existence of numerical problems, we have scaled zero coupon bond prices by using the implied yield at a certain time, to compute the bond price at another time, assuming the bond was held to maturity (mark to model).

This scaling procedure proved to be an important tool in the context of historical simulation value at risk (VaR) for portfolios with bonds.

In a joint work with Prof. Manuel Esquível and Prof. Pedro Corte Real, we have sold the authors wrights of an historical simulation value at risk implementation, for portfolios with bonds (among other securities), to a private bank, by 50.100,00 EUR. That implementation was based on this paper.

The main contributions of this paper are:

- Adjust bonds historical returns so that the adjusted returns can be used directly to compute VaR by historical simulation;
- Using real bond prices, to show that the developed method provides results consistent with the usual market observed trend, in which shorter times to maturity imply smaller yields, carrying smaller risk and consequently having smaller VaR;
- Using real bond prices, to show that the developed method strongly preserves the market implicit correlations between the instruments in the portfolio.

Abstract

Bonds historical returns can not be used directly to compute Value at Risk (VaR) by historical simulation because the maturities of the yields implied by the historical prices are not the relevant maturities at time VaR is computed.

In this paper we adjust bonds historical returns so that the adjusted returns can be used directly to compute VaR by historical simulation.

The adjustment is based on using implied historical yields to mark to model the bonds at the times to maturity relevant for the VaR computation.

We show that the obtained VaR values agree with the usual market trend of shorter times to maturity being traded with smaller yields, hence, carrying smaller risk and consequently having a smaller VaR.

5.2 Introduction

Despite all criticisms (Pritsker 2006), historical simulation is by far the most popular VaR method (Pérignon and Smith 2010).

It is well known that VaR computation, by historical simulation, of bond portfolios differs in important ways from VaR computation of stock portfolios (Darbha 2001). Essentially, this is because the maturities of the yields implied by bonds historical prices are not the relevant maturities, at time VaR is computed. They are greater than the relevant maturities because they correspond to historical past times when the time to maturity was greater than it is when VaR is computed. Since time to maturity is a critical factor of bonds risk, this moves away the possibility of using bonds historical returns, directly in VaR computation by historical simulation.

The popular method to overcome this issue of cash flow mapping in risk factors, besides ignoring the portfolio specific VaR, being subjective, complex (Alexander 2009) and using lots of information sources, ruins the objectivity and the simplicity of the historical simulation method.

In this paper we develop a method of adjusting bonds historical returns so that they can be used directly in VaR computations by historical simulation.

The method is based on computing the returns of the prices obtained by marking to model the bonds at the times to maturity relevant for the VaR computation. The historical prices implied yields are used as if the bonds were in a hold-to-maturity portfolio.

We show that the developed method provides results consistent with the usual market observed trend, in which shorter times to maturity imply smaller yields, carrying smaller risk and consequently having smaller VaR. We also show that the developed method strongly preserves the market implicit correlations between the instruments in the portfolio.

5.3 Time to maturity adjusted bond returns

Consider the VaR computation at day n_{VaR}, with time horizon N days, and confidence level α percent, of a portfolio with an alive zero coupon bond with maturity $T > n_{VaR} + N$ and principal P. See the time line in Exhibit 5.1 for a graphical representation of these instants. Clearly, the relevant maturities for this VaR computation are $T - n_{VaR}$ and $T - (n_{VaR} + N)$.

Following the general historical simulation[1] method (J. C. Hull 2008), the bond's N days market observed historical returns should be used to compute VaR. Denoting by $p(n)$, the historical price of the bond, at day $N < n < n_{VaR}$, and denoting by $HR(n, N)$ the N days historical return at day n, defined as in (J. C. Hull 2008), the N days possibly overlapping historical returns are given by:

$$\begin{aligned} HR(n,N) &= \frac{p(n) - p(n-N)}{p(n-N)} + 1 \\ &= \frac{p(n)}{p(n-N)}, \qquad\qquad n = N+1, \cdots, n_{VaR} - 1 \end{aligned} \tag{5.1}$$

[1] VaR historical simulation method is referred by some authors as non-parametric VaR.

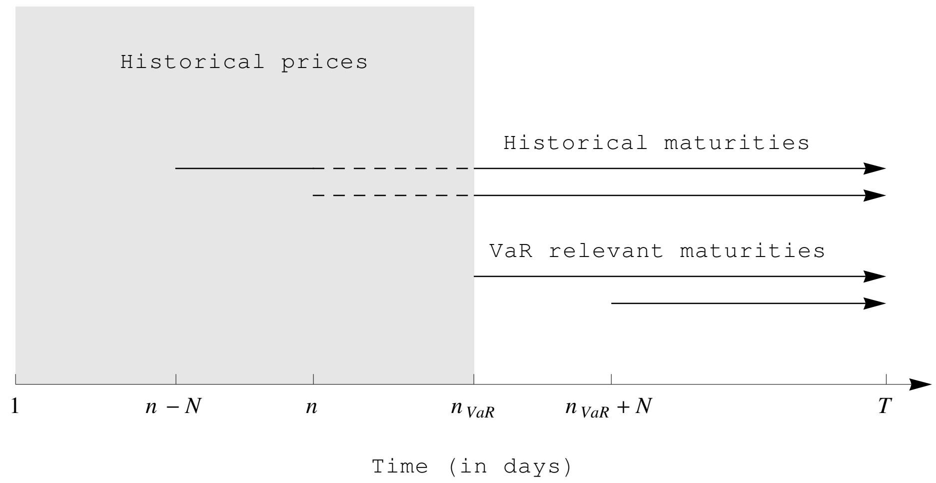

Figure 5.1: VaR computation time line. The gray zone represents the time interval where there are historical prices available. The dashed part of the historical maturities arrows means that those arrows can extend to all the gray zone.

These market observed historical returns should be applied to the bond market value at day n_{VaR} as follows:

$$p(n_{VaR})HR(n, N) = p(n_{VaR})\frac{p(n)}{p(n-N)}, \quad n = N+1, \cdots, n_{VaR}-1 \tag{5.2}$$

The resulting values define an empirical distribution of possible N days bond profits and losses, at time n_{VaR}. The VaR should be the potential loss of the $1-\alpha/100$ quantile of this empirical distribution.

But the historical price sequence $p(n)$ for $1 \leq n < n_{VaR}$, used in Equation (5.2), implies a sequence of daily compounded yields[2] $r(n)$, given by:

$$r(n) = \left(\frac{P}{p(n)}\right)^{\frac{1}{T-n}} - 1 \tag{5.3}$$

And the problem with this general approach is that the maturities of these implied historical yields are $T-n$, which are greater than the relevant maturities for the VaR computation (as can be observed in Exhibit 5.1). This is why the bond's historical returns can not be used directly in VaR computation.

Nevertheless, the implied historical yields, for times $1 \leq n < n_{VaR}$, provide bond valuation at future times n_{VaR} and $n_{VaR}+N$ by marking to model the bond

[2]Daily compounding is used because typical VaR time horizons are specified in days.

with constant daily yield, as if it was part of a held to maturity portfolio.

Denoting by $v(m,n)$ the future value at time $n_{VaR} \leq m < T$ of the bond, bought at time $1 \leq n < n_{VaR}$, at historical price $p(n)$, $v(m,n)$ is given by the valuation of the future cash flow at maturity time, with daily compounded yield $r(n)$, implied by price $p(n)$:

$$v(m,n) = \frac{P}{(1+r(n))^{T-m}} = \frac{P}{\left(\frac{P}{p(n)}\right)^{\frac{T-m}{T-n}}} \tag{5.4}$$

The possibility of a default event is assumed to be implicitly incorporated in the price $p(n)$ itself.

In this paper we adjust the historical returns of Equation 5.1 in the following way:

Step 1 for each historical price $p(n)$, in Equation 5.1, we compute the corresponding future value $v(n_{VaR}+N,n)$, at time $n_{VaR}+N$, matching the VaR relevant maturity $T-(n_{VaR}+N)$;

Step 2 for each historical price $p(n-N)$, in Equation 5.1, we compute the corresponding future value $v(n_{VaR}, n-N)$, at time n_{VaR}, matching the VaR relevant maturity $T-n_{VaR}$;

Step 3 we adjust each historical return, in Equation 5.1, for the VaR relevant maturities by replacing $p(n)$ by $v(n_{VaR}+N,n)$ and $p(n-N)$ by $v(n_{VaR}, n-N)$.

Denoting by $AHR(n,N,n_{Var})$ the N days adjusted historical return, at day n, adjusted for time n_{Var}, and using Equation 5.4, $AHR(n,N,n_{Var})$ is given by:

$$\begin{aligned} AHR(n,N,n_{Var}) &= \frac{v(n_{VaR}+N,n)}{v(n_{VaR},n-N)} \\ &= \frac{\left(\frac{P}{p(n-N)}\right)^{\frac{T-n_{VaR}}{T-(n-N)}}}{\left(\frac{P}{p(n)}\right)^{\frac{T-(n_{VaR}+N)}{T-n}}}, \quad n = N+1,\cdots,n_{VaR}-1 \end{aligned} \tag{5.5}$$

Note that each $AHR(n,N,n_{Var})$ value if fixed by historical market prices $p(n-N)$ and $p(n)$, thus capturing the market changes between times $n-N$ and n, while being adjusted to the VaR computation relevant maturities $T-n_{VaR}$ and $T-(n_{VaR}+N)$ thus avoiding the pull-to-par effect.

Our proposal is to use the adjusted historical returns $AHR(n, N, n_{Var})$ directly in the VaR computation. Therefore the VaR is given by the potential loss of the $1 - \alpha/100$ quantile of the following time to maturity adjusted empirical distribution:

$$p(n_{VaR}) \frac{\left(\frac{P}{p(n-N)}\right)^{\frac{T-n_{VaR}}{T-(n-N)}}}{\left(\frac{P}{p(n)}\right)^{\frac{T-(n_{VaR}+N)}{T-n}}}, \quad n = N+1, \cdots, n_{VaR}-1 \tag{5.6}$$

5.4 Extensions

In this section we discuss the usage of the proposed adjustment to other scenarios beside computing the VaR of portfolios with zero coupon bonds at time the historical sequence of prices ends.

5.4.1 Coupon bonds

The extension to portfolios with coupon bonds is straight forward. In order to compute the future value of a coupon bond at time m, based on the market price of the bond at time $n < m$, two differences from the zero coupon bond case arise:

1. the yield to maturity at time n is computed using the bond's dirty price and accounting for all future cash flows after time n;
2. the value of the bond at time m accounts for all future cash flows after time m.

Then, the adjusted historical returns are defined by Equation (5.5) as in the case of a zero coupon bond and the VaR is computed in the same way.

5.4.2 Adjusting for past times

Consider a bond B that has already expired. Consider also a new bond, B_1, from the same issuer, equal to bond B, i.e., with the same type, principal, maturity, number of coupons and coupon rate (if applicable), etc. The VaR of a portfolio containing bond B_1 is to be computed by historical simulation at day $n_{VaR} = 1$ of bond's B_1 life. Suppose that the only historical prices available from bond's B_1 issuer are those of bond B.

In this limit situation, the VaR computation relevant maturities are $T - 1$ and $T - (1 + N)$, which, as can be observed in Exhibit 5.2, are greater than the maturities of the historical implied yields of bond B (with exception of the return at time

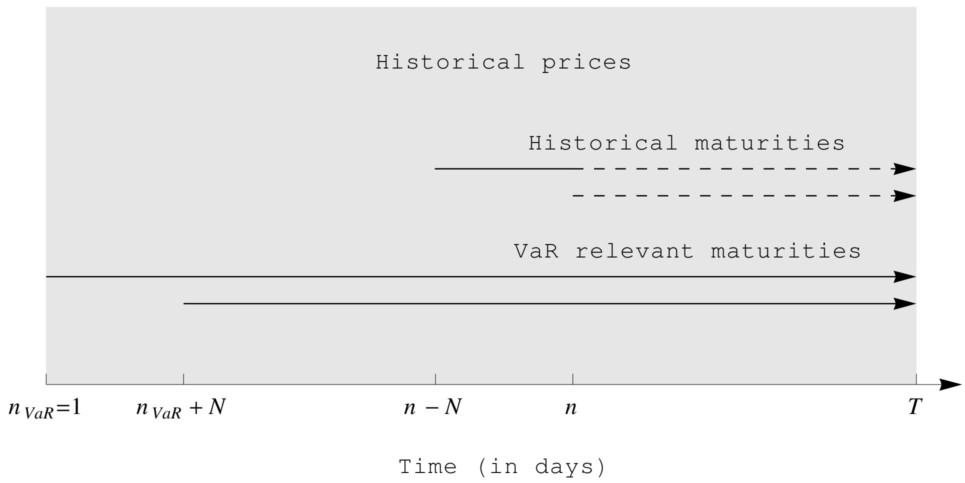

Figure 5.2: VaR computation time line for $n_{VaR} = 1$. The gray zone represents the time interval where there are historical prices available. The dashed part of the historical maturities arrows means that those arrows can extend to all the gray zone.

$N + 1$ where the maturities are equal). Therefore, the adjustment of the historical returns of bond B is now for past times, corresponding to greater maturities.

The adjustment method proposed in section 5.3 can still be applied in this situation. In the computation of the $v(n_{VaR} + N, n)$, $n_{VaR} + N$ is now smaller than n and in the computation of $v(n_{VaR}, n - N)$, n_{VaR} is also smaller than $n - N$. But the time to maturity adjustment works as in the case of the future times.

In the general situation of n_{VaR} lying across the life of bond B_1 the bond B historical returns corresponding to maturities greater than the VaR relevant maturities $T - n_{VaR}$ and $T - (n_{VaR} + N)$, will be adjusted for future times, while the ones corresponding to maturities smaller than the VaR relevant maturities will be adjusted for past times. This ensures the adjustment of all bond B historical returns for the VaR relevant maturities.

5.5 Application

In this section we illustrate the usage of the bond adjusted historical returns of Equation (5.5), by computing the VaR of the simplest possible portfolio, namely, a portfolio with a unique real zero coupon bond. We use a sequence of real historical prices of an alive zero coupon bond and compute the VaR at time the historical prices sequence ends.

The VaR parameters used are:

Time horizon $N = 30$ days[3];

Confidence level $\alpha = 99\%$.

We first present the used portfolio. Then we detail the adjustment of a single historical return for purposes of illustrating the adjustment process. Finally, we adjust all the available historical returns and compute the VaR.

Additionally we illustrate the adjustment for past values by computing the VaR at the time the historical prices sequence begins. As if the used bond was already expired and the VaR of a portfolio with a new bond, from the same issuer, equal to the expired one, was to be computed at the starting time of the new bond.

5.5.1 Portfolio

The portfolio used for VaR computation has a single instrument: the real alive zero coupon bond, B, with principal $P = 1000$, maturing at day $T = 731$. Exhibit 5.3 shows the available real historical prices at day VaR is computed, $n_{VaR} = 372$. The prices were obtained from a quote service that delivers market prices aggregated from different dealers responsible for trading (market makers) this particular bond.

The prices in Exhibit 5.3 imply the market observed yields presented in Exhibit 5.4. Recall from Exhibit 5.4 that each day corresponds to a different time to maturity.

Exhibit 5.4 clearly shows the usual trend observed in the market, in which shorter time to maturities are traded with smaller yields.

5.5.2 Adjustment of a single return

In order to illustrate the historical returns adjustment process, in this section, the computations of a single return are detailed. We picked a bond's B, $N = 30$ days, historical return approximately in the middle of the sequence of available historical prices. The historical return picked was the one at day $n = 190$, which, according to Equation 5.1, is determined by the historical prices $p(n) = p(190)$ and $p(n - N) = p(190 - 30) = p(160)$. Exhibit 5.5 shows the maturities of the corresponding implied yields as well as the maturities at time VaR is computed.

[3]We use a time horizon of 30 days instead of other typical values, such as 1 or 10 days, because the difference between computation steps 1 and 2 becomes much more clear graphically.

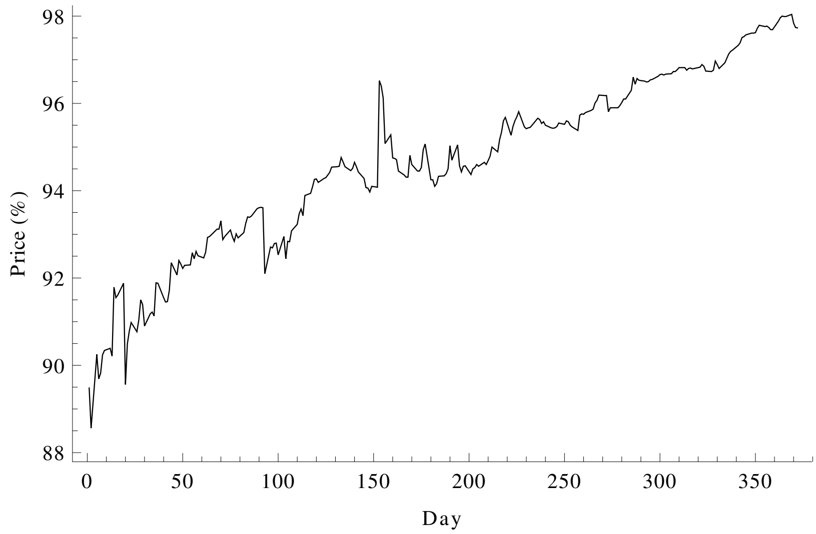

Figure 5.3: Real historical prices of a zero coupon bond with principal $P = 1000$ maturing at day $T = 731$. The prices are in percentage of the principal.

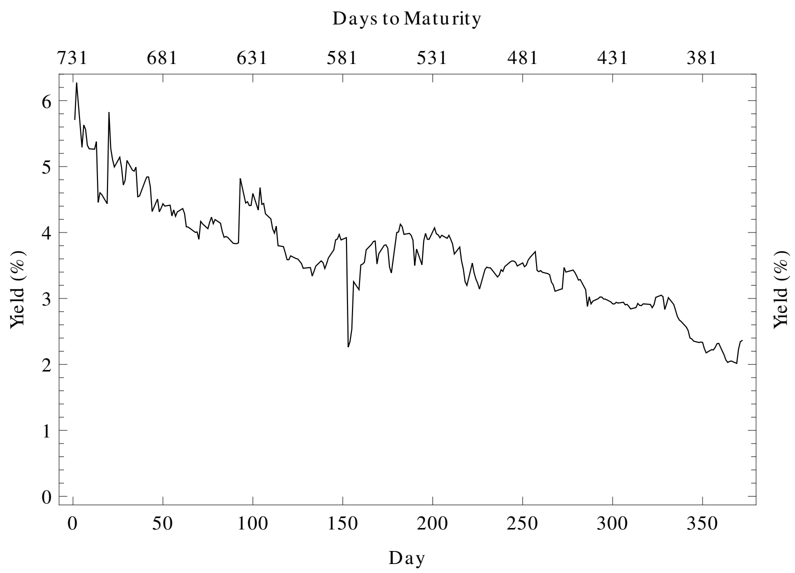

Figure 5.4: Daily compounded annualized implied yields from the historical prices of Exhibit 5.3, as a function of both time and time to maturity.

(a) **Historical Return Maturities**			
Day		Days to Maturity	
n	$= 190$	$T - n = 731 - 190$	$= \mathbf{541}$
$n - N = 190 - 30$	$= 160$	$T - (n - N) = 731 - 160$	$= \mathbf{571}$

(b) **VaR Relevant Maturities**			
Day		Days to Maturity	
n_{VaR}	$= 372$	$T - n_{VaR} = 731 - 372$	$= \mathbf{359}$
$n_{VaR} + N = 372 + 30$	$= 402$	$T - (n_{VaR} + N) = 731 - 402$	$= \mathbf{329}$

Figure 5.5: (a) $HR(n = 190, N = 30)$ historical return maturities. (b) VaR computed at time $n_{VaR} = 372$ relevant maturities.

Comparing Exhibit 5.5 (a) and (b), it is clear that the historical return underlying maturities are greater than the VaR relevant maturities.

The adjustment of the historical return at day $n = 190$, with Equation 5.5, matching the VaR relevant maturities, is detailed in Exhibit 5.6.

The prices that determine this historical return are highlighted in Exhibit 5.7 with the circles. The corresponding future values, used to compute the adjusted return, are highlighted with the squares.

As it can be observed from Exhibit 5.6 the adjusted return is closer to one than the historical return. This is in accordance with the trend observed in Exhibit 5.4. Once the yields of shorter times to maturity tend to be smaller, the returns at time to maturity $T - (n_{VaR} + N) = 372 + 30 = 329$ should be closer to one (smaller) than those at time to maturity 541.

5.5.3 Portfolio VaR

The $N = 30$ time horizon, $\alpha = 99\%$ confidence level, VaR, of the described portfolio, is computed at day $n_{VaR} = 372$ using the empirical distribution of the $N = 30$ adjusted returns of Equation (5.5). In order to obtain this distribution the adjustment of the single return detailed in the previous section is repeated for all available historical returns, following the steps described in section 5.3.

Step 1 Exhibit 5.8 shows the future value $v(n_{VaR}+N, n)$, at time $n_{VaR}+N = 402$, given each historical price $p(n)$ for $N = 30 < n < n_{VaR}(= 372)$, matching the VaR relevant maturity $T - (n_{VaR} + N) = 329$;

Step 2 Exhibit 5.9 shows the future value $v(n_{VaR}, n - N)$, at time $n_{VaR} = 372$, given each historical price $p(n)$ for $1 < n < n_{VaR} - N(= 342)$, matching the VaR relevant maturity $T - n_{VaR} = 359$;

(a) **Historical Return**				
Day	Price	Day-Time Horizon	Price	Return
n	$p(n)$	$n-N$	$p(n-N)$	$HR(n,N)=\frac{p(n)}{p(n-N)}$
190	95.03	160	94.75	**1.00296**

(b) **Step 1**				
Day	Price	Yield (%)	Future Day	Future Value
n	$p(n)$	$r(n)$	$n_{VaR}+N$	$v(n_{VaR}+N,n)$
190	95.03	3.499	402	96.947

(c) **Step 2**				
Day	Price	Yield (%)	Future Day	Future Value
$n-N$	$p(n-N)$	$r(n-N)$	n_{VaR}	$v(n_{VaR},n-N)$
160	94.75	3.507	372	96.666

(d) **Step 3**		
$v(n_{VaR}+N,n)$	$v(n_{VaR},n-N)$	$AHR(n,N,n_{Var})=\frac{v(n_{VaR}+N,n)}{v(n_{VaR},n-N)}$
96.947	96.666	**1.00291**

Figure 5.6: (a) $N = 30$ days market observed historical return at day $n = 190$. (b) Day $n = 190$ implied yield and future value at time $n_{VaR} + N = 402$. (c) Day $n = 160$ implied yield and future value at time $n_{VaR} = 372$. (d) The adjusted historical return for time $n_{VaR} = 372$.

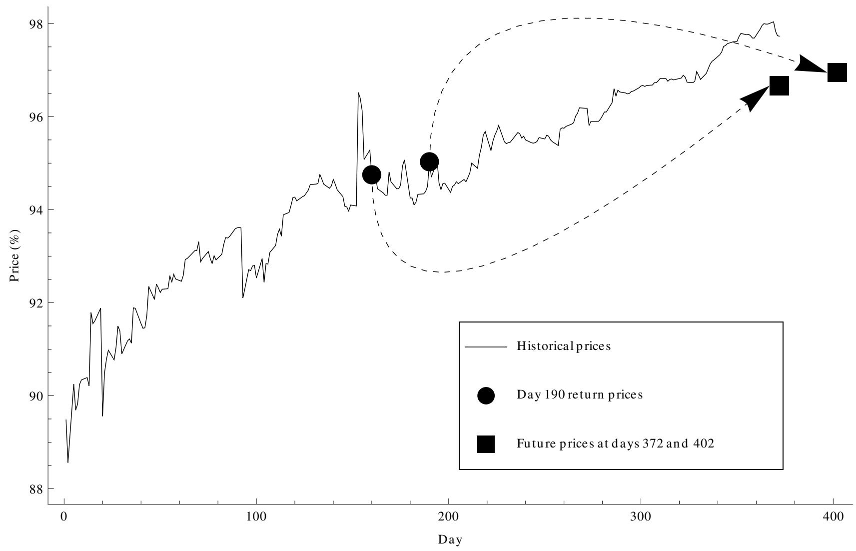

Figure 5.7: The prices that determine the $N = 30$ days historical return at time $n = 190$, the corresponding future prices at times $n_{VaR} = 372$ and $n_{VaR}+N = 402$, along with the historical prices sequence. The arrows represent future values.

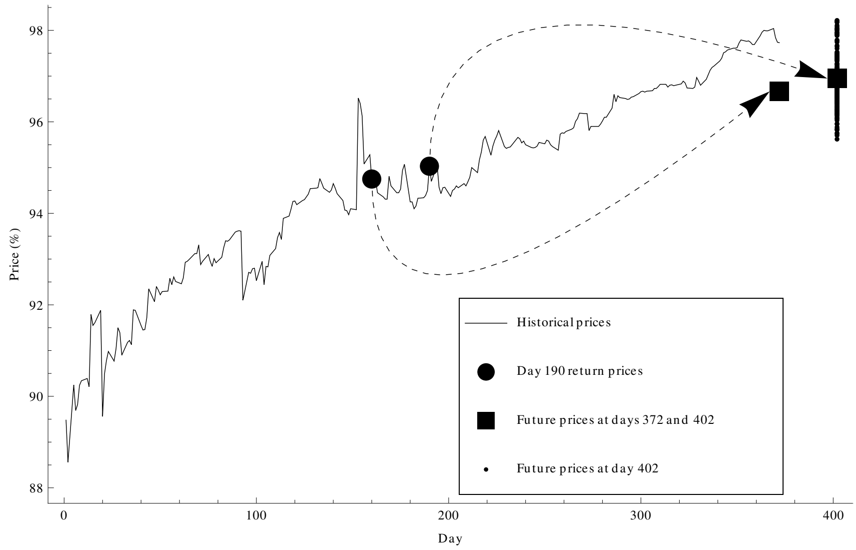

Figure 5.8: **Step 1** – Future values $v(n_{VaR} + N, n)$ at time $n_{VaR} + N = 402$, along with the historical prices sequence.

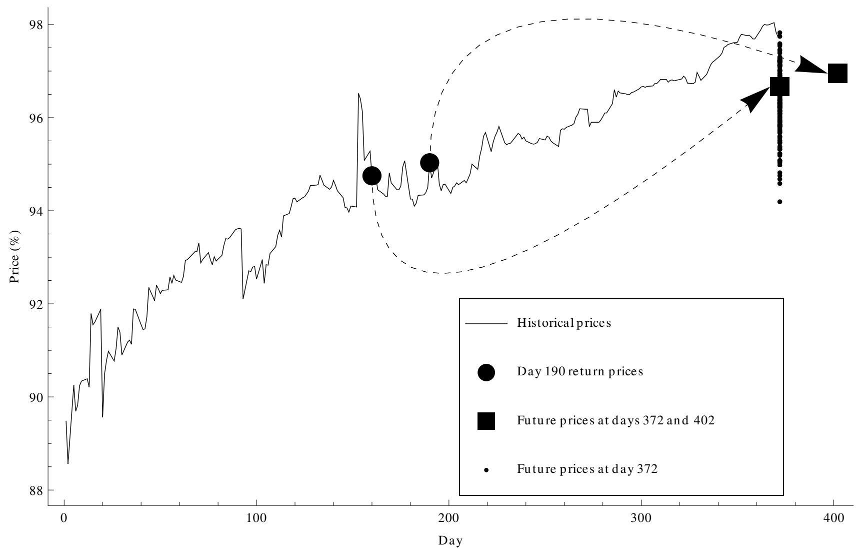

Figure 5.9: **Step 2** – Future values $v(n_{VaR}, n - N)$ at time $n_{VaR} = 372$, along with historical prices sequence.

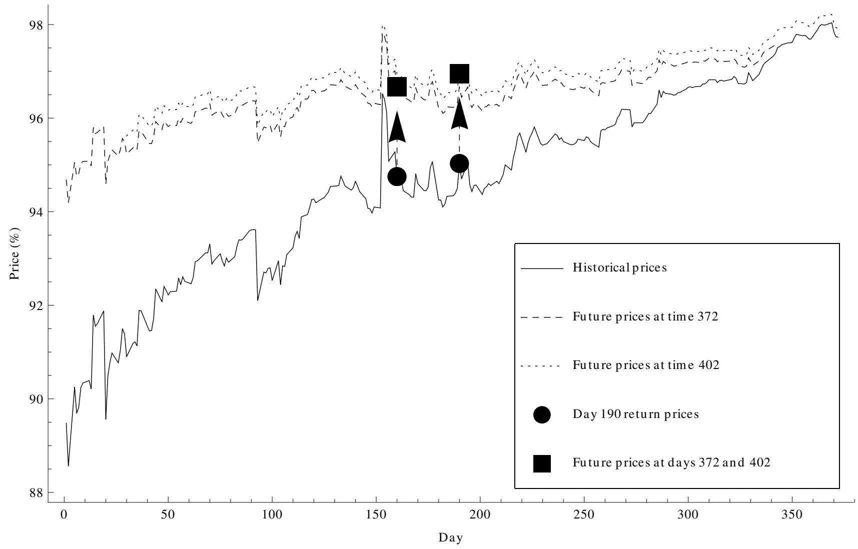

Figure 5.10: Future values $v(m,n)$, at times $m = n_{VaR} + N = 402$ (Step 1) and $m = n_{VaR} = 372$ (Step 2), along with the historical prices sequence. The future values are plotted as a function of the time n, of the historical price $p(n)$, that fixed the future value.

For purposes of comparison, Exhibit 5.10 shows the real, market observed historical prices, and also, the corresponding future values, $v(m,n)$ of Equation (5.4), at days $m = 402$ (Step 1) and $m = 372$ (Step 2). In this figure, the future values are plotted as a function of the day n of the historical price $p(n)$ which fixes the future value $v(m,n)$. The values highlighted in Exhibit 5.7 with the circles and squares are highlighted again in Exhibit 5.10, but now plotted as a function of n, too.

Step 3 Exhibit 5.11 shows the sequence of the adjusted historical returns for $n_{VaR} = 372$, computed from the future values of Steps 1 and 2, and the historical returns sequence for purposes of comparison.

Exhibit 5.12 shows the histograms of the adjusted and historical returns of Exhibit 5.11.

Finally, Exhibit 5.13 shows the VaR value computed from the empirical distribution of the overlapping adjusted returns of Equation (5.6), along with the possible loss corresponding to $1 - \alpha/100$ quantile of the overlapping historical returns empirical distribution of Equation (5.2), for comparison purposes. It also shows the correlation coefficient between the original and the adjusted returns.

Exhibit 5.11 shows that the adjusted returns are closer to one (are smaller) than the historical ones. This can be observed again in Exhibit 5.12 where the adjusted returns histogram is more concentrated towards one than the historical returns

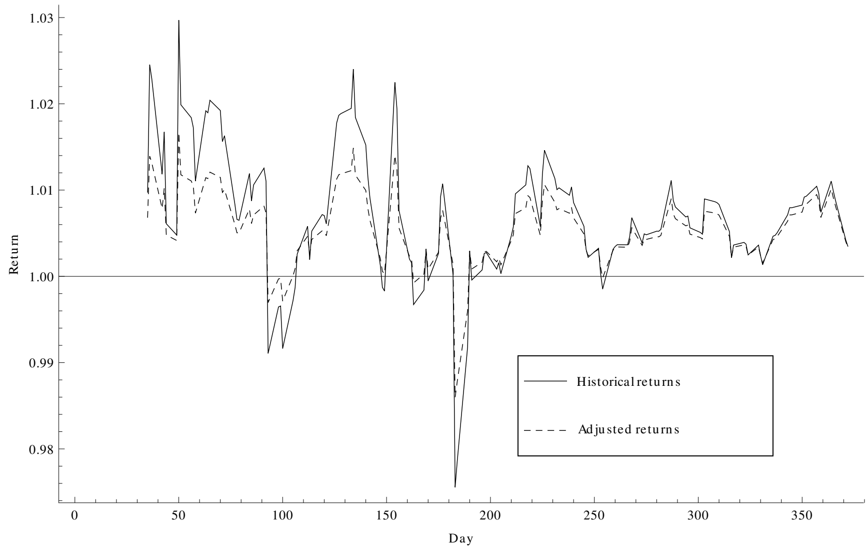

Figure 5.11: **Step 3** – Sequence of adjusted historical returns for $n_{VaR} = 372$ along with the corresponding historical returns sequence.

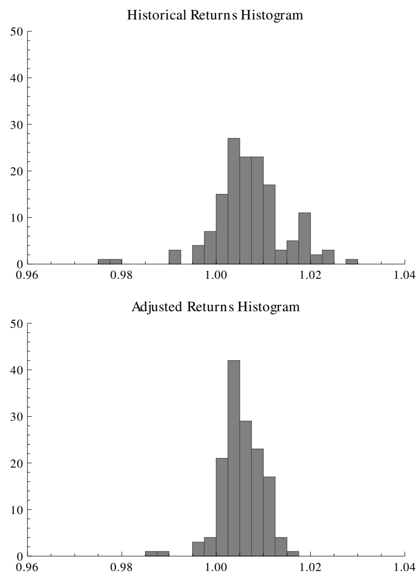

Figure 5.12: Adjusted returns for $n_{VaR} = 372$ and the historical returns histograms.

Time horizon	Confidence level	VaR	Possible Loss	Correlation
$N = 30$	$\alpha = 99\%$	$-\mathbf{1.222}$ %	$-\mathbf{2.147}$ %	0.988

Figure 5.13: Time horizon $N = 30$, confidence level $\alpha = 99\%$, bond B VaR, computed at day $n_{VaR} = 372$ by historical simulation using adjusted historical returns.

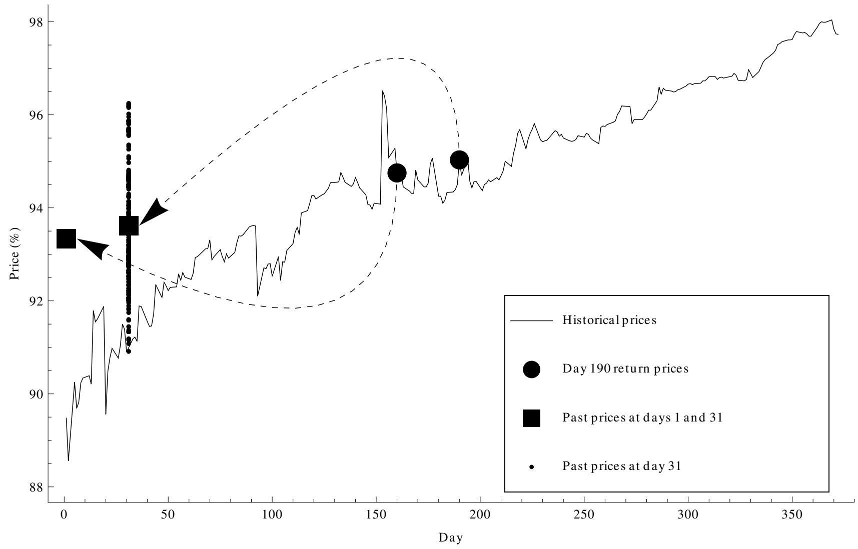

Figure 5.14: **Step 1** – Past values $v(n_{VaR}+N, n)$ at time $n_{VaR}+N = 31$, along with the historical prices sequence.

histogram. This results in a VaR value smaller than the possible loss corresponding to the $1-\alpha/100$ quantile of the historical returns empirical distribution (see Exhibit 5.13). Again, this result conforms with Exhibit 5.4 which shows a clear decreasing trend in implied yields as time to maturity decreases.

5.5.4 Adjusting for past times

In this section we repeat the VaR computation of the previous section, for $n_{VaR} = 1$, as if the bond B was already expired, a new bond B_1 equal to bond B was issued by the same issuer, the VaR of a portfolio with bond B_1 was to be computed by historical simulation and the only historical prices available from the bonds issuer were those of bond B, presented in Exhibit 5.3.

Following section 5.4.2 the past values, $v(m, n)$, of Equation (5.4), with $m = 1 \leq n$, are used to compute the adjusted historical returns of Equation (5.5) and the VaR is computed from the resulting empirical distribution.

Exhibits 5.14 to 5.18 illustrate the VaR computations and Exhibit 5.19 presents the results.

It can be observed from Exhibit 5.18 that the adjusted returns are now less concentrated towards one than the historical returns. This results in a VaR value, showed in Exhibit 5.19, which is now greater than the possible loss corresponding to the $1-\alpha/100$ quantile of the historical returns empirical distribution. Again, this is in accordance with Exhibit 5.4 and the fact that VaR relevant maturities

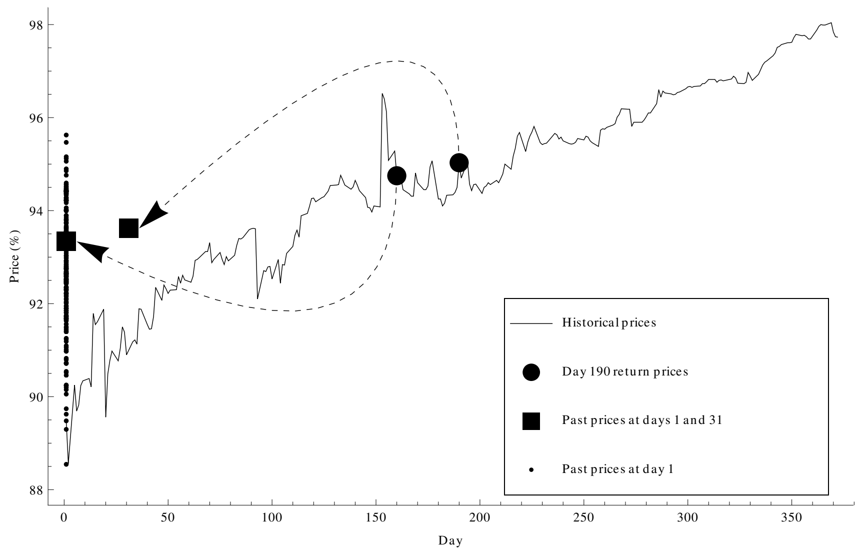

Figure 5.15: **Step 2** – Past values $v(n_{VaR}, n - N)$ at time $n_{VaR} = 1$, along with historical prices sequence.

Figure 5.16: Past values $v(m, n)$, at times $m = n_{VaR} + N = 31$ (Step 1) and $m = n_{VaR} = 1$ (Step 2), along with the historical prices sequence. The past values are plotted as a function of the time n, of the historical price $p(n)$, that fixed the past value.

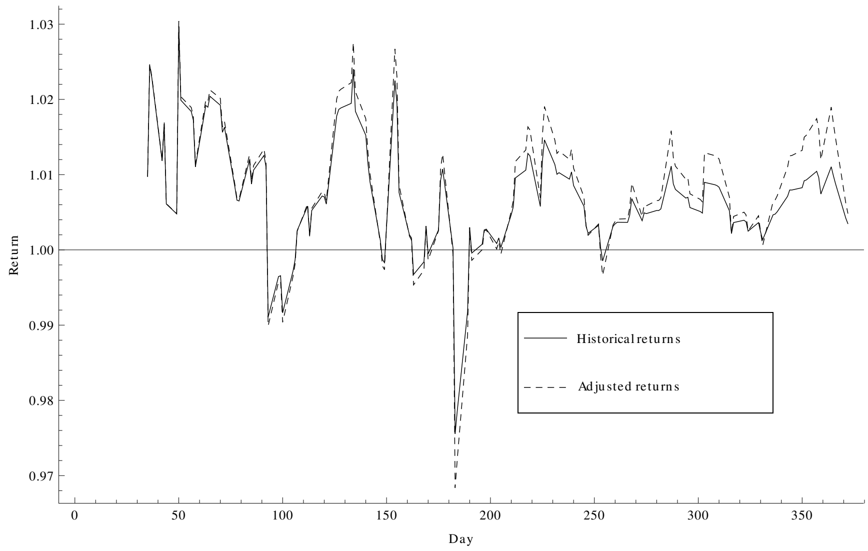

Figure 5.17: **Step 3** – Sequence of adjusted historical returns for $n_{VaR} = 1$ along with the corresponding historical returns sequence.

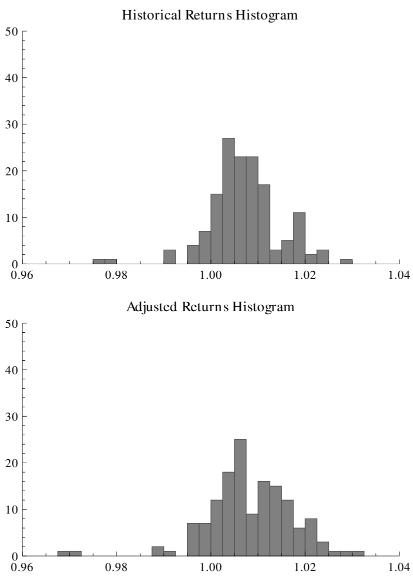

Figure 5.18: Adjusted returns for $n_{VaR} = 1$ and the historical returns histograms.

Time horizon	Confidence level	VaR	Possible Loss	Correlation
$N = 30$	$\alpha = 99\%$	$-\mathbf{2.791}$ %	$-\mathbf{2.147}$ %	0.980

Figure 5.19: Time horizon $N = 30$, confidence level $\alpha = 99\%$, bond B VaR, computed at day $n_{VaR} = 1$ by historical simulation using adjusted historical returns.

are now greater then the times to maturity underlying the available historical returns[4].

5.6 Conclusions

Bond historical returns can not be used directly to compute VaR by historical simulation because the maturities of the yields implied by the historical prices are not the relevant maturities at time VaR is computed.

In this paper we adjust bonds historical returns to the VaR relevant maturities so that the adjusted returns can be used directly to compute VaR by historical simulation while preserving the market conditions underlying the historical prices sequences. The adjustment is based on using implied historical yields to mark to model the bonds at the times to maturity relevant for the VaR computation.

The proposed method has the following features:

- Time to maturity adjusted bond returns are used directly in the VaR historical simulation computation.
- VaR of portfolios with bonds can be computed by historical simulation keeping the simplicity of the historical simulation method.
- Portfolio's specific VaR are obtained.
- VaR values obtained are consistent with the usual market trend of shorter times to maturity being traded with smaller yields, therefore carrying smaller risk and having a smaller VaR.
- The only source of information used is the market, through the bonds historical prices.
- The correlation between each bond returns and the returns of the other instruments in the portfolio is strongly preserved.
- The VaR for the desired time horizon is computed directly with no VaR time scaling approximations.

We left for future work the research of the non-linear mathematical properties of the developed method, and also back-testing the method with benchmark portfolios.

[4]With the exception of the single historical return at time $n = 31$, where the maturities are equal.

Part II

Unsubmitted Papers

6

Machine Learning Gaussian Short Rate

6.1 Preamble

Using a single Vasicek short rate factor, under the risk neutral measure, the machine learning Gaussian short rate model can't solve the term structure fitting issue mentioned in Section 1.2. In this paper a sum of Vasicek short rate factors is proposed in order to solve that problem.

The main contributions of this paper are:

- Propose a sum of Vasicek short rate factors, under the risk neutral measure, as a prior to Gaussian processes for machine learning regression;
- Obtain the zero coupon bond mean and covariance functions of the prior.

Abstract

In this paper we model the short rate, under the risk neutral measure, as a Gaussian process conditioned by the logarithm of market observed zero coupon bonds prices. The model is based on Gaussian processes for machine learning, using N addictive Vasicek factors as prior.

The model automatically fits all observed zero coupon bond log prices, in particular those which define the current term structure of interest rates.

The number of factors needed is equal to the maximum number of zero coupon bonds maturities, observed in a single time instant.

All model parameters are learned directly under the risk neutral measure, using zero coupon bonds log prices, exclusively.

6.2 Introduction

The one factor machine learning Gaussian short rate model presented in (Sousa, Esquível, and Gaspar 2013) has the limitation of supporting a single T-maturity zero coupon bond log price in each time instant. Consequently it can not fit the current term structures of interest rates, observed in the market.

In this paper we extend the one factor machine learning Gaussian short rate model, considering a sum of N Vasicek factors as prior. The extended model supports a maximum of N distinct zero coupon bond prices in each time instant. Therefore it can fit the current term structures of interest rates observed in the market.

All the model parameters are obtained directly under the risk neutral model through maximization of the likelihood of market observed zero coupon bond log prices given the parameters. Besides zero coupon bond prices, no other sources of information are needed.

6.3 Short rate prior

The short rate prior, $r(t)$, is a sum of N Vasicek factors, under the arbitrage free risk neutral measure:

$$r(t) = \sum_{j=1}^{N} r_j(t) \tag{6.1}$$

Each Vasicek factor $r_j(t)$ follows an Ornstein-Uhlenbeck mean-reverting process, under the risk neutral measure, defined by the stochastic differential equation (SDE):

$$\mathrm{d}r_j(t) = k_j(\theta_j - r_j(t))\mathrm{d}t + \sigma_j \mathrm{d}W_j(t). \tag{6.2}$$

For each factor j, parameter k_j is the mean reversion velocity, θ_j is the mean interest rate level, σ_j is the volatility and $W_j(t)$ the Wiener process. Parameters k_j and σ_j are positive.

The factors Wiener processes $W_j(t)$ are correlated with correlation coefficient ρ_{ij}:

$$\rho_{ij}\mathrm{d}t = \mathrm{d}W_i(t)\mathrm{d}W_j(t). \tag{6.3}$$

Parameters ρ_{ij} are $-1 < \rho_{ij} < 1$.

Since, for $0 < s < t$, the SDE solution of Equation 6.2 is:

$$r_j(t) = r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)}) + \sigma_j e^{-k_j t}\int_s^t e^{k_j u}\mathrm{d}W_j(u) \tag{6.4}$$

the short rate is given by:

$$r(t) = \sum_{j=1}^{N}\left(r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)}) + \sigma_j e^{-k_j t}\int_s^t e^{k_j u}\mathrm{d}W_j(u)\right) \tag{6.5}$$

6.3.1 Short rate prior mean

Each factor mean, $m_{r_j}(t)$, is given by:

$$\begin{aligned} m_{r_j}(t) \quad = \quad & E[r_j(t)] = E\left[r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)}) + \sigma_j e^{-k_j t}\int_s^t e^{k_j u}\mathrm{d}W_j(u)\right] \\ & r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)}) + \sigma_j e^{-k_j t}E\left[\int_s^t e^{k_j u}\mathrm{d}W_j(u)\right] \\ & = r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)}). \end{aligned} \tag{6.6}$$

Therefore the short rate mean, $m_r(t)$, is given by

$$\begin{aligned} m_r(t) &= E\left[r(t)\right] = E\left[\sum_{j=1}^{N} r_j(t)\right] = \sum_{j=1}^{N} E\left[r_j(t)\right] \\ &= \sum_{j=1}^{N} m_{r_j}(t) \\ &= \sum_{j=1}^{N}\left(r_j(s)e^{-k_j(t-s)} + \theta_j(1 - e^{-k_j(t-s)})\right) \end{aligned} \tag{6.7}$$

6.3.2 Short rate prior covariance

Each factor covariance, $\mathrm{cov}_{r_j}(t_1,t_2)$, is, by definition:

$$\mathrm{cov}_{r_j}(t_1,t_2) = E[(r_j(t_1) - m_{r_j}(t_1))(r_j(t_2) - m_{r_j}(t_2))].$$

The term $(r_j(t) - m_{r_j}(t))$ is given by

$$r_j(t) - m_{r_j}(t) = \sigma_j e^{-k_j t}\int_s^t e^{k_j u}\mathrm{d}W_j(u).$$

Therefore each factor covariance is given by

$$\begin{aligned}
\mathrm{cov}_{r_j}(t_1,t_2) &= E\left[\left(\sigma_j e^{-k_j t_1}\int_s^{t_1} e^{k_j u}\mathrm{d}W_j(u)\right)\left(\sigma_j e^{-k_j t_2}\int_s^{t_2} e^{k_j u}\mathrm{d}W_j(u)\right)\right] \\
&= \sigma_j^2 e^{-k_j(t_1+t_2)} E\left[\left(\int_s^{t_1} e^{k_j u}\mathrm{d}W_j(u)\right)\left(\int_s^{t_2} e^{k_j u}\mathrm{d}W_j(u)\right)\right] \\
&= \sigma_j^2 e^{-k_j(t_1+t_2)} E\left[\left(\int_s^{\min(t_1,t_2)} e^{k_j u}\mathrm{d}W_j(u)\right)^2\right] \\
&= \sigma_j^2 e^{-k_j(t_1+t_2)} \int_s^{\min(t_1,t_2)} E\left[\left(e^{k_j u}\right)^2\right]\mathrm{d}u \\
&= \sigma_j^2 e^{-k_j(t_1+t_2)} \int_s^{\min(t_1,t_2)} e^{2k_j u}\mathrm{d}u \\
&= \sigma_j^2 e^{-k_j(t_1+t_2)} \frac{1}{2k_j}\left(e^{2k_j \min(t_1,t_2)} - e^{2k_j s}\right) \qquad (6.8)
\end{aligned}$$

The short rate covariance, $\mathrm{cov}_r(t_1,t_2)$, is, by definition:

$$\mathrm{cov}_r(t_1,t_2) = E[(r(t_1) - m_r(t_1))(r(t_2) - m_r(t_2))].$$

The term $(r(t) - m_r(t))$ is given by

$$r(t) - m_r(t) = \sum_{j=1}^{N}\left(\sigma_j e^{-k_j t}\int_s^t e^{k_j u}\mathrm{d}W_j(u)\right).$$

The short rate covariance is given by

$$\begin{aligned}
&\mathrm{cov}_r(t_1,t_2)\\
&= E\left[\left(\sum_{j=1}^{N}\left(\sigma_j e^{-k_j t_1}\int_s^{t_1} e^{k_j u}\mathrm{d}W_j(u)\right)\right)\left(\sum_{j=1}^{N}\left(\sigma_j e^{-k_j t_2}\int_s^{t_2} e^{k_j u}\mathrm{d}W_j(u)\right)\right)\right]\\
&= \sum_{j=1}^{N}\mathrm{cov}_{r_j}(t_1,t_2)+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\mathrm{cov}_{r_i,r_j}(t_1,t_2)+\mathrm{cov}_{r_j,r_i}(t_1,t_2)\right) \qquad (6.9)
\end{aligned}$$

where $\mathrm{cov}_{r_i,r_j}(t_1,t_2)$ is the covariance between factors r_i and r_j.
The covariance $\mathrm{cov}_{r_i,r_j}(t_1,t_2)$ is given by:

$$\begin{aligned}
&\mathrm{cov}_{r_i,r_j}(t_1,t_2)\\
&= E\left[\left(\sigma_i e^{-k_i t_1}\int_s^{t_1} e^{k_i u}\mathrm{d}W_i(u)\right)\left(\sigma_j e^{-k_j t_2}\int_s^{t_2} e^{k_j u}\mathrm{d}W_j(u)\right)\right]\\
&= \sigma_i\sigma_j e^{-k_i t_1-k_j t_2}E\left[\left(\int_s^{\min(t_1,t_2)} e^{k_i u}\mathrm{d}W_i(u)\right)\left(\int_s^{\min(t_1,t_2)} e^{k_j u}\mathrm{d}W_j(u)\right)\right]\\
&= \sigma_i\sigma_j e^{-k_i t_1-k_j t_2}\int_s^{\min(t_1,t_2)} e^{(k_i+k_j)u}\rho_{ij}\mathrm{d}u\\
&= \rho_{ij}\frac{\sigma_i\sigma_j e^{-k_i t_1-k_j t_2}}{k_i+k_j}\left(e^{(k_i+k_j)\min(t_1,t_2)}-e^{(k_i+k_j)s}\right) \qquad (6.10)
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathrm{cov}_r(t_1,t_2)\\
&= \sum_{j=1}^{N}\sigma_j^2 e^{-k_j(t_1+t_2)}\frac{1}{2k_j}\left(e^{2k_j\min(t_1,t_2)}-e^{2k_j s}\right)\\
&\quad+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\rho_{ij}\frac{\sigma_i\sigma_j e^{-k_i t_1-k_j t_2}}{k_i+k_j}\left(e^{(k_i+k_j)\min(t_1,t_2)}-e^{(k_i+k_j)s}\right)\right.\\
&\quad\left.+\,\rho_{ji}\frac{\sigma_j\sigma_i e^{-k_j t_1-k_i t_2}}{k_j+k_i}\left(e^{(k_j+k_i)\min(t_1,t_2)}-e^{(k_j+k_i)s}\right)\right)\\
&= \sum_{j=1}^{N}\sigma_j^2 e^{-k_j(t_1+t_2)}\frac{1}{2k_j}\left(e^{2k_j\min(t_1,t_2)}-e^{2k_j s}\right)\\
&\quad+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\rho_{ij}\frac{\sigma_i\sigma_j}{k_i+k_j}\left(e^{(k_i+k_j)\min(t_1,t_2)}-e^{(k_i+k_j)s}\right)\left(e^{-k_i t_1-k_j t_2}+e^{-k_j t_1-k_i t_2}\right)
\end{aligned} \qquad (6.11)$$

6.4 Zero coupon bond prices prior

Under the risk neutral measure, the price, at time t, of a zero coupon bond that pays 1 at maturity T, is given by

$$p(t,T) = E\left[e^{-\int_t^T r(u)\mathrm{d}u}\right], \tag{6.12}$$

where the expectation E is to be taken under the risk neutral measure Q.

Since $r(t)$ is a Gaussian process and that $\int_t^T r(u)\mathrm{d}u$ is also a Gaussian process, the bond prices are the expected value of the exponential of a Gaussian random variable.

It is known that if

$$X \sim \mathcal{N}(m, v^2)$$

then

$$E\left[e^X\right] = e^{m+\frac{1}{2}v^2}$$

Therefore, in order to get the bond prices we need to compute the mean, $m_x(t)$, and variance, $\mathrm{var}_x(t)$ of

$$x(t) = -\int_t^T r(u)\mathrm{d}u$$

6.4.1 $x(t)$ mean

The mean, $m_x(t)$, is given by

$$\begin{aligned} m_x(t) &= E[x(t)] = E\left[-\int_t^T r(u)\mathrm{d}u\right] = E\left[-\int_t^T \sum_{j=1}^N r_j(u)\mathrm{d}u\right] \\ &= -\sum_{j=1}^N \int_t^T E\left[r_j(u)\right]\mathrm{d}u \\ &= \sum_{j=1}^N -\int_t^T m_{r_j}(u)\mathrm{d}u \end{aligned}$$

Denoting by $m_{x_j}(t)$ the integral of factor j mean

$$\begin{aligned}
m_{x_j}(t) &= -\int_t^T m_{r_j}(u)\mathrm{d}u \\
&= -\int_t^T \left(r_j(t)e^{-k_j(u-t)} + \theta_j(1-e^{-k_j(u-t)})\right)\mathrm{d}u \\
&= -r_j(t)\frac{1}{k_j}\left(1-e^{-k_j(T-t)}\right) - \theta_j(T-t) + \theta_j\frac{1}{k_j}\left(1-e^{-k_j(T-t)}\right)
\end{aligned}$$

Finally

$$\begin{aligned}
m_x(t) &= \sum_{j=1}^{N}\Bigg(-r_j(t)\frac{1}{k_j}\left(1-e^{-k_j(T-t)}\right) \\
&\quad -\theta_j(T-t) + \theta_j\frac{1}{k_j}\left(1-e^{-k_j(T-t)}\right)\Bigg)
\end{aligned} \tag{6.13}$$

6.4.2 $x(t)$ variance

The variance $\mathrm{var}_x(t)$ is, by definition

$$\mathrm{var}_x(t) = E\left[(x(t)-m_x(t))^2\right]$$

The term $(x(t)-m_x(t))$ is given by

$$\begin{aligned}
&x(t)-m_x(t) \\
&\quad= -\int_t^T r(u)\mathrm{d}u - E\left[-\int_t^T r(u)\mathrm{d}u\right] \\
&\quad= -\int_t^T r(u)\mathrm{d}u - \int_t^T -E\left[r(u)\right]\mathrm{d}u \\
&\quad= -\int_t^T r(u) - m_r(u)\mathrm{d}u \\
&\quad= -\int_t^T \left(\sum_{j=1}^{N}\sigma_j e^{-k_j u}\int_t^u e^{k_j s}\mathrm{d}W_j(s)\right)\mathrm{d}u
\end{aligned}$$

The variance of $x(t)$ is given by

$$
\begin{aligned}
\mathrm{var}_x(t) &= E\left[(x(t)-m_x(t))^2\right] \\
&= E\left[\left(-\int_t^T \left(\sum_{j=1}^N \sigma_j e^{-k_j u}\int_t^u e^{k_j s}\mathrm{d}W_j(s)\right)\mathrm{d}u\right)^2\right] \\
&= E\left[\left(-\int_t^T \left(\sum_{j=1}^N \sigma_j e^{-k_j u}\int_t^u e^{k_j s}\mathrm{d}W_j(s)\right)\mathrm{d}u\right)\right. \\
&\quad \left.\left(-\int_t^T \left(\sum_{j=1}^N \sigma_j e^{-k_j v}\int_t^v e^{k_j s}\mathrm{d}W_j(s)\right)\mathrm{d}v\right)\right] \\
&= \int_t^T\int_t^T E\left[\left(\sum_{j=1}^N \sigma_j e^{-k_j u}\int_t^u e^{k_j s}\mathrm{d}W_j(s)\right)\right. \\
&\quad \left.\left(\sum_{j=1}^N \sigma_j e^{-k_j v}\int_t^v e^{k_j s}\mathrm{d}W_j(s)\right)\right]\mathrm{d}u\mathrm{d}v \\
&= \int_t^T\int_t^T \mathrm{cov}_r(u,v)\mathrm{d}u\mathrm{d}v \\
&= \int_t^T\int_t^T \left(\sum_{j=1}^N \mathrm{cov}_{r_j}(u,v) + \sum_{i=1}^N\sum_{j=i+1}^N \left(\mathrm{cov}_{r_i,r_j}(u,v)+\mathrm{cov}_{r_j,r_i}(u,v)\right)\right)\mathrm{d}u\mathrm{d}v \\
&= \sum_{j=1}^N \int_t^T\int_t^T \mathrm{cov}_{r_j}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad + \sum_{i=1}^N\sum_{j=i+1}^N \int_t^T\int_t^T \mathrm{cov}_{r_i,r_j}(u,v)+\mathrm{cov}_{r_j,r_i}(u,v)\mathrm{d}u\mathrm{d}v
\end{aligned}
\tag{6.14}
$$

According to the one factor model, the first sum in the previous equations is given by:

$$
\begin{aligned}
&\sum_{j=1}^N \int_t^T\int_t^T \mathrm{cov}_{r_j}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad = \sum_{j=1}^N \left(\frac{\sigma_j^2}{2k_j^3}\left(2k_j(T-t)+4e^{-k_j(T-t)}-e^{-2k_j(T-t)}-3\right)\right)
\end{aligned}
\tag{6.15}
$$

The covariance $\mathrm{cov}_{r_i,r_j}(u,v)$, in Equation 6.14, is a function of $\min(u,v)$. In order compute the double integral, we first split the inner integral in the intervals from t to v, and from v to T. In the first interval we have $\min(u,v)=u$ while in the second $\min(u,v)=v$. A similar step is taken in the one factor model, in order to compute the double integral over $\mathrm{cov}_{r_j}(u,v)$.

$$\begin{aligned}
&\int_t^T \int_t^T \mathrm{cov}_{r_i,r_j}(u,v) + \mathrm{cov}_{r_j,r_i}(u,v)\mathrm{d}u\mathrm{d}v \\
&= \int_t^T \int_t^v \mathrm{cov}_{r_i,r_j}(u,v) + \mathrm{cov}_{r_j,r_i}(u,v)\mathrm{d}u\mathrm{d}v \\
&\quad + \int_t^T \int_v^T \mathrm{cov}_{r_i,r_j}(u,v) + \mathrm{cov}_{r_j,r_i}(u,v)\mathrm{d}u\mathrm{d}v \\
&= \int_t^T \int_t^v \rho_{ij}\frac{\sigma_i\sigma_j}{k_i+k_j}\left(e^{(k_i+k_j)\min(u,v)} - e^{(k_i+k_j)t}\right)\left(e^{-k_iu-k_jv} + e^{-k_ju-k_iv}\right)\mathrm{d}u\mathrm{d}v \\
&\quad + \int_t^T \int_v^T \rho_{ij}\frac{\sigma_i\sigma_j}{k_i+k_j}\left(e^{(k_i+k_j)\min(u,v)} - e^{(k_i+k_j)t}\right)\left(e^{-k_iu-k_jv} + e^{-k_ju-k_iv}\right)\mathrm{d}u\mathrm{d}v \\
&= \int_t^T \int_t^v \rho_{ij}\frac{\sigma_i\sigma_j}{k_i+k_j}\left(e^{(k_i+k_j)u} - e^{(k_i+k_j)t}\right)\left(e^{-k_iu-k_jv} + e^{-k_ju-k_iv}\right)\mathrm{d}u\mathrm{d}v \\
&\quad + \int_t^T \int_v^T \rho_{ij}\frac{\sigma_i\sigma_j}{k_i+k_j}\left(e^{(k_i+k_j)v} - e^{(k_i+k_j)t}\right)\left(e^{-k_iu-k_jv} + e^{-k_ju-k_iv}\right)\mathrm{d}u\mathrm{d}v \\
&= \frac{2\rho_{ij}\sigma_i\sigma_j e^{-(k_i+k_j)(t+3T)}}{k_i^2k_j^2(k_i+k_j)}\left(-e^{(k_i+k_j)(t+3T)}\left(k_i^2(k_jt-k_jT+1)+k_ik_j(k_jt-k_jT+1)+k_j^2\right)\right. \\
&\quad \left. +k_ik_j\left(-e^{2(k_i+k_j)(t+T)}\right) + k_j(k_i+k_j)e^{2k_i(t+T)+k_j(t+3T)} + k_i(k_i+k_j)e^{k_i(t+3T)+2k_j(t+T)}\right)
\end{aligned} \tag{6.16}$$

Therefore:

$$\begin{aligned}
&\sum_{i=1}^N \sum_{j=i+1}^N \int_t^T \int_t^T \mathrm{cov}_{r_i,r_j}(u,v) + \mathrm{cov}_{r_j,r_i}(u,v)\mathrm{d}u\mathrm{d}v \\
&= \sum_{i=1}^N \sum_{j=i+1}^N \frac{2\rho_{ij}\sigma_i\sigma_j e^{-(k_i+k_j)(t+3T)}}{k_i^2k_j^2(k_i+k_j)} \\
&\quad \left(-e^{(k_i+k_j)(t+3T)}\left(k_i^2(k_jt-k_jT+1)+k_ik_j(k_jt-k_jT+1)+k_j^2\right)\right. \\
&\quad \left. +k_ik_j\left(-e^{2(k_i+k_j)(t+T)}\right) + k_j(k_i+k_j)e^{2k_i(t+T)+k_j(t+3T)} + k_i(k_i+k_j)e^{k_i(t+3T)+2k_j(t+T)}\right)
\end{aligned} \tag{6.17}$$

Finally

$$
\begin{aligned}
&\mathrm{var}_x(t)\\
&= \sum_{j=1}^{N}\left(\frac{\sigma_j^2}{2k_j^3}\left(2k_j(T-t)+4e^{-k_j(T-t)}-e^{-2k_j(T-t)}-3\right)\right)\\
&\quad +\sum_{i=1}^{N}\sum_{j=i+1}^{N}\Bigg(\frac{2\rho_{ij}\sigma_i\sigma_j e^{-(k_i+k_j)(t+3T)}}{k_i^2k_j^2(k_i+k_j)}\\
&\quad \left(-e^{(k_i+k_j)(t+3T)}\left(k_i^2(k_jt-k_jT+1)+k_ik_j(k_jt-k_jT+1)+k_j^2\right)\right.\\
&\quad \left.+k_ik_j\left(-e^{2(k_i+k_j)(t+T)}\right)+k_j(k_i+k_j)e^{2k_i(t+T)+k_j(t+3T)}+k_i(k_i+k_j)e^{k_i(t+3T)+2k_j(t+T)}\right)\Bigg)
\end{aligned}
\tag{6.18}
$$

Given Equations 6.4, 6.13 and 6.18:

$$
p(t,T) = e^{\sum_{j=1}^{N} A_j(t,T)-\sum_{j=1}^{N} B_j(t,T)r_j(t)+\sum_{i=1}^{N}\sum_{j=i+1}^{N} C_{ij}(t,T)} \tag{6.19}
$$

where

$$
B_j(t,T) = \frac{1}{k_j}\left(1-e^{-k_j(T-t)}\right) \tag{6.20}
$$

and

$$
A_j(t,T) \;=\; \left(\theta_j-\frac{\sigma_j^2}{2k_j^2}\right)\left(B_j(t,T)-T+t\right)-\frac{\sigma_j^2}{4k_j}B_j^2(t,T) \tag{6.21}
$$

and

$$
\begin{aligned}
&C_{ij}(t,T)\\
&= \frac{\rho_{ij}\sigma_i\sigma_j e^{-(k_i+k_j)(t+3T)}}{k_i^2k_j^2(k_i+k_j)}\left(-e^{(k_i+k_j)(t+3T)}\left(k_i^2(k_jt-k_jT+1)+k_ik_j(k_jt-k_jT+1)+k_j^2\right)\right.\\
&\quad \left.+k_ik_j\left(-e^{2(k_i+k_j)(t+T)}\right)+k_j(k_i+k_j)e^{2k_i(t+T)+k_j(t+3T)}+k_i(k_i+k_j)e^{k_i(t+3T)+2k_j(t+T)}\right)
\end{aligned}
\tag{6.22}
$$

It is clear from Equation 6.19 that the model has an affine term structure and that the logarithm of the zero coupon bonds prices is a Gaussian process.

6.4.3 Zero coupon bond log prices prior mean

The logarithm of zero coupons bonds prices is given by

$$\log p(t,T) = \sum_{j=1}^{N} A_j(t,T) - \sum_{j=1}^{N} B_j(t,T) r_j(t) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T) \tag{6.23}$$

The mean function, $m_p(t,T)$, of $\log p(t,T)$ is given by

$$\begin{aligned} m_p(t,T) &= E[\log p(t,T)] \\ &= E\left[\sum_{j=1}^{N} A_j(t,T) - \sum_{j=1}^{N} B_j(t,T) r_j(t) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T)\right] \\ &= \sum_{j=1}^{N} A_j(t,T) - \sum_{j=1}^{N} B_j(t,T) E\left[r_j(t)\right] + \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T) \\ &= \sum_{j=1}^{N} A_j(t,T) - \sum_{j=1}^{N} B_j(t,T) m_{r_j}(t) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T) \end{aligned}$$

6.4.4 Zero coupon bond log prices prior covariance

The covariance function, $\text{cov}(t_1, T_1, t_2, T_2)$, of $\log p(t,T)$ is, by definition

$$\begin{aligned} &\text{cov}(t_1, T_1, t_2, T_2) \\ &= E\left[\left(\log p(t_1, T_1) - m_p(t_1, T_1)\right)\left(\log p(t_2, T_2) - m_p(t_2, T_2)\right)\right] \end{aligned} \tag{6.24}$$

The term $\log p(t,T) - m_p(t,T)$ is given by

$$\begin{aligned} &\log p(t,T) - m_p(t,T) \\ &= \sum_{j=1}^{N} A_j(t,T) - \sum_{j=1}^{N} B_j(t,T) r_j(t) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T) \\ &\quad - \sum_{j=1}^{N} A_j(t,T) + \sum_{j=1}^{N} B_j(t,T) m_{r_j}(t) - \sum_{i=1}^{N} \sum_{j=i+1}^{N} C_{ij}(t,T) \\ &= - \sum_{j=1}^{N} B_j(t,T)\left(r_j(t) - m_{r_j}(t)\right) \end{aligned} \tag{6.25}$$

Therefore

$$
\begin{aligned}
&\mathrm{cov}(t_1, T_1, t_2, T_2) \\
&= E\left[\left(-\sum_{j=1}^{N} B_j(t_1, T_1)(r_j(t_1) - m_{r_j}(t_1))\right)\right. \\
&\qquad \left.\left(-\sum_{j=1}^{N} B_j(t_2, T_2)(r_j(t_2) - m_{r_j}(t_2))\right)\right] \\
&= E\left[\sum_{i=1}^{N}\sum_{j=1}^{N} B_i(t_1, T_1)B_j(t_2, T_2)(r_i(t_1) - m_{r_i}(t_1))(r_j(t_2) - m_{r_j}(t_2))\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} B_i(t_1, T_1)B_j(t_2, T_2)E\left[(r_i(t_1) - m_{r_i}(t_1))(r_j(t_2) - m_{r_j}(t_2))\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} B_i(t_1, T_1)B_j(t_2, T_2)\mathrm{cov}_{r_i,r_j}(t_1, t_2) \qquad (6.26)
\end{aligned}
$$

In matrix terms, given vectors $\mathbf{B}_N(t_1, T_1)$, $\mathbf{B}_N(t_2, T_2)$, and matrix $\mathbf{M}_N(t_1, t_2)$

$$\mathbf{B}_N(t_1, T_1) = [B_1(t_1, T_1) \quad B_2(t_1, T_1) \quad \cdots \quad B_N(t_1, T_1)] \qquad (6.27)$$

$$\mathbf{B}_N(t_2, T_2) = [B_1(t_2, T_2) \quad B_2(t_2, T_2) \quad \cdots \quad B_N(t_2, T_2)]^\top \qquad (6.28)$$

$$\mathbf{M}_N(t_1, t_2) = \begin{bmatrix} \mathrm{cov}_{r_1,r_1}(t_1, t_2) & \mathrm{cov}_{r_1,r_2}(t_1, t_2) & \cdots & \mathrm{cov}_{r_1,r_N}(t_1, t_2) \\ \mathrm{cov}_{r_2,r_1}(t_1, t_2) & \mathrm{cov}_{r_2,r_2}(t_1, t_2) & \cdots & \mathrm{cov}_{r_2,r_N}(t_1, t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}_{r_N,r_1}(t_1, t_2) & \mathrm{cov}_{r_N,r_2}(t_1, t_2) & \cdots & \mathrm{cov}_{r_N,r_N}(t_1, t_2) \end{bmatrix} \qquad (6.29)$$

the covariance function is given by

$$\mathrm{cov}(t_1, T_1, t_2, T_2) = \mathbf{B}_N(t_1, T_1)\mathbf{M}_N(t_1, t_2)\mathbf{B}_N(t_2, T_2) \qquad (6.30)$$

6.5 Machine learning Gaussian short rate

As in (Sousa, Esquível, and Gaspar 2013), let:

- $\mathbf{x} = [t \quad T]^\top$;

- $y = \log p(\mathbf{x}) = \log p(t, T)$;
- $m_p(\mathbf{x}) = m_p(t, T)$;
- $cov_p(\mathbf{x}_i, \mathbf{x}_j) = cov_p(t_i, T_i, t_j, T_j)$;
- matrix $\mathbf{X}$ collects a set of vectors $\mathbf{x}^\diamond$ where the values of zero coupon log prices were observed;
- vector $\mathbf{y}$ collects the corresponding values $y^\diamond = \log p(\mathbf{x}^\diamond)$.

The machine learning Gaussian short rate is the Gaussian short rate process, $r_\mathcal{D}(t)$, underlying the zero coupon bond prices

$$p_\mathcal{D}(t, T) = E\left[e^{-\int_t^T r_\mathcal{D}(u)du}\right] \tag{6.31}$$

where $\log p_\mathcal{D}(t, T) = \log p_\mathcal{D}(\mathbf{x})$ is the conditioned on zero coupon bonds log prices Gaussian process

$$\log p_\mathcal{D}(\mathbf{x}) \sim \mathcal{GP}\left(m_{p_\mathcal{D}}(\mathbf{x}), cov_{p_\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)\right). \tag{6.32}$$

Functions $m_{p_\mathcal{D}}(\mathbf{x})$ and $cov_{p_\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j)$ are given by

$$m_{p_\mathcal{D}}(\mathbf{x}) = m_p(\mathbf{x}) + \mathbf{K}_{\mathbf{X},\mathbf{x}}^\top \mathbf{K}^{-1}(\mathbf{y} - \mathbf{m}) \tag{6.33}$$

and

$$cov_{p_\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j) = cov_p(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{X},\mathbf{x}_i}^\top \mathbf{K}^{-1}\mathbf{K}_{\mathbf{X},\mathbf{x}_j} \tag{6.34}$$

where

- $m_p(\mathbf{x})$ is given by Equation 6.24.
- $cov_p(\mathbf{x}_i, \mathbf{x}_j)$ is given by Equation 6.26.
- $\mathbf{m}$ is the prior mean on the training set. It results from applying $m_p(\mathbf{x})$ function (Equation 3.16) on all $\mathbf{X}$ collected vectors;
- $\mathbf{K}$ is the prior covariance matrix on the training set. It results from applying $cov_p(\mathbf{x}_i, \mathbf{x}_j)$ function (Equation 3.18) on all pairs of $\mathbf{X}$ collected vectors;
- $\mathbf{K}_{\mathbf{X},\mathbf{x}}$ is the prior covariance between every vector in the training set and $\mathbf{x}$. It results from applying $cov_p(\mathbf{x}_i, \mathbf{x}_j)$ function (Equation 3.18) on all pairs composed by each $\mathbf{X}$ collected vector, and the $\mathbf{x}$ vector.

6.6 Conclusions

In this paper we extend the one factor machine learning Gaussian short rate model by considering a sum of N Vasicek factors as prior. The prior zero coupon bond log prices mean and covariance functions are obtained.

The extended model supports a maximum of N distinct zero coupon bond prices in each time instant. Therefore it can fit the current term structures of interest rates, observed in the market. We show, using simulated data, calibration examples of term structures with distinct shapes.

All the model parameters are obtained directly under the risk neutral model by maximizing the likelihood of market observed zero coupon bond log prices given the parameters. Besides zero coupon bond prices, no other sources of information are needed.

7

Interest Rate Market Changes Detection

7.1 Preamble

A common problem that arises in mathematical finance when using models with parameters estimated from market data, available until a certain time, is that of checking the necessity of using new parameters as newer data become available.

In this paper we use a covariance matrix statistical test to evaluate the necessity of using new parameters of a machine learning Gaussian short rate model of the Euribor, as newer data become available. Whenever we detect such necessity, we say that the market conditions have changed.

The main contributions of this paper are:

- Obtain the likelihood ratio criterion to test if a covariance matrix, Σ , is equal to a given matrix, Σ_0, as a decomposition of simpler tests.

- Propose a machine learning Gaussian short rate model with Vasicek short rate noise in the observations.

- Using real data, model the Euribor with the proposed model and apply the changes detection procedure to the credit crisis years of 2007 and 2008.

Abstract

In this paper we check for interest rate market changes, using the distribution of the likelihood ratio criterion, to test if a covariance matrix, Σ , is equal to a given matrix, Σ_0.

We start by transforming the original test into the equivalent test $\Sigma = I$. Then, the test $\Sigma = I$ is decomposed into two conditional independent tests, namely, the sphericity test $\Sigma = \sigma^2 I$, and the test $\sigma^2 = 1$, given that the data are spherical. The distribution moments and characteristic function are obtained. The characteristic function inversion is done numerically.

We apply the covariance matrix test to check interest rate market changes using Euribor real data. We model the Euribor with a one factor machine learning Gaussian short rate model, using the Vasicek short rate model as prior, and assuming Vasicek short rate noise in the observations. In the beginning we calibrate the model to get a reference parameters set. Then, in the presence of newer data, we recalibrate the model and get a newer parameters set. We check the validity of the reference parameters set, using the statistical test applied to the model observations covariance matrix computed with both sets of parameters. Whenever the newer covariance matrix is not equal to the reference one, we say that the market conditions have changed.

Keywords: Covariance matrix test, Conditionally independent tests, Sphericity test, Identity matrix test.

7.2 Introduction

Let $\mathbf{Y}$ be a $p \times 1$ multivariate Normal random vector, with mean $\boldsymbol{\mu}_Y$ and variance-covariance matrix Σ_Y:

$$\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}_Y, \Sigma_Y). \tag{7.1}$$

If we are interested in testing the null hypothesis

$$H_0 : \Sigma_Y = \Sigma_0 \,,$$

since Σ_0 is positive definite, Σ_0^{-1} exists and also $\Sigma_0^{-1/2}$, and thus we may transform the original random vector $\mathbf{Y}$ to

$$\mathbf{X} = \Sigma_0^{-1/2}\mathbf{Y},$$

where

$$\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma),$$

with $\Sigma = \Sigma_0^{-1/2}\Sigma_Y\Sigma_0^{-1/2}$.

Therefore, testing $H_0 : \Sigma_Y = \Sigma_0$ is equivalent to test

$$H_0 : \Sigma = \Sigma_0^{-1/2}\Sigma_0\Sigma_0^{-1/2} = I_p\,. \tag{7.2}$$

The test of the null hypothesis H_0 in (7.2) can be decomposed by testing $H_{01} : \Sigma = \sigma^2 I_p$ (the sphericity test) in the first place, and then, if H_{01} is not rejected, testing $H_{02|01} : \sigma^2 = 1$. This decomposition is denoted by

$$H_0 = H_{02|01} \circ H_{01}. \tag{7.3}$$

Using this decomposition, the statistic for testing $\Sigma = \Sigma_0$, is obtained. Then, the moments and characteristic function of the corresponding distribution are also obtained.

7.3 Likelihood ratio test statistic

Given a sample of size $n+1$ of the random vector $\mathbf{X}$, the likelihood ratio criterion to test $H_{01} : \Sigma = \sigma^2 I_p$, the sphericity test, is (T. W. Anderson 2003)

$$\Lambda_1 = \frac{|A|^{\frac{1}{2}(n+1)}}{(\frac{1}{p}\mathrm{tr}A)^{\frac{1}{2}p(n+1)}}, \tag{7.4}$$

where A is the matrix of sums of squares and cross products of deviations to the sample mean $\bar{\mathbf{x}}$. Namely,

$$A = \sum_{k=1}^{n+1}(\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})' \tag{7.5}$$

and

$$\bar{\mathbf{x}} = \frac{1}{n+1}\sum_{k=1}^{n+1}\mathbf{x}_k. \tag{7.6}$$

Using the maximum likelihood estimator of σ^2, under $H_{01} : \Sigma = \sigma^2 I_p$ (see Appendix B), the likelihood ratio criterion to test $H_{02|01} : \sigma^2 = 1$, given that $\Sigma = \sigma^2 I_p$, is

$$\Lambda_2 = \left(\frac{\mathrm{tr} A}{p(n+1)} \right)^{\frac{1}{2}p(n+1)} e^{\frac{1}{2}(p(n+1) - \mathrm{tr}A)}, \tag{7.7}$$

so that the likelihood ratio criterion to test H_0 in (7.2) is (T. W. Anderson 2003)

$$\Lambda_0 = \Lambda_1 \Lambda_2 = \frac{\left(\frac{e}{n+1}\right)^{\frac{1}{2}p(n+1)} |A|^{\frac{1}{2}(n+1)}}{e^{\frac{1}{2}\mathrm{tr}A}}. \tag{7.8}$$

However, according to (T. W. Anderson 2003) and (Sugiura and Nagao 1968) the test based on Λ_0 in (7.8) is biased. In order to have an unbiased test, as suggested by (Sugiura and Nagao 1968), we will use the statistic

$$\Lambda_* = \frac{\left(\frac{e}{n}\right)^{\frac{1}{2}pn} |A|^{\frac{1}{2}n}}{e^{\frac{1}{2}\mathrm{tr}A}} .$$

7.4 Moments of Λ_*

Since the matrix A has a Wishart distribution $W(A|\Sigma, n)$, the moments $E\{\Lambda_*^h\}$ of Λ_* are given by

$$E\{\Lambda_*^h\} = \int \Lambda_*^h f_A(a; \Sigma, n)\, da \,,$$

where

$$f_A(a; \Sigma, n) = \frac{|a|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}a)}}{2^{\frac{1}{2}pn} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{j=1}^{p} \Gamma(\frac{1}{2}(n+1-j))}$$

is the p.d.f. of a $W(A|\Sigma, n)$ distribution.

$E\{\Lambda_*^h\}$ thus becomes

$$E\{\Lambda_*^h\} = \left(\frac{e}{n}\right)^{\frac{1}{2}pnh} \int \frac{|A|^{\frac{1}{2}(n+nh-p-1)} e^{-\frac{1}{2}\mathrm{tr}((\Sigma^{-1}+hI)A)}}{2^{\frac{1}{2}pn} \pi^{p(p-1)/4} |\Sigma|^{\frac{1}{2}n} \prod_{j=1}^{p} \Gamma(\frac{1}{2}(n+1-j))}\, da \,,$$

where we recognize the numerator of the integrand as a the numerator of the p.d.f. of a $W(A|(\Sigma^{-1}+hI)^{-1}, n+(n+1)h)$ distribution. Multiplying and dividing the whole expression by the denominator of $W(A|(\Sigma^{-1}+hI)^{-1}, n+(n+1)h)$ p.d.f.

and rearranging the terms, we obtain

$$\begin{aligned}
&E\{\Lambda_*^h\} \\
&= \left(\frac{2e}{n}\right)^{\frac{1}{2}pnh} \frac{\prod_{j=1}^p \Gamma(\frac{1}{2}(n+nh+1-j))}{|\Sigma|^{\frac{1}{2}n}|\Sigma^{-1}+hI|^{\frac{1}{2}(n+nh)}\prod_{j=1}^p \Gamma(\frac{1}{2}(n+1-j))} \\
&\qquad \underbrace{\int f_A(a;(\Sigma^{-1}+hI)^{-1}, n+nh)\, da}_{=1} \\
&= \left(\frac{2e}{n}\right)^{\frac{1}{2}pnh} \frac{\prod_{j=1}^p \Gamma(\frac{1}{2}(n+nh+1-j))}{|\Sigma|^{\frac{1}{2}n}|\Sigma^{-1}+hI|^{\frac{1}{2}(n+nh)}\prod_{j=1}^p \Gamma(\frac{1}{2}(n+1-j))}.
\end{aligned}$$

Finally, since under H_0 in (7.2) $\Sigma = I$,

$$E\{\Lambda_*^h\} = \left(\frac{2e}{n}\right)^{\frac{1}{2}pnh} (1+h)^{-\frac{1}{2}p(n+nh)} \prod_{j=1}^p \frac{\Gamma(\frac{1}{2}(n+nh+1-j))}{\Gamma(\frac{1}{2}(n+1-j))}. \tag{7.9}$$

7.5 Characteristic function of $W = -\log \Lambda_*$

If we take $W = -\log \Lambda_*$, the characteristic function of W is, from (7.9), given by

$$\begin{aligned}
\phi_W(t) &= E\{e^{itW}\} = E\{e^{-it\log \Lambda_*}\} = E\{\Lambda_*^{-it}\} \\
&= \left(\frac{2e}{n}\right)^{-\frac{1}{2}pnit} (1-it)^{-\frac{1}{2}p(n-nit)} \prod_{j=1}^p \frac{\Gamma(\frac{1}{2}(n-nit+1-j))}{\Gamma(\frac{1}{2}(n+1-j))}.
\end{aligned} \tag{7.10}$$

7.6 Market changes detection

A common problem that arises in mathematical finance when using models with parameters estimated from market data, available until a certain time, is that of checking the necessity of using new parameters as newer data become available.

In this section we use the $\Sigma = \Sigma_0$ test to evaluate the necessity of using new parameters of a machine learning Gaussian short rate model of the Euribor (Euro Interbank Offered Rate), as newer data become available.

We use the Euribor data available online at the Euribor-EBF site (Euribor-EBF 2013). We model the short rate, using the 1, 6 and 12 months Euribor rates, daily quoted by the 48 banks in the Euribor contributor banks panel, during 2007 and 2008, the credit crisis years.

The machine learning Gaussian short rate model has as inputs the logarithm of zero coupon bonds. Therefore, all the Euribor quotes are first converted to

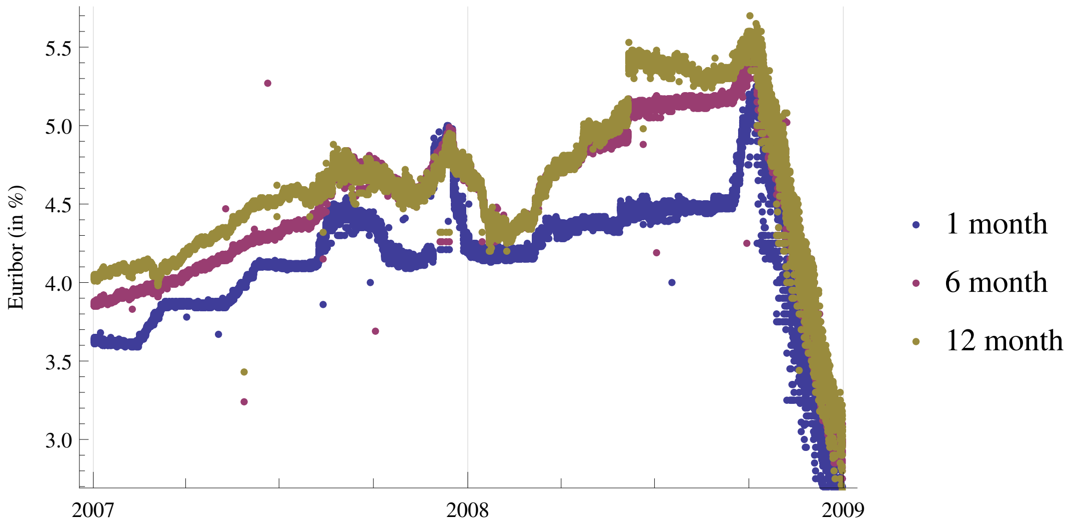

Figure 7.1: Euribor rates with maturities of 1, 6 and 12 months, quoted by the Euribor contributor banks in 2007 and 2008.

the corresponding 1, 6 and 12 months maturity zero coupon bonds prices. Then, the logarithm of the prices is taken. The model parameters are obtained using periods of 5 consecutive days.

All the computations were done using Wolfram Mathematica 8 (Wolfram Research 2011).

7.6.1 Euribor data

Figure 7.1 shows the 1, 6 and 12 months maturity Euribor raw rates quoted by all the Euribor contributor banks[1] in 2007 and 2008. Figure 7.2 shows the corresponding zero coupon bonds log prices.

In our model each 5 consecutive days period is a random vector. Every maturity in each of the 5 days is a marginal random variable. Since there are 3 maturities in each day, the random vector dimension is 15. Therefore, the number of variables is $p = 15$.

The Mardia goodness of fit test (Mardia 1970)(Mardia 1974) was used to test the data normality. The test was first carried over each univariate marginal random variable, and then over each random vector.

Table 7.1 shows the normality goodness of fit test results for the univariate marginal random variables.

The small percentage of normal variables, observed in Table 7.1 forces us to

[1]During 2007 and 2008 the number of banks in the Euribor contributors panel that supplied quotes varied, day by day, ranging from 39 to 48 banks.

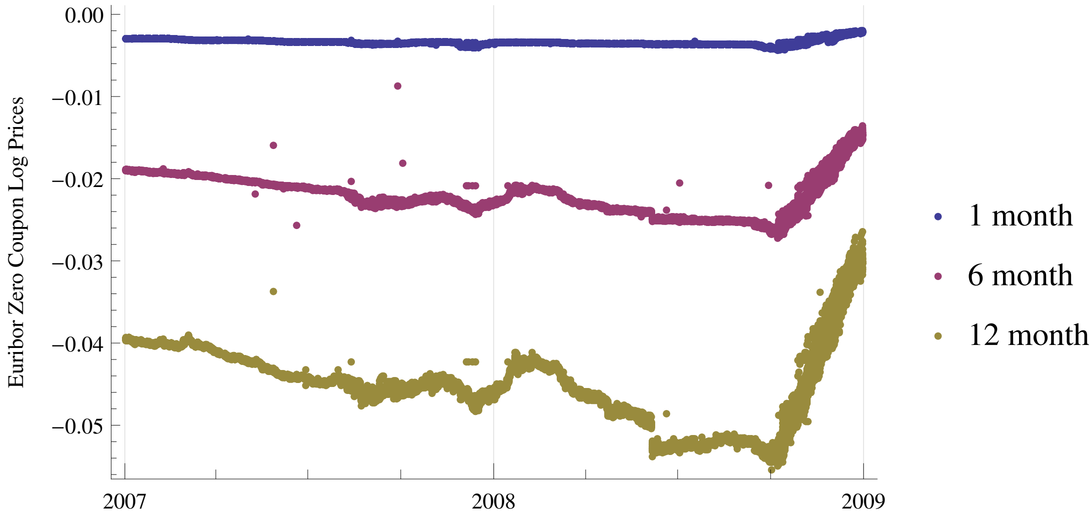

Figure 7.2: Zero coupon bond log prices with maturities of 1, 6 and 12 months, computed from the corresponding Euribor rates, quoted by the Euribor contributor banks in 2007 and 2008.

# Variables	# Normal Variables	Normal Variables (in %)
1533	49	3.20

Table 7.1: Marginals random variables normality distribution fit test results.

conclude that the data are not normal. Given this fact, the model is applied not directly to Euribor zero coupon log prices, computed from the Euribor rates quoted by the contributor banks, but to randomly selected portfolios of banks. The construction of randomly selected portfolios of banks, definitely makes sense from a diversification point of view, and should provide normal data (due to the central limit theorem) as required by our model.

We have chosen to construct 16 randomly selected portfolios of 24 banks. The number of portfolios results in a number of observations of $n + 1 = p$ which is required by the test statistics. The number of banks in each portfolio was empirically defined to maximize the number of normal random variables obtained.

Figure 7.3 shows the 1, 6 and 12 months maturity zero coupon bonds portfolios log prices obtained.

Table 7.2 shows the normality goodness of fit test results for the portfolios marginal random variables as well as for the random vectors.

The large percentage of random variables and random vectors in Table 7.2 allows us to assume that the portfolios data are normal, as required by our model.

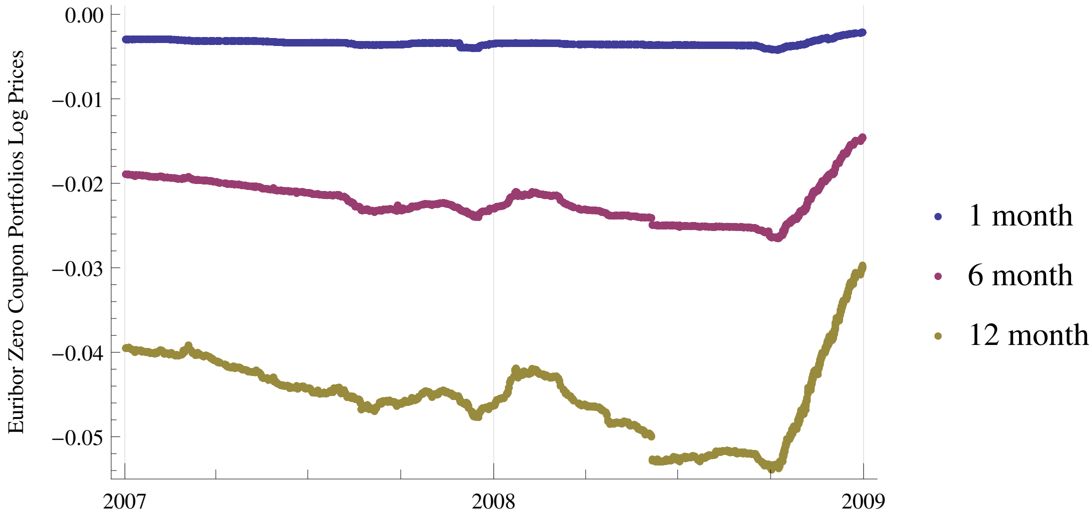

Figure 7.3: Zero coupon bond portfolios log prices with maturities of 1, 6 and 12 months, computed from the corresponding Euribor rates, quoted by the Euribor contributor banks in 2007 and 2008. The number of portfolios at each time and each maturity is 16. Each portfolio price is the average of 24 randomly selected zero coupon bonds, computed from the quoted contributor banks rates.

Portfolios Marginal Variables

# Variables	# Normal Variables	Normal Variables (in %)
1533	1382	90.15

Portfolios Vector Variables

# Vectors	# Normal Vectors	Normal Vectors (in %)
507	507	100.00

Table 7.2: Portfolios normality fit test results.

7.6.2 Short rate model

The short rate model considered is a one factor machine learning Gaussian short rate, under the risk neutral measure, with noise in the observations.

Under the one factor machine learning Gaussian short rate (Sousa, Esquível, and Gaspar 2013), the short rate prior, $r(t)$, follows an Ornstein-Uhlenbeck mean-reverting process, under the risk neutral measure, defined by the stochastic differential equation (Vasicek 1977):

$$dr(t) = k(\theta - r(t))dt + \sigma dW(t). \tag{7.11}$$

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

Considering the vector $\mathbf{x} = [t \quad T]^\top$, where t is time and T is zero coupon bond maturity, the short rate, $r_{\mathcal{D}}(t)$, is the Gaussian process underlying the conditioned on zero coupon log prices Gaussian process with mean and covariance functions given by

$$m_{\mathcal{D}}(\mathbf{x}) = m(\mathbf{x}) + \mathbf{K}_{\mathbf{X},\mathbf{x}}^T \mathbf{K}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \tag{7.12}$$

and

$$cov_{\mathcal{D}}(\mathbf{x}_i, \mathbf{x}_j) = cov(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{X},\mathbf{x}_i}^T \mathbf{K}^{-1} \mathbf{K}_{\mathbf{X},\mathbf{x}_j} \tag{7.13}$$

where $m(\mathbf{x})$ is the zero coupon bonds log prices prior mean

$$\begin{aligned} m(\mathbf{x}) &= m\left(\begin{bmatrix} t \\ T \end{bmatrix}\right) \\ &= \left(\theta - \frac{\sigma^2}{2k^2}\right)\left(\frac{1 - e^{-k(T-t)}}{k} - T + t\right) \\ &\quad - \frac{\sigma^2 \left(1 - e^{-k(T-t)}\right)^2}{4k^3} \\ &\quad - \frac{\left(1 - e^{-k(T-t)}\right)\left(r(0)e^{-kt} + \theta\left(1 - e^{-kt}\right)\right)}{k}, \end{aligned} \tag{7.14}$$

$cov(\mathbf{x}_i, \mathbf{x}_j)$ is the zero coupon bonds log prices prior covariance

$$\begin{aligned} cov(\mathbf{x}_i, \mathbf{x}_j) &= cov\left(\begin{bmatrix} t_i \\ T_i \end{bmatrix}, \begin{bmatrix} t_j \\ T_j \end{bmatrix}\right) \\ &= \frac{\sigma^2 e^{-k(t_i+t_j)}\left(1 - e^{-k(T_i-t_i)}\right)\left(1 - e^{-k(T_j-t_j)}\right)\left(e^{2k\min(t_i,t_j)} - 1\right)}{2k^3}, \end{aligned} \tag{7.15}$$

matrix $\mathbf{X}$ collects all the vectors $\mathbf{x}$ where zero coupon bond log prices were observed, $\mathbf{K}_{\mathbf{X},\mathbf{x}}$ is the covariance vector between every observation and $\mathbf{x}$, $\mathbf{K}$ is the covariance matrix between observations, $\mathbf{y}$ is the vector of observed zero coupon log prices and $\boldsymbol{\mu}$ is the mean vector of observations.

This model requires a single observation of zero coupon bond prices for each time/maturity pair. Otherwise, matrix $\mathbf{K}$ would have some equal lines, and would not be invertible.

To support several zero coupon bond prices for each time/maturity pair, as it is the case with Euribor data, we will assume noise in the observations, following the machine learning procedure in (Rasmussen and Williams 2005).

But the proposal of constant variance noise in (Rasmussen and Williams 2005) makes no sense in the context of a risk neutral short rate model, as it would result in zero coupon bond log prices with non zero variance at maturity. And this would constitute a no arbitrage violation once the prices at maturity must equal the face value with zero variance.

To overcome this issue our proposal is to assume addictive independent Gaussian Vasicek zero coupon bond log prices process noise in the observations.

The proposed noise mean interest rate level is $\theta_n = 0$ because the noise is modeling the dispersion around the short rate only, not the mean. The proposed mean reversion velocity is $k_n = 1$ because the noise observations are independent of each other, providing no characterization of a reversion velocity. The only extra parameter introduced is the noise volatility, σ_n.

Given this noise model, the covariance function, $cov_o(\mathbf{x}_i, \mathbf{x}_j)$, between observations is given by:

$$\begin{aligned} cov_o(\mathbf{x}_i, \mathbf{x}_j) &= cov_o\left(\begin{bmatrix} t_i \\ T_i \end{bmatrix}, \begin{bmatrix} t_j \\ T_j \end{bmatrix}\right) \\ &= \frac{\sigma^2 e^{-k(t_i+t_j)}\left(1-e^{-k(T_i-t_i)}\right)\left(1-e^{-k(T_j-t_j)}\right)\left(e^{2k\min(t_i,t_j)}-1\right)}{2k^3} \\ &\quad + \frac{\sigma_n^2\left(1-e^{-(T_i-t_i)}\right)^2\left(1-e^{-2t_i}\right)}{2}\delta_{ij} \end{aligned} \tag{7.16}$$

where δ_{ij} is the Kronecker delta, which is one if i equals j, and zero otherwise.

We have considered the initial time sufficiently far away, so that the initial short rate value $r(0)$, in Equation 7.14, has no numerical influence in the model. This allows us to ignore parameter $r(0)$. All remaining parameters are obtained by maximizing the likelihood of observed zero coupon bond log prices using the conjugate gradients method and the closed forms of the derivatives of the likelihood with respect to the parameters.

7.6.3 Experimental procedure

In order to evaluate whether or not new parameters should be used as newer data become available, the following experimental procedure was followed.

Setup

- Set the confidence level to 0.05.

Initialization

- Obtain the initial reference model parameters using the first five days period data.

Evaluation

- Slide the 5 days period window 1 day.
- Obtain new model parameters using the new five days period data.
- Compute new period data observations covariance matrix, Σ, using the new parameters.
- Compute new period data observations covariance matrix, Σ_0, using the reference parameters.
- Apply the test $\Sigma = \Sigma_0$ to obtain the corresponding p-value.

Update

- Substitute the reference parameters by the new parameters.
- Proceed to the Evaluation step while there are more data.

The inversion of the characteristic function of Equation 7.10 was done numerically using the algorithm in (Abate and Valkó 2004).

Figure 7.4 illustrates the proposed model for a particular 5 day period. The data adjustment clearly stands out.

7.6.4 Results

Figure 7.5 shows the parameters sequences obtained, while Figure 7.6 shows the covariance matrix test results p-values sequence.

Compared with the parameters sequences, that exhibit some smoothness during relatively large periods of time, the covariance matrix test results are quite unstable. Recall that whenever the p-value is less than the confidence level we say that the market conditions have changed. In general, the detected changes are mixed with no changes detections. No clear periods of changes or no changes were obtained, suggesting an over-fitting behavior. An exception to this is the end of 2007 where a clear period of changes detections is observable. This was the period when the Bank of Canada, the Bank of England, the European Central Bank, the Federal Reserve, and the Swiss National Bank announced joint actions to address high pressures in short-term funding markets (Bank 2007), thus confirming that it was a worldwide turbulent period.

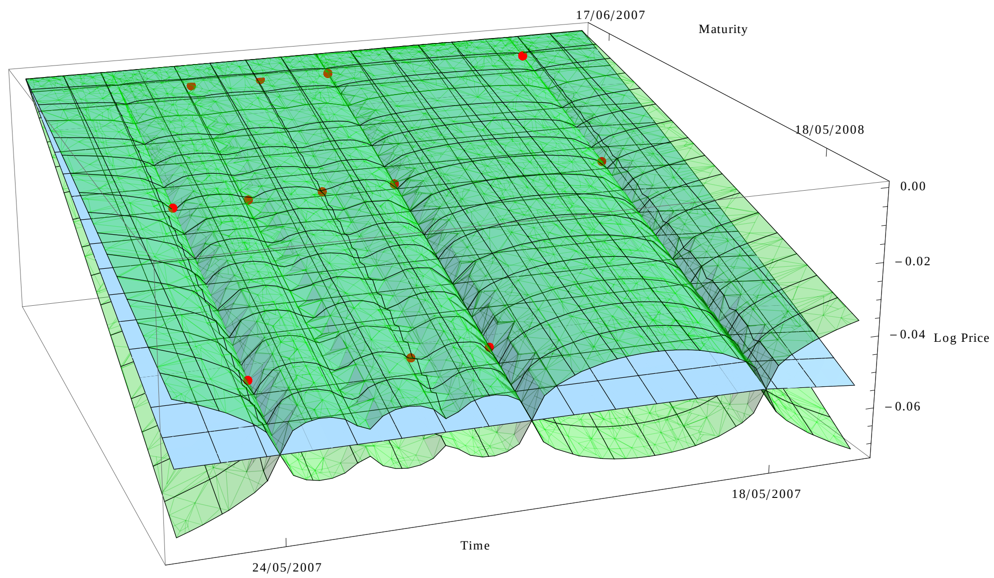

Figure 7.4: Euribor's 5 day short rate model zero coupon bond log prices: data (red); mean (blue); 2 standard deviation surface (green).

7.7 Conclusions

In this paper we obtained the statistic for testing if a covariance matrix is equal to a given matrix, as the decomposition of simpler tests. We also obtained the moments and the characteristic function of the test distribution.

We proposed the extension of the one factor machine learning Gaussian short rate model by assuming noise on the observations. The proposed model is suitable for situations where there are several prices for a zero coupon bond at the same time, such as the quotes provided by the Euribor banks panel.

We modeled the Euribor during the 2007 and 2008 years, the credit crisis years, with the proposed model, using sliding 5 days periods of data. We used the covariance matrix test to detect Euribor market changes.

The market changes detection results obtained are quite unstable suggesting an over-fitting behavior. Nevertheless the turbulent short-term funding period of the end of 2007 stands out as a period of consecutive market changes.

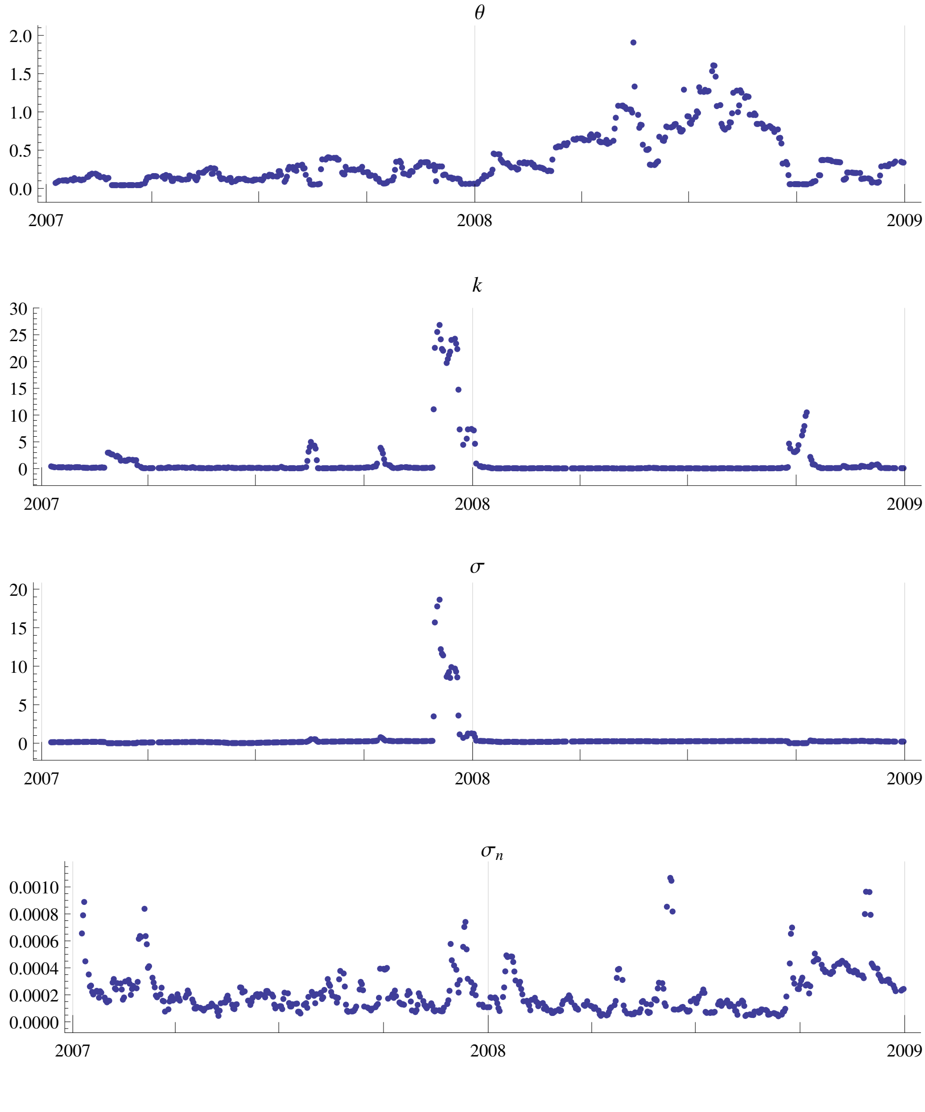

Figure 7.5: Short rate model parameters sequences.

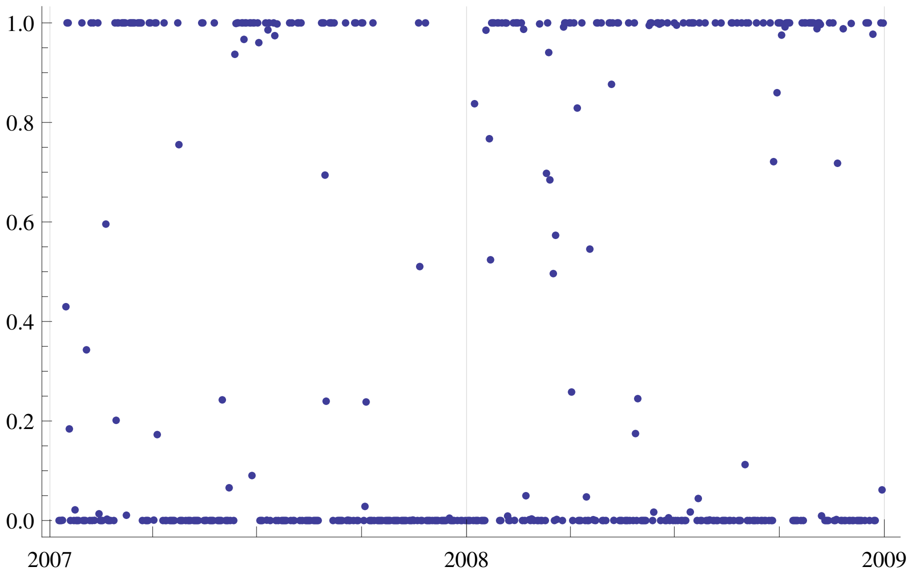

Figure 7.6: $\Sigma = \Sigma_0$ test, p-values sequence.

8

Conclusions and Future Work

8.1 Thesis contributions

The main contribution of this thesis is the proposal of a Gaussian short rate model, under the risk neutral measure. The model is a conditioned on zero coupon bond log prices Gaussian process, based on Gaussian processes for machine learning regression, with N addictive risk neutral Vasicek short rate factors as prior. It has the following features:

- All model parameters are obtained directly in the risk neutral measure using market observed zero coupon bond prices,exclusively. No other sources of information are needed;
- The model automatically fits by its construction all the zero coupon bond prices observed in the market, in particular, those that define the current term structure of interest rates;
- The number of factors needed equals the maximum number of distinct maturities observed in a single time.

Other contributions also emerged during this research and development, namely:

- The proposal of a path dependent Gaussian trajectories vectorial simulation framework;

- The proposal of a historical simulation value at risk methodology for bonds;
- The detection of interest rate market changes with a statistical test to a covariance matrix;
- The extension of the proposed short rate model to the case where there are more than one price of T-maturity zero coupon bonds at time t, considering Vasicek short rate noise in the observations.

8.2 Future work

Regarding the unsubmitted papers in part II of this thesis, the additional work we plan to include before submitting them to publication, is composed by the following improvements.

- Machine Learning Gaussian Short Rate
  - Proceeding in the same way in (Sousa, Esquível, and Gaspar 2013) obtain the stochastic differential equation parameters of the short rate.
  - Obtain the derivatives with respect to the parameters of the likelihhod of the training data given the parameters.
  - Show, by simulation, that the risk neutral parameters are properly obtained from zero coupon log prices.
  - Use the model with real data.
- Interest Rate Market Changes Detection

  All the development of the statistical test presented in this paper, as a decomposition of simpler conditional independent tests, has the purpose of factorizing the test statistic in two parts. One that can be inverted analytically and another that can not. The part that can't be analytically inverted, can be approximated by Gamma densities mixture to produce a near exact test distribution the same way other near exact distribution were developed (Marques and Coelho 2008)(Coelho and Marques 2010).

  Preliminary results show that the numerical inversion of the characteristic function of the near exact distribution perform faster and more accurate than the numerical inversion used in this paper.

  The additional work to include in this paper is to formalize the near-exact distribution and use it in the presented application.

Bibliography

Abate, J. and P. Valkó (2004). “Multi-precision Laplace transform inversion”. In: *International Journal for Numerical Methods in Engineering* 60.5, pp. 979–993.

Alabert, F. (1987). “The practice of fast conditional simulations through the LU decomposition of the covariance matrix”. In: *Mathematical Geology* 19.5, pp. 369–386.

Alexander, C. (2009). *Market Risk Analysis, Value at Risk Models*. John Wiley & Sons.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Third. Wiley series in probability and statistics. New York: John Wiley & Sons. ISBN: 0471360910.

Bank, E. C. (Dec. 2007). *Measures designed to address elevated pressures in short-term funding markets*. URL: http://www.ecb.europa.eu/press/pr/date/2007/html/pr071212.en.html.

Bingham, N. and R. Kiesel (2004). *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer Finance. Springer.

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

Black, F., E. Derman, and W. Toy (1990). “A one-factor model of interest rates and its application to treasury bond options”. In: *Financial analysts journal*, pp. 33–39.

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

Broadie, M. and P. Glasserman (1997). “Pricing American-style securities using simulation”. In: *Journal of Economic Dynamics and Control* 21.8, pp. 1323–1352.

Coelho, C. A. and F. J. Marques (2010). “Near-exact distributions for the independence and sphericity likelihood ratio test statistics”. In: *Journal of Multivariate Analysis* 101.3, pp. 583–593.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). "A theory of the term structure of interest rates". In: *Econometrica: Journal of the Econometric Society*, pp. 385–407.

Darbha, G. (2001). *Value-at-Risk for Fixed Income portfolios – A comparison of alternative models*. Tech. rep. National Stock Exchange, Mumbai.

Davis, M. W. (1987). "Production of conditional simulations via the LU triangular decomposition of the covariance matrix". In: *Mathematical Geology* 19.2, pp. 91–98.

Dietrich, C. and G. Newsam (1996). "A fast and exact method for multidimensional Gaussian stochastic simulations: Extension to realizations conditioned on direct and indirect measurements". In: *Water resources research* 32.6, pp. 1643–1652.

Dothan, L. U. (1978). "On the term structure of interest rates". In: *Journal of Financial Economics* 6.1, pp. 59–69.

Euribor-EBF (Sept. 2013). *Euribor*. URL: http://www.euribor-ebf.eu/euribor-org/euribor-rates.html.

Garland, M., S. Le Grand, J. Nickolls, J. Anderson, J. Hardwick, S. Morton, E. Phillips, Y. Zhang, and V. Volkov (2008). "Parallel computing experiences with CUDA". In: *Micro, IEEE* 28.4, pp. 13–27.

Gillespie, D. T. (1996). "Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral". In: *Physical review E* 54.2, p. 2084.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability) (v. 53)*. Springer.

Group, Q. (2012). *QuantLib - a free/open-source library for quantitative finance*. URL: http://quantlib.org/.

Ho, T. S. and S.-B. LEE (1986). "Term structure movements and pricing interest rate contingent claims". In: *The Journal of Finance* 41.5, pp. 1011–1029.

Hoffman, Y. and E. Ribak (1991). "Constrained realizations of Gaussian fields-A simple algorithm". In: *The Astrophysical Journal* 380, pp. L5–L8.

Hull, J. C. (2008). *Options, Futures, and Other Derivatives*. 7th ed. Prentice Hall.

Hull, J. and A. White (1990a). "Pricing Interest-Rate-Derivative Securities". In: *Review of Financial Studies* 3.4, pp. 573–92.

— (1990b). "Pricing interest-rate-derivative securities". In: *Review of financial studies* 3.4, pp. 573–592.

Jean-François, C. (2000). "Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study". In: *Journal of statistical software* 5, pp. 1–53.

Kasdin, N. J. (1995). "Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation". In: *Proceedings of the IEEE* 83.5, pp. 802–827.

Kloeden, P. E. and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. Springer.

Ltaief, H., S. Tomov, R. Nath, and J. Dongarra (2010). "Hybrid multicore cholesky factorization with multiple gpu accelerators". In: *IEEE Transaction on Parallel and Distributed Systems*.

Mandelbrot, B. B. (1971). "A fast fractional Gaussian noise generator". In: *Water Resources Research* 7.3, pp. 543–553.

Mardia, K. V. (1970). "Measures of multivariate skewness and kurtosis with applications". In: *Biometrika* 57.3, pp. 519–530.

— (1974). "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies". In: *Sankhya: The Indian Journal of Statistics, Series B*, pp. 115–128.

Marques, F. J. and C. A. Coelho (2008). "Near-exact distributions for the sphericity likelihood ratio test statistic". In: *Journal of Statistical Planning and Inference* 138.3, pp. 726–741.

Moskowitz, B. and R. E. Caflisch (1996). "Smoothness and dimension reduction in quasi-Monte Carlo methods". In: *Mathematical and Computer Modelling* 23.8, pp. 37–54.

Pang, K. (1998). "Calibration of Gaussian Heath, Jarrow and Morton and random field interest rate term structure models". In: *Review of Derivatives Research* 2.4, pp. 315–345.

Paxson, V. (1997). "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic". In: *ACM SIGCOMM Computer Communication Review* 27.5, pp. 5–18.

Pérignon, C. and D. R. Smith (2010). "The level and quality of Value-at-Risk disclosure by commercial banks". In: *Journal of Banking & Finance* 34.2, pp. 362–377.

Pritsker, M. (Feb. 2006). "The hidden dangers of historical simulation". In: *Journal of Banking & Finance* 30.2, pp. 561–582.

Rainer, M. (2009). "Calibration of stochastic models for interest rate derivatives". In: *Optimization* 58.3, pp. 373–388.

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

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

Research, W. (2013). *Mathematica: Version 9*.

Sousa, J. B., M. L. Esquível, and R. M. Gaspar (2012). "Machine Learning Vasicek Model Calibration with Gaussian Processes". In: *Communications in Statistics - Simulation and Computation* 41.6, pp. 776–786.

Sousa, J. B., M. L. Esquível, and R. M. Gaspar (2013). "One Factor Machine Learning Gaussian Short Rate". In: *Communications in Statistics-Simulation and Computation*. Submitted.

Sugiura, N. and H. Nagao (1968). "Unbiasedness of some test criteria for the equality of one or two covariance matrices". In: *The Annals of Mathematical Statistics* 39.5, pp. 1686–1692.

Vasicek, O. (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5.2, pp. 177–188.

Volkov, V. and J. Demmel (2008). "LU, QR and Cholesky factorizations using vector capabilities of GPUs". In: *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2008-49, May*, pp. 2008–49.

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

— (2011). *Mathematica Edition: Version 8.0.4.0*.

— (2012). *Mathematica Edition: Version 9.0.1.0*.

A

Wolfram Mathematica Sources

Listing A.1: Public section of "VasicekPackage" symbolic package.

```
BeginPackage["VasicekPackage`"]

(* in this package Zcb means "Zero Coupon Bond" *)






VasicekShortRateMean::usage = "VasicekShortRateMean[t0, r0, t, k,
  \[Theta]] gives the Vasicek short rate mean at time t, given the
  initial time t0, the initial short rate r0, the mean short rate
  level \[Theta] and the mean revearsion velocity k.";

VasicekShortRateCovariance::usage = "VasicekShortRateCovariance[t0,
  t1, t2, k, \[Sigma]] gives the Vasicek short rate covariance
  beteween times t1 and t2, given the initial time t0, the mean
  revearsion velocity k and the volatility \[Sigma].";

VasicekBFunction::usage = "VasicekBFunction[t, T, k] is the auxiliary
  Vasicek B(t,T) function."

VasicekAFunction::usage = "VasicekAFunction[t, T, k, \[Theta],
  \[Sigma]] is the auxiliary Vasicek A(t,T) function."

VasicekLogZcbMeanByDefinition::usage =
  "VasicekLogZcbMeanByDefinition[t0, r0, t, T, k, \[Theta], \[Sigma]]
  gives the mean of the logarithm of the price of a T maturity zero
  coupon bond at time t, given the initial time t0, the initial short
  rate value r0, the mean interest rate level \[Theta], the mean
  revearsion velocity k and the volatility \[Sigma]."
```

```
VasicekLogZcbMean::usage = "VasicekLogZcbMean[t0, r0, t, T, k,
\[Theta], \[Sigma]] is a FullSimplify version of
VasicekLogZcbMeanByDefinition[t0, r0, t, T, k, \[Theta], \[Sigma]]."

VasicekLogZcbCovarianceByDefinition::usage =
  "VasicekLogZcbCovarianceByDefinition[t0, t1, t2, T1, T2, k,
  \[Sigma]] gives the covariance at times t1 and t2 of the logarithm
  of the price of a T1 maturity zero coupon bond and a T2 maturity
  zero coupon bond given the initial time t0, the mean revearsion
  velocity k and the volatility \[Sigma]."

VasicekLogZcbCovariance::usage = "VasicekLogZcbCovariance[t0, t1, t2,
  T1, T2, k, \[Sigma]] is a FullSimplify version of
  VasicekLogZcbCovarianceByDefinition[t0, t1, t2, T1, T2, k,
  \[Sigma]]."




Begin["`Private`"]

(* Private context to big to print *)




End[]




EndPackage[]
```

Listing A.2: Public section of "OneFactorWithNoiseShortRate" numeric package.

```
BeginPackage["OneFactorWithNoiseShortRate`"]




SetData::usage = "SetData[tListN, TListN, yListN] sets all the
internal packed arrays needed to learn the model parameters. The
argument tListN is the list of time instants. The argument TListN is
the list of maturity times. The argument yListN is the list of zero
coupon bond log prices. Each triplet {tVectorN[[k]], TVectorN[[k]],
yVectorN[[k]]} is a T maturity zero coupon bond log price y at time
t. All arguments lists must be lists of machine numbers with the same
length. ";

TestInternalPackedArrays::usage = "Tests if internal variables are
packed arrays.";
```

```
ComputeMeanVector::usage = "ComputeMeanVector[r0, kr,
thetar, sigmar, sigman] computes the mean vector.";

GetMeanVector::usage = "GetMeanVector[] returns the
internal mean vector.";

ComputeCovarianceMatrix::usage = "ComputeCovarianceMatrix[r0, kr,
thetar, sigmar, sigman] computes the covariance matrix.";

GetCovarianceMatrix::usage = "GetCovarianceMatrix[] returns the
internal covariance matrix.";

ComputeLikelihood::usage = "ComputeLikelihood[r0, kr, thetar, sigmar,
sigman] returns the likelioohd of the previous seted data given the
parameters arguments.";

ComputeLikelihoodNumericWrapper::usage =
"ComputeLikelihoodNumericWrapper[r0, kr, thetar, sigmar, sigman]
is a numeric wrapper for ComputeLikelihood[r0, kr, thetar, sigmar,
sigman]. Maximizing the likelihhod using FindMaximum should use this
wrapper in order to avoid symbolic computations (that grow memory to
infinite).";




ComputeLikelihoodDr0::usage = "ComputeLikelihoodDr0[r0, kr, thetar,
sigmar, sigman] returns the likelioohd derivative w.r.t. r0.";

ComputeLikelihoodDkr::usage = "ComputeLikelihoodDkr[r0, kr, thetar,
sigmar, sigman] returns the likelioohd derivative w.r.t. kr.";

ComputeLikelihoodDthetar::usage = "ComputeLikelihoodDthetar[r0, kr,
thetar, sigmar, sigman] returns the likelioohd derivative
w.r.t. thetar.";

ComputeLikelihoodDsigmar::usage = "ComputeLikelihoodDsigmar[r0, kr,
thetar, sigmar, sigman] returns the likelioohd derivative
w.r.t. sigmar.";

ComputeLikelihoodDsigman::usage = "ComputeLikelihoodDsigman[r0, kr,
thetar, sigmar, sigman] returns the likelioohd derivative
w.r.t. sigman.";




ComputeLikelihoodDr0NumericWrapper::usage = "The numeric wrapper for
ComputeLikelihoodDr0.";

ComputeLikelihoodDkrNumericWrapper::usage = "The numeric wrapper for
ComputeLikelihoodDkr.";

ComputeLikelihoodDthetarNumericWrapper::usage = "The numeric wrapper for
```

```
ComputeLikelihoodDthetar.";

ComputeLikelihoodDsigmarNumericWrapper::usage = "The numeric wrapper
for ComputeLikelihoodDsigmar.";

ComputeLikelihoodDsigmanNumericWrapper::usage = "The numeric wrapper
for ComputeLikelihoodDsigman.";





ConditionedOnDataMeanFunction::usage =
"ConditionedOnDataMeanFunction[t, T, r0, kr, thetar, sigmar, sigman]
returns the conditioned on data mean value on time t, maturity T.";

ConditionedOnDataCovarianceFunction::usage =
"ConditionedOnDataCovarianceFunction[t1, T1, t2, T2, r0, kr, thetar,
sigmar, sigman] returns the conditioned on data covariance value
between time t1, maturity T1 and time t2, maturity T2.";

SetObservationsFactor::usage = "SetObservationsFactor[factorNumber]
sets the factor to be multiplyed by the covariance matrix, in case of
independent vector observations";




Begin["`Private`"]

(* Private context to big to print *)




End[]




EndPackage[]
```

B

Maximum likelihood estimator of σ^2 in test $H_{02|01}$

In this appendix we deduce the maximum likelihood estimator of σ^2, in case the null hypothesis $H_{02|01} : \sigma^2 = 1$, given that $H_{01} : \Sigma = \sigma^2 I_p$ was not rejected, is rejected.

Let $\mathbf{X}$ be a $p \times 1$ multivariate Normal random vector, with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\sigma^2 I_p$:

$$\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \sigma^2 I_p). \tag{B.1}$$

Given a sample of $n + 1$ vectors $\mathbf{x}_1, ..., \mathbf{x}_n, \mathbf{x}_{n+1}$ of $\mathbf{X}$, the likelihood function L is

$$L = \prod_{k=1}^{n+1} \frac{1}{(2\pi)^{\frac{1}{2}p} |\sigma^2 I_p|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}_k - \boldsymbol{\mu})'(\sigma^2 I_p)^{-1}(\mathbf{x}_k - \boldsymbol{\mu})}. \tag{B.2}$$

Since the random vector $\mathbf{X}$ is Normal with a diagonal covariance matrix, the its components are independent. As so, we can rewrite L as

$$L = \prod_{k=1}^{n+1} \prod_{l=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(\mathbf{x}_{kl} - \boldsymbol{\mu}_l)^2}{\sigma^2}}, \tag{B.3}$$

where $\mathbf{x}_{kl}$ is the l component of vector $\mathbf{x}_k$ and $\boldsymbol{\mu}_l$ is the l component of vector $\boldsymbol{\mu}$.

The logarithm of the likelihood function is

$$\log L = \sum_{k=1}^{n+1}\sum_{l=1}^{p}\left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\frac{(\mathbf{x}_{kl}-\boldsymbol{\mu}_l)^2}{\sigma^2}\right). \tag{B.4}$$

Since $\log L$ is an increasing function of L, the values of $\boldsymbol{\mu}_l$ and σ^2 that maximizes $\log L$ also maximizes L.

Denoting by $\hat{\boldsymbol{\mu}}_m$ $(m = 1, ..., p)$ the value of $\boldsymbol{\mu}_m$ that maximizes $\log L$, and by $\hat{\sigma}^2$ the value of σ^2 that maximizes $\log L$, $\hat{\boldsymbol{\mu}}_m$ and $\hat{\sigma}^2$ can be found by setting the partial derivatives of $\log L$ w.r.t $\hat{\boldsymbol{\mu}}_m$ and w.r.t. $\hat{\sigma}^2$, respectively, equal to zero.

The derivative of $\log L$ w.r.t $\hat{\boldsymbol{x}}_m$ is

$$\begin{aligned}\frac{\partial \log L}{\partial \hat{\boldsymbol{\mu}}_m} &= \sum_{k=1}^{n+1}\frac{\mathbf{x}_{km}-\hat{\boldsymbol{\mu}}_m}{\hat{\sigma}^2} && \text{(B.5)}\\ &= \frac{\sum_{k=1}^{n+1}\mathbf{x}_{km}-(n+1)\hat{\boldsymbol{\mu}}_m}{\hat{\sigma}^2}. && \text{(B.6)}\end{aligned}$$

Setting the derivative of $\log L$ w.r.t $\hat{\boldsymbol{x}}_m$ equal to zero yields

$$\hat{\boldsymbol{\mu}}_m = \frac{\sum_{k=1}^{n+1}\mathbf{x}_{km}}{n+1}. \tag{B.7}$$

The derivative of $\log L$ w.r.t $\hat{\sigma}^2$ is

$$\begin{aligned}\frac{\partial \log L}{\partial \hat{\sigma}^2} &= \sum_{k=1}^{n+1}\sum_{l=1}^{p}\left(-\frac{1}{2\hat{\sigma}^2}+\frac{1}{2}\frac{(\mathbf{x}_{kl}-\hat{\boldsymbol{\mu}}_l)^2}{(\hat{\sigma}^2)^2}\right) && \text{(B.8)}\\ &= -\frac{p(n+1)}{2\hat{\sigma}^2}+\frac{1}{2}\frac{\sum_{k=1}^{n+1}\sum_{l=1}^{p}(\mathbf{x}_{kl}-\hat{\boldsymbol{\mu}}_l)^2}{(\hat{\sigma}^2)^2}. && \text{(B.9)}\end{aligned}$$

Setting the derivative of $\log L$ w.r.t $\hat{\sigma}^2$ equal to zero yields

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{n+1}\sum_{l=1}^{p}(\mathbf{x}_{kl}-\hat{\boldsymbol{\mu}}_l)^2}{p(n+1)}. \tag{B.10}$$

Recognizing $\hat{\boldsymbol{\mu}}_m$ as the m component of the sample mean vector $\bar{\mathbf{x}}$ (equation 7.6), we can rewrite $\hat{\boldsymbol{\mu}}_m$ and $\hat{\sigma}^2$ as:

$$\hat{\boldsymbol{\mu}}_m = \bar{\boldsymbol{x}}_m \tag{B.11}$$

$$\hat{\sigma}^2 = \frac{\sum_{l=1}^{p} \sum_{k=1}^{n+1} (\mathbf{x}_{kl} - \bar{\boldsymbol{x}}_l)^2}{p(n+1)} \tag{B.12}$$

$$= \frac{\sum_{l=1}^{p} A_{ll}}{p(n+1)} \tag{B.13}$$

$$= \frac{1}{p(n+1)} trA, \tag{B.14}$$

where matrix A is defined by equation 7.5.

Equation B.14 defines the maximum likelihood estimator of σ^2, in case the null hypothesis $H_{02|01} : \sigma^2 = 1$, given that $H_{01} : \Sigma = \sigma^2 I_p$ was not rejected, is rejected.