Kriging of financial term-structures

Areski Cousin, Hassan Maatouk, Didier Rullière

▶ **To cite this version:**

Areski Cousin, Hassan Maatouk, Didier Rullière. Kriging of financial term-structures. European Journal of Operational Research, 2016, 255 (2), pp.631-648. 10.1016/j.ejor.2016.05.057 . hal-01206388v2

HAL Id: hal-01206388
https://hal.science/hal-01206388v2

Submitted on 7 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kriging of financial term-structures

Areski Cousin[†1], Hassan Maatouk[‡2] and Didier Rullière[†3]

(†) Université de Lyon, Université Claude Bernard Lyon 1, ISFA, Laboratoire SAF EA2429, 50 av. Tony Garnier, 69366 Lyon, France
(‡) Mines de Saint-Etienne, 158 Cours Fauriel, 42023 Saint-Étienne, France

Abstract

Due to the lack of reliable market information, building financial term-structures may be associated with a significant degree of uncertainty. In this paper, we propose a new term-structure interpolation method that extends classical spline techniques by additionally allowing for quantification of uncertainty. The proposed method is based on a generalization of kriging models with linear equality constraints (market-fit conditions) and shape-preserving conditions such as monotonicity or positivity (no-arbitrage conditions). We define the most likely curve and show how to build confidence bands. The Gaussian process covariance hyper-parameters under the construction constraints are estimated using cross-validation techniques. Based on observed market quotes at different dates, we demonstrate the efficiency of the method by building curves together with confidence intervals for term-structures of OIS discount rates, of zero-coupon swaps rates and of CDS implied default probabilities. We also show how to construct interest-rate surfaces or default probability surfaces by considering time (quotation dates) as an additional dimension.

JEL classification C63; E43; G12

Keywords Model risk; interest-rate curve; yield curve; OIS discount curve; implied default distribution; kriging; no-arbitrage constraints

1 Introduction

Constructing term-structures is at the heart of asset pricing and risk management. A term-structure is a curve which describes the evolution of some financial or economic quantities as a function of time horizon. Typical examples are the term-structure of risk-free interest-rates, the term-structure of bond yields or credit spreads, the term-structure of default probabilities or the term-structure of stock return implied volatilities. These curve are typically not directly observed in the market. Thus, the curve construction is based on a benchmark set of contingent financial instruments whose values explicitly depend on some part of the curve. In practice, market quotes of these products only provide a partial information on the term-structure since they can only be considered to be reliable for a small set of liquid maturities. The problem is then to transform a small set of market quotes into a continuum set of values representing the

[1]areski.cousin@univ-lyon1.fr
[2]hassan.maatouk@mines-stetienne.fr
[3]didier.rulliere@univ-lyon1.fr

evolution of the underlying quantity of interest with respect to time horizon.

On practical grounds, the curve is assumed to belong to a family of parametric functions (Nielson-Siegel functional Nelson and Siegel (1987), polynomial splines, Smith and Wilson (2001)) and its construction consists in finding the underlying parameters that best fits observed market quotes for all available maturities. In de Andrés Sánchez and Gómez (2004), the interest-rate term-structure is estimated from bid-ask spreads of underlying instruments using fuzzy regression techniques. Hagan and West (2006) provide a review of different interpolation techniques for curve construction. They introduce a monotone convex method and postulate a series of quality criterion such as ability to fit market quotes, arbitrage-freeness, smoothness, locality of interpolation scheme, stability of forward rate and consistency of hedging strategies. Andersen (2007) analyzes the use of hyperbolic tension splines for construction of interest-rate term structures. The underlying optimization allows the user to control the relative importance of fit precision with respect to shape preservation (smoothness of the curve, penalization of oscillations and excess convexity/concavity). In the same vein, Chiu et al. (2008) shows that L_1 cubic splines minimizes the curve oscillation without sacrificing good approximation of the data. Iwashita (2013) makes a survey of non-local spline interpolation techniques which preserve stability of forward rates. Other papers such as Ametrano and Bianchetti (2009), Chibane et al. (2009), Kenyon and Stamm (2012) or Fries (2013) are concerned with the adaptation of curve construction methods in a multi-curve interest-rate environment. Note that, in terms of interpolation scheme, there is no consensus towards a particular best practice method in all circumstances. In addition, the previous approaches does not account for the uncertainty embedded in the process of curve construction. This could be of primary importance given that the market inputs may be unreliable or even inexistent for some maturities.

This issue is related to the study of model uncertainty and its impact on risk management. This topic has been studied since a certain time period and, following the recent financial crisis, has received a particular interest. Impact of model risk on valuation and hedging of financial derivatives have been treated by, among others, Derman (1996), Eberlein and Jacod (1997), El Karoui et al. (1998), Green and Figlewski (1999), Branger and Schlag (2004), Cont (2006), Davis and Hobson (2007), Henaff (2010), Morini (2011). In most papers, the question of model risk is restricted to the class of derivative products. One of the main objective is to quantify model uncertainty, for instance to obtain bounds for the arbitrage free value of some derivative instruments, given some information on the underlying securities, such as marginal distribution of its price at some particular time horizons. In contrast, the question of model risk embedded in the construction of marginal distribution or term-structure function themselves has not been investigated as a main object, whatever it may concern discount curves, zero-coupon curves, swap basis curves, bond term structures or CDS-implied survival curves.

From Kimeldorf and Wahba (1970) and Mardia et al. (1996), it is well-known that spline fitting is a special case of kriging (see also Bay et al., 2015; Bay et al., 2016). In addition, kriging allows to account for quantification of uncertainty. Kriging has been developed in geosatistics to estimate the density of some mineral resource in the ground given a relatively small set of borehole, see Krige (1951), Matheron (1963), Cressie (1990). Its principle relies on the determination of the conditional distribution of a spatial random field given a set of observed values of the field. The main interest of this method is that it allows to build a predictor of

quantities of interest at other locations, as well as uncertainties relying on this prediction.

Kriging is now widely used in many fields like hydrology, air pollution, epidemiology, weather prediction, etc. to interpolate some quantity of interest given some known values at different locations. Despite its popularity, there are relatively few works concerning kriging in actuarial sciences or in finance. Many reference academic journal of these fields give only few or even no entry corresponding to the word "kriging". The method is however sometimes referred to using the terms "Gaussian Processes" and "machine learning". Some existing works using kriging methodology in actuarial sciences and finance concern for example dynamic lifetime adjustments (Debón et al., 2010), variable annuities valuation (Guojun, 2013; Gan and Sheldon Lin, 2015), nested simulation of expected shortfall (Liu and Staum, 2010), Vasicek model calibration (Sousa et al., 2012), stock market linkages (Asgharian et al., 2013) or credit scoring (Fernandes and Artes, 2015). Other works using spatial techniques are Kanevski et al. (2008) on interest rates, Benth (2015) for energy futures prices. Some preprints or conference papers also mention the fit of some financial models (Stutvoet, 2007), spatial insurance (Paulson and Hart, 2006), trading and hedging strategies (Baysal et al., 2008), valuation of Bermudan options (Ludkovski, 2015). Kriging methods naturally rely on some assumptions on the underlying random fields, and one must carefully consider all conditions that must be satisfied before constructing a kriging model.

In practice, the term-structure under construction has to satisfy several type of conditions. One of the most important condition is the compatibility of the curve with market data, i.e., if the curve is used to value a benchmark set of instruments (under a specific pricing rule), the resulting values shall be as close as possible to the observed market quotes. In many classical situations, the market-fit condition translates into a system of linear constraints which can be easily incorporated in kriging techniques. In addition, kriging can also handle the presence of noisy observations (using the so-called *nugget effect*). This may be relevant in situation where, due to the lack of liquidity, market quotes cannot be considered to be reliable. It is then possible to incorporate an additional level of uncertainty (degree of confidence) associated with market observations. Monotonicity constraints also appears to be important in many applications. For instance, the price of default-free zero-coupon bonds (or risk-free discount factors) is a non-increasing function of time-to-maturities under no-arbitrage assumption. Survival functions inferred from CDS spread term-structures are $[0, 1]$-valued non-increasing functions.

Recently, some authors have studied the integration of monotonicity constraints into Gaussian process emulators, see e.g. Golchi et al. (2015) and Kleijnen and Van Beers (2012). However, these methods do not guaranty monotonicity constraints in the entire domain. The article Abrahamsen and Benth (2001) also deals with the introduction of constraints at some location of a Gaussian process. In Maatouk and Bay (2014b), classical kriging has been improved to tackle monotonicity, positivity constraints or bounds constraints on the curve values. In the present paper, we show how "constrained" kriging techniques can be used to extend the classical spline interpolation approaches by additionally quantifying the uncertainty in some illiquid part of the curve.

The paper is organized as follows. Section 2 states the term-structure construction problem and gives specific examples of market-fit conditions and shape preserving constraints. In Section 3, we briefly recall Gaussian process modeling with interpolation conditions or with more

general linear equality constraints. In Section 4, we present the model defined in Maatouk and Bay (2014b) to incorporate monotonicity constraints into a Gaussian process emulator. We then study the associated properties such as convergence to the constrained interpolation spline and the estimation of the covariance hyper-parameters. In Section 5, based on real market data, we construct curves together with confidence intervals for different financial term-structures such as OIS discount curves, zero-coupon swap curves and CDS-implied default distributions.

2 The term-structure construction problem

The main ingredients in the construction of a term-structure function is a set of financial products whose value depends on some points of the curve. Then, observing the price of these products provides an indirect (and partial) information on the curve. The first step is then to specify the relation between the value of these products and the values of the curve at different time horizons. In this paper, we restrict ourselves to situations where this relation is linear. As we will see, this is the case in many practical situations such as the construction of corporate or sovereign bond yield curve, the construction of OIS discount curves, the construction of forward curves based on fixed-vs-floating interest-rate swaps or the construction of implied default rates based on CDS spreads.

2.1 Market-fit and shape-preserving conditions

The aim is to construct at some quotation date t a term-structure function $T \to P(t,T)$, based on the observation of a series of market quotes $S_1(t) \ldots, S_n(t)$ corresponding to the market value of n financial instruments with time-to-maturities $T_1, \ldots, T_n$. In what follows, the quantity T denotes a time length (as opposed to a calendar date), so that $P(t,T)$ corresponds to the value of the curve at time horizon T or at calendar date $t+T$. The observation of market quotes at time t provides a partial information on the curve at a set of time horizons or points $X = (\tau_1, \ldots, \tau_m)$, i.e., at some calendar dates $t+\tau_1, \ldots, t+\tau_m$ which typically correspond to payment dates of cash-flows.

The curve is (fully) compatible at time t with market observations if the vector $P(t,X) := (P(t,\tau_1), \ldots, P(t,\tau_m))^\top$ satisfies a linear system of the form

$$A_t \cdot P(t,X) = \boldsymbol{b}_t, \tag{1}$$

where A_t is $n \times m$ real-valued matrix and $\boldsymbol{b}_t$ is a n-dimensional column vector. Of course, A_t and $\boldsymbol{b}_t$ may depend on market quotes $S_1(t) \ldots, S_n(t)$, on the characteristics of the product cash-flows but also on the hypothesis made for assessing the value of these products at time t. Note that the number of observations n is typically smaller than the number of points m, so that the solution of system (1) lives in a linear space with dimension $m-n$.

When constructing a financial term-structure, one may consider some additional information on the shape of the curve. For instance, the function $T \to P(t,T)$ may be known to be decreasing with respect to time horizon T and its values may be bounded and belong to the interval $[0,1]$. This is typically the case when one wants to construct a curve of discount factors

(default-free zero-coupon bond prices) or an implied survival function (survival probabilities of a CDS reference entity). Violating these kinds of shape-preserving conditions results in term-structure functions that are typically not arbitrage-free.

Remark 2.1. *The term-structure construction problem can be stated in a two-dimensional setting when the evolution of time (quotation dates) is added as a second dimension. In that case, the aim is to construct a surface* $(t, T) \to P(t, T)$ *based on a series of market quotes* $S_1(t) \dots, S_n(t)$ *observed at several quotation dates* $t = t_1, \dots, t_N$*. Note that the cash-flows characteristics of these products may depend on time* t*. In particular, the underlying maturity dates could be different at every time* t*. If for any date* $t = t_1, \dots, t_N$*, the market-fit condition translates into a linear system, then this condition can again be expressed by a single system by concatenating for every time* $t = t_1, \dots, t_N$ *the* N *systems given as in Equation (1). The shape-preserving condition can be expressed as the intersection of shape-preserving conditions at time* $t = t_1, \dots, t_N$*.*

In the following, we give some examples where the construction of term-structures involves combination of linear equality constraints and shape preserving conditions such as monotony or positivity.

2.2 Classical examples of term-structures

In what follows, t denotes a particular quotation date, i.e., the date at which the market quotes are observed for the underlying family of contingent products. For each example, we only consider a single financial product in this family and we provide the linear relation which characterize its value. These relation then corresponds to one particular row of the linear system (1).

Corporate or sovereign bond yield curves

Let S be the observed market price of a corporate or a sovereign bond with time-to-maturity T and with a fixed coupon rate c. The price S and the coupon rate c are expressed in percentage of invested nominal. The set of coupon payment dates is given by $(t + \tau_1, \dots, t + \tau_p)$ where $\tau_1 < \dots < \tau_p = T$. The year fraction δ_k represents the time length $\tau_k - \tau_{k-1}$, $k = 1, \dots, p$ where $\tau_0 = 0$. The present value of this bond can be defined as a linear combination of some default-free zero-coupon bonds, i.e.,

$$c \sum_{k=1}^{p} \delta_k P^B(t, \tau_k) + P^B(t, \tau_p) = S, \tag{2}$$

where $P^B(t, \tau)$ represents the price at time t of a default-free zero-coupon bond with time-to-maturity τ. Note that, even if representation (2) obviously relies on a default-free assumption, it is commonly employed as an intermediary step in the computation of the so-called bond yield-to-maturity.[4] In this example, the curve $T \to P^B(t, T)$ will be inferred from a set of market fit conditions similar to (2), each of them corresponding to specific debt products but with different maturities. As a result, this set of conditions can be easily represented in the

[4]The bond yield associated with time-to-maturity T is defined as the constant rate of return $Y(t, T)$ such that the present value relation (2) holds exactly when all the involved ZC bonds have this rate of return.

form of a linear system as in Equation (1). In addition, the default-free assumption implies that the curve $T \to P^B(t,T)$ is decreasing if arbitrage opportunities are precluded.

Discount curves based on Overnight Indexed Swaps par rates

Due to legal terms of standard collateral agreements, a possible choice to build discount curves is to use market quotes of OIS-like instruments (see, for instance Hull and White, 2013, for more details). Let S be the par swap rate of an overnight indexed swap with maturity T and fixed leg payment dates $\tau_1 < \ldots < \tau_p = T$. The year fraction δ_k represents the time length $\tau_k - \tau_{k-1}$, $k = 1, \ldots, p$ where $\tau_0 = 0$. For overnight-index swaps this time length is typically equal to one year. The swap equilibrium relation takes the following linear form

$$S \sum_{k=1}^{p} \delta_k P^D(t, \tau_k) = 1 - P^D(t, T), \tag{3}$$

where $P^D(t,\tau)$ is the discount factor at time t associated with a time horizon τ. In the previous equation, the left hand side represents the fixed leg present value whereas the right hand side corresponds to the floating leg present value. For more details on the derivation of (3), the reader is referred to Fujii et al. (2010).

Under some circumstances, discount cuves can also be extracted from par rates of fixed versus Euribor swaps with Euribor tenor of 3 months or 6 months. This is the case for instance in the LTGA framework of Solvability 2 prudential regulation where the basis risk-free rates used for euro are constructed from the euro swap rates with a small adjustment for credit spread (see, e.g. CFO Forum and CRO Forum, 2010). The resulting market-fit condition will exactly have the same form as (3).

Note that, in the banking industry, discount curves are now understood as OIS based curves, see e.g. Pallavicini and Tarenghi (2010). The next example explains how to infer Euribor forward rates from quoted OIS rates and Euribor Swap rates.

Forward curves based on OIS and fixed versus Ibor-floating interest-rate swaps

Let S be the observed par rate of an interest rate swap with maturity time T and floating payments linked to a Libor or an Euribor rate associated with a tenor j (typically, $j = 3$ months or $j = 6$ months). The fixed-leg payment scheme is given by $\tau_1 < \cdots < \tau_p = T$ and the floating-leg payment scheme is given by $\tilde{\tau}_1 < \cdots < \tilde{\tau}_q = T$. For most liquid products, payment on the fixed leg are made with an annual frequency, so that τ_k corresponds to k years ahead from the current date t. The year fraction δ_k represents the time length $\tau_k - \tau_{k-1}$, $k = 1, \ldots, p$ ($\tau_0 = 0$) whereas the year fraction $\tilde{\delta}_i$ represents the time length $\tilde{\tau}_i - \tilde{\tau}_{i-1}$, $i = 1, \ldots, q$ ($\tilde{\tau}_0 = 0$). Note that the length between two consecutive dates on the floating leg should correspond to the Libor or Euribor tenor, i.e., depending on the case, $\tilde{\delta}_i \simeq 3$ months or $\tilde{\delta}_i \simeq 6$ months. As a result, given an OIS discount curve P^D, the swap equilibrium relation can be represented in a linear form with respect to some forward Libor or Euribor rates, i.e.,

$$S \sum_{k=1}^{p} \delta_k P^D(t, \tau_k) = \sum_{i=1}^{q} P^D(t, \tilde{\tau}_i) \tilde{\delta}_i F_j(t, \tilde{\tau}_i), \tag{4}$$

where $P^D(t,\tilde{\tau})$ is a risk-free discount factor at time t for maturity τ and $F_j(t,\tilde{\tau}_i) := F(t,\tilde{\tau}_{i-1},\tilde{\tau}_i)$ is the forward Libor or Euribor rate defined as the fixed rate to be exchanged at time $\tilde{\tau}_i$ against the j-tenor Libor or Euribor rate established at time $\tilde{\tau}_{i-1}$ so that the swap has zero value at time t. As in the previous example, the left hand side of 4 represents the fixed leg present value whereas the right hand side corresponds to the floating leg present value. For more details, see, for instance Chibane et al. (2009). Given a (pre-constructed) discount curve P^D and a set of swap par rates $S = S_1, \ldots, S_n$ corresponding to time-t market quotes of Euribor or Libor swaps with maturities $T = T_1, \ldots, T_n$, the forward curve $T \to F_j(t,T)$ under construction has to satisfy a linear market-fit condition. This condition takes the form of a linear system whose each line is given by (4). In addition, one can require that forward rates are positive, so that the underlying pseudo zero-coupon prices form a decreasing function of time horizon. This particular shape preserving property (positivity) can also be enforced in the interpolation procedure we will described in the next section.

Credit curves based on Credit Default Swaps spreads

Let S be the fair spread of a credit default swap with protection time horizon T and with premium payment dates $\tau_1 < \cdots < \tau_p = T$. If we denote by R the expected recovery rate of the reference entity and by δ_k the year fraction corresponding to time length $\tau_k - \tau_{k-1}$ ($\tau_0 = 0$), then the CDS swap equilibrium relation can be expressed as

$$S \sum_{k=1}^{p} \delta_k P^D(t,\tau_k) Q(t,\tau_k) = -(1-R) \int_0^T P^D(t,\tau) dQ(t,\tau), \tag{5}$$

where $P^D(t,\tau)$ is the risk-free discount factor at time t for time horizon τ and where $Q(t,\tau)$ is the probability (at time t) that the underlying reference entity has not defaulted before time horizon τ. Then, $Q(t,\tau)$ is the survival probability of the debt issuer in the time horizon τ. The left hand side of (5) represents the premium leg present value whereas the right hand side corresponds to the protection leg (or default leg) present value. We implicitly assume here that recovery, default and interest rates are stochastically independent. Using an integration by parts, it is straightfoward to show that survival probabilities $Q(t,\tau)$, $0 \le \tau \le T$, are linked through the following linear relation :

$$\begin{aligned} S \sum_{k=1}^{p} \delta_k P^D(t,\tau_k) Q(t,\tau_k) + (1-R) P^D(t,T) Q(t,T) \\ + (1-R) \int_0^T f^D(t,\tau) P^D(t,\tau) Q(t,\tau) d\tau = 1 - R \end{aligned} \tag{6}$$

where $f^D(t,\tau)$ is the instantaneous forward rate[5] at time t for time horizon τ. In practice, the integral involved in the expression of the protection leg present value is classically discretized on the premium time grid $\tau_1 < \cdots < \tau_n = T$, so that the continuous linear condition (6) can be stated as a discrete one given by

$$\begin{aligned} \sum_{k=1}^{p-1} \Big(S\delta_k P^D(t,\tau_k) + (1-R)(P^D(t,\tau_{k-1}) - P^D(t,\tau_k)) \Big) Q(t,\tau_k) \\ + \Big(S\delta_p P^D(t,T) + (1-R) P^D(t,\tau_{p-1}) \Big) Q(t,T) = 1 - R. \end{aligned} \tag{7}$$

[5]Instantaneous forward rates can be derived from discount factors through the following relation : $f^D(t,\tau) P^D(t,\tau) = -\frac{\partial P}{\partial \tau}(t,\tau)$.

At some fixed quotation date t, CDS protection is usually available for a set of liquid maturities $T_1, \ldots, T_n$. Then the construction of an implied survival function $T \to Q(t,T)$ consists in building a $[0,1]$-valued decreasing function that satisfies a system of n linear equality constraints as of (7).

3 Kriging under linear equality constraints

The term-structure construction approach we propose relies on an extension of kriging to linear equality and shape preserving constraints. In this section, we give a formal presentation of this interpolation technique when only linear equality constraints are considered. Section 4 explains how this technique can be generalized when some monotonicity constraints are added.

Kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian Process (GP) with a prior covariance function. This method is widely used in the domain of spatial analysis and computer experiments (see, e.g. Rasmussen and Williams, 2005). More formally, we consider the model $y = f(\boldsymbol{x})$, where f is an unknown real-valued function of a d-dimensional input variable $\boldsymbol{x} \in \mathbb{R}^d$. In the case where the computation of f is expensive and time-consuming or in the case where f is known only at some input locations, this function can be estimated using so-called *kriging* techniques. In that case, f is seen as a realization of a Gaussian Process (GP) Y defined as

$$Y(\boldsymbol{x}) := \mu(\boldsymbol{x}) + Z(\boldsymbol{x}),$$

where the deterministic function $\mu \ : \ \boldsymbol{x} \in \mathbb{R}^d \longrightarrow \mu(\boldsymbol{x}) \in \mathbb{R}$ is the mean of Y and Z is a zero-mean GP with covariance function

$$K \ : \ (\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{R}^d \times \mathbb{R}^d \longrightarrow K(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) \in \mathbb{R}.$$

The quantity $K(\boldsymbol{x}, \boldsymbol{x}')$ is thus the covariance of the values of the random field Y at two input locations $\boldsymbol{x}$ and $\boldsymbol{x}'$. We assume that the covariance function K is such that the random field Y have continuous and differentiable sample paths with probability one, see Abrahamsen (1997). Figure 1 (right) gives an illustration of some Gaussian processes sample paths. In numerical illustrations of Section 5, we consider Gaussian processes with d-dimensional covariance functions given as a tensor product, i.e., for $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_d)$:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \prod_{i=1}^{d} C_i(x_i - x'_i, \theta_i),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$ and σ^2 are respectively called *length* and *variance hyper parameters*. The function C_i are kernel correlation functions which depend on the length parameter θ_i and on $x_i - x'_i$, $i = 1, \ldots, d$, see Table 1 for some popular kernel correlation functions. Note that the length parameter θ can be interpreted as a correlation parameter[6] as it controls the degree of dependence amongst the values of the Gaussian process at any two points. The parameter σ controls the initial Gaussian process variance.

[6]For a Gaussian covariance kernel, it can be shown that, for any n, increasing $\boldsymbol{\theta}$ yields an increase of the vector $(Y(x^1), \ldots, Y(x^n))$ with respect to the supermodular order, a stochastic order which is well-known for comparing the degree of dependence.

3.1 Classical kriging

In classical kriging, the real function f is known to take some values $y_1, \ldots, y_n$ at some d-dimensional design points $x^{(1)}, \ldots, x^{(n)}$, so that $f(X) = \boldsymbol{y}$, where the design points are given as the rows of the $n \times d$ matrix $X = \left(x^{(1)}, \ldots, x^{(n)}\right)^\top \in \mathbb{R}^{n \times d}$, $f(X) = \left(f(x^{(1)}), \ldots, f(x^{(n)})\right)^\top \in \mathbb{R}^n$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$. One advantage of using a GP emulator is that, conditionally to the observation data $\boldsymbol{y}$, the conditional process $Y \mid Y(X) = \boldsymbol{y}$ is still a GP. This process is characterized by its (marginal) mean

$$\eta(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^\top \mathbb{K}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \quad \boldsymbol{x} \in \mathbb{R}^d \tag{8}$$

and its covariance function $\tilde{K}$ given by

$$\tilde{K}(\boldsymbol{x}, \boldsymbol{x}') = K(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}(\boldsymbol{x})^\top \mathbb{K}^{-1} \boldsymbol{k}(\boldsymbol{x}'), \quad \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d \tag{9}$$

where $\boldsymbol{\mu} = \mu(X) = \left(\mu(x^{(1)}), \ldots, \mu(x^{(n)})\right)^\top \in \mathbb{R}^n$ is the trend vector at the design points, $\mathbb{K}$ is the covariance matrix of $Y(X)$ and $\boldsymbol{k}(\boldsymbol{x}) = \left(K\left(\boldsymbol{x}, x^{(1)}\right), \ldots, K\left(\boldsymbol{x}, x^{(n)}\right)\right)^\top$ is the vector of covariance between $Y(\boldsymbol{x})$ and $Y(X)$. The conditional mean $\eta(\boldsymbol{x})$ given the observation data $Y(X) = \boldsymbol{y}$ is the Best Linear Unbiased Estimator (BLUE) of $Y(\boldsymbol{x})$, which is known as *kriging mean* (see Jones et al., 1998). One remarkable property is that the covariance function $\tilde{K}$ of the conditional Gaussian process does not depend on the observation data $\boldsymbol{y}$. In addition, the regularity of the kriging mean predictor function η inherits from the regularity of the mean function μ and from the regularity of the covariance function K of the original GP Y. Then, the choice of this covariance function is essential because it drives the smoothness of the kriging metamodel. Table 1 gives some popular kernel correlation functions, ordered by decreasing degree of smoothness. Figure 1 shows three alternative covariance functions with associated Gaussian process sample paths.

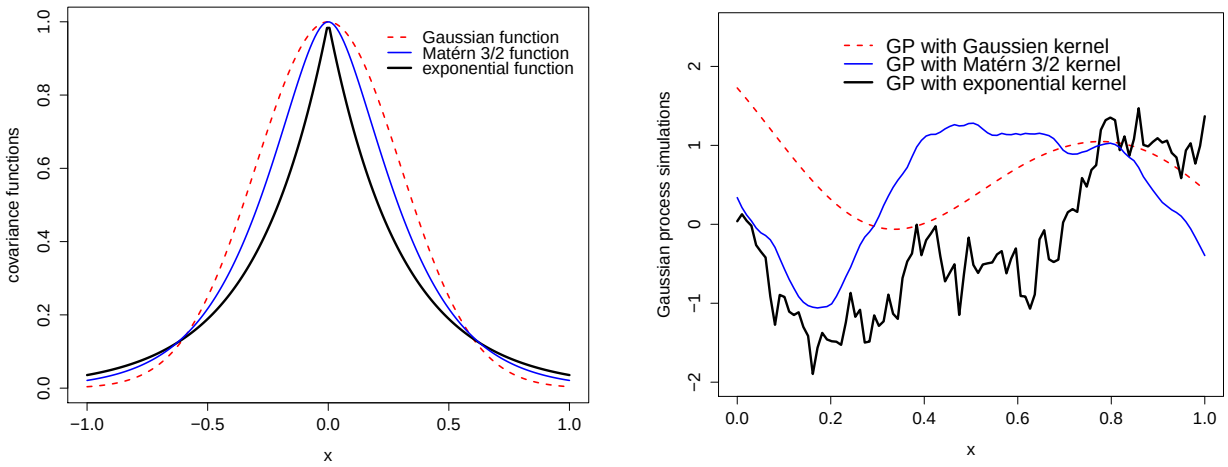

Figure 1: Some Gaussian process covariance functions (left) and associated sample paths (right). The covariance parameters are fixed to $(\theta, \sigma) = (0.3,\ 1.0)$.

Table 1: Some popular kernel correlation functions $C(x - x', \theta)$ used in kriging methods.

Name	Expression	Class
Gaussian	$\exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$	$\mathcal{C}^\infty$
Matérn 5/2	$\left(1 + \frac{\sqrt{5}\lvert x-x'\rvert}{\theta} + \frac{5(x-x')^2}{3\theta^2}\right)\exp\left(-\frac{\sqrt{5}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^2$
Matérn 3/2	$\left(1 + \frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right)\exp\left(-\frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^1$
Exponential	$\exp\left(-\frac{\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^0$

3.2 Extension to linear equality constraints

The previous setting can be generalized by considering linear equality constraints instead of pure interpolation constraints. This is of primary importance if one wants to construct term-structures which are compatible with linear market-fit conditions. We are under the situation where n relevant financial products are considered to construct the curve and their market quotes provide information on the curve at the m points $x^{(1)}, \dots, x^{(m)}$. Then, the (unknown) real function f satisfies some linear constraints of the form

$$A \cdot f(X) = \boldsymbol{b}, \tag{10}$$

where A is a given matrix of dimension $n \times m$, $n, m \in \mathbb{N}$, $X = \left(x^{(1)}, \dots, x^{(m)}\right)^\top \in \mathbb{R}^{m \times d}$, $f(X) = \left(f(x^{(1)}), \dots, f(x^{(m)})\right)^\top \in \mathbb{R}^m$ and $\boldsymbol{b} \in \mathbb{R}^n$. In that case, the conditional process $Y \mid A \cdot Y(X) = \boldsymbol{b}$ is still a Gaussian process with mean

$$\eta(\boldsymbol{x}) = \mu(\boldsymbol{x}) + (A\boldsymbol{k}(\boldsymbol{x}))^\top \left(A\mathbb{K}A^\top\right)^{-1}(\boldsymbol{b} - A\boldsymbol{\mu}), \quad \boldsymbol{x} \in \mathbb{R}^d \tag{11}$$

and covariance function

$$\tilde{K}(\boldsymbol{x}, \boldsymbol{x}') = K(\boldsymbol{x}, \boldsymbol{x}') - (A\boldsymbol{k}(\boldsymbol{x}))^\top \left(A\mathbb{K}A^\top\right)^{-1} A\boldsymbol{k}(\boldsymbol{x}'), \quad \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d \tag{12}$$

where $\boldsymbol{\mu} = \mu(X) = \left(\mu(x^{(1)}), \dots, \mu(x^{(m)})\right)^\top \in \mathbb{R}^m$ is the trend vector at the design points, $\mathbb{K}$ is the covariance matrix of $Y(X)$ and $\boldsymbol{k}(\boldsymbol{x}) = \left(K\left(\boldsymbol{x}, x^{(1)}\right), \dots, K\left(\boldsymbol{x}, x^{(m)}\right)\right)^\top$ is the vector of covariance between $Y(\boldsymbol{x})$ and $Y(X)$. Note that when A is a square identity matrix, the linear constraints become interpolation constraints.

3.3 Kriging in the presence of noisy observations

The previous framework can be further extended to situations where market observations cannot be considered to be fully reliable due to, e.g., market microstructure effect. We assume that this uncertainty blurred the market information on f in such a way that the previous measurement equation (10) is flawed by an additive error term ϵ :

$$\boldsymbol{b} = A \cdot f(X) + \epsilon.$$

We consider that this error term is one realization of an independent zero-mean Gaussian noise $\boldsymbol{\varepsilon}$ in $\mathbb{R}^m$ with covariance matrix Σ. If the measurement error $\boldsymbol{\varepsilon}$ is assumed to be independent

of the Gaussian process Y, the conditional process $Y \mid \boldsymbol{b} = A \cdot Y(X) + \boldsymbol{\varepsilon}$ is still a Gaussian process with mean

$$\eta(\boldsymbol{x}) = \mu(\boldsymbol{x}) + (A\boldsymbol{k}(\boldsymbol{x}))^\top \left(A\mathbb{K}A^\top + \Sigma\right)^{-1} (\boldsymbol{b} - A\boldsymbol{\mu}), \quad \boldsymbol{x} \in \mathbb{R}^d \tag{13}$$

and covariance function

$$\tilde{K}(\boldsymbol{x}, \boldsymbol{x}') = K(\boldsymbol{x}, \boldsymbol{x}') - (A\boldsymbol{k}(\boldsymbol{x}))^\top \left(A\mathbb{K}A^\top + \Sigma\right)^{-1} A\boldsymbol{k}(\boldsymbol{x}'), \quad \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d \tag{14}$$

where $\boldsymbol{\mu} = \mu(X) = \left(\mu(x^{(1)}), \ldots, \mu(x^{(m)})\right)^\top \in \mathbb{R}^m$ is the trend vector at the design points, $\mathbb{K}$ is the covariance matrix of $Y(X)$ and $\boldsymbol{k}(\boldsymbol{x}) = \left(K\left(\boldsymbol{x}, x^{(1)}\right), \ldots, K\left(\boldsymbol{x}, x^{(m)}\right)\right)^\top$ is the vector of covariance between $Y(\boldsymbol{x})$ and $Y(X)$. Note that the only difference compared to noiseless kriging equations (11) and (12) is the replacement, at every occurrence, of the covariance matrix $A\mathbb{K}A^\top$ by an inflated matrix $A\mathbb{K}A^\top + \Sigma$. In addition, the curve associated with the kriging mean function (13) does not satisfy the noiseless market fit condition.

The possibility to handle the presence of noisy observations is quite important. The presence of measurement errors and observations with contaminations may have important consequences on the dynamics of yield curves, as noticed for example in Laurini and Ohashi (2015). The previous framework can be used to construct term structure functions in the presence of illiquid securities. Assume that the vector $\boldsymbol{b}$ represents mid-prices of some reference products in a market situation where bid-ask spreads are large. For instance, the covariance matrix Σ can be chosen in such a way that the range of $(\boldsymbol{b} - \boldsymbol{\varepsilon})$ is concentrated on the bid-ask interval. One possible way to do that is to define Σ as a diagonal matrix where each positive term is the square difference between ask and mid prices.

Note moreover that this extension can also be useful when one wants to explicitly look for an attractive trade-off between precision and curve smoothness. Indeed, considering that observed information is uncertain allows to weaken the market-fit condition which in turn benefits to curve smoothness. Our methodology can be then adapted both for best-fitting of noisy market prices and for the construction of an exact interpolatory term structure to a set of liquid instruments.

4 Kriging under additional monotonicity constraints

As mentioned in Section 2, the studied real function f may be known to satisfy some shape-preserving constraints such as monotonicity or positivity. *No-Arbitrage constraints* lead to a theoretical need of monotonicity in curve constructions. For example, quantities such as prices of default-free zero-coupon bond, discount factors or implied survival probabilities shall be non-increasing with respect to time horizons under the no-arbitrage condition. In this section, we propose to extend constrained spline techniques to constrained kriging in order to build term-structure functions.

The use of monotonic splines for term-structure construction has been a subject of great interest in the literature. This has been investigated by among other Barzanti and Corradi (1999), Ramponi (2003), Chiu et al. (2008). The relationship of shape and monotonicity restrictions

with no-arbitrage conditions is detailed for example in the recent literature, see Laurini and Moura (2010) and Fengler and Hin (2015). Additional motivations of the use of monotonicity constraints will be developed in the numerical Section 5.4.

A natural extension of kriging under monotonicity constraints is to consider conditional Gaussian processes with monotone paths. However, the difficulty is that the conditional monotone process is not a Gaussian process any more. We adopt here the approach introduced by Maatouk and Bay (2014b) where Gaussian processes are approximated by finite-dimensional versions, so that the monotonicity constraints can be checked very efficiently in the entire domain.

In the following, we consider linear equality constraints as described in Section 3.2 and we explain how to incorporate supplementary monotonicity constraints. Note that, even if we focus on monotonicity constraints, other shape-preserving constraints can be incorporated using similar ideas (see Maatouk and Bay, 2014b). We first introduce monotonicity constraints under the one-dimensional case, where one has to retrieve a monotonic curve at a given quotation date (Section 4.1). We then explain how to extend the one-dimensional kriging construction method to dimension two, where information at several quotation dates can be jointly used (Section 4.2).

4.1 One dimensional case

In this section, we assume that the input variable x belongs to an interval $D = [\underline{x}, \overline{x}]$ of $\mathbb{R}$ and we consider an original Gaussian process Y with covariance function K. For simplicity, we assume that Y is a zero-mean GP. The aim of this section is to explain how to construct a process that both satisfies linear equality constraints and monotonicity constraints.

The first step is to approximate the original Gaussian process Y by a finite-dimensional version which is monotonic under some simple finite-dimensional linear inequality conditions. We begin by discretizing the input interval D into a regular subdivision $u_0 < \ldots < u_N$ with $u_0 = \underline{x}$, $u_N = \overline{x}$ and with a constant mesh δ, so that $u_j = u_0 + j\delta$, $j = 0, \ldots, N$. We then consider an associated set of basis functions ϕ_j, $j = 0, \ldots, N$ defined as

$$\phi_j(x) = \int_{\underline{x}}^{x} h_j(u)du, \qquad x \in D,$$

where $h_j(x) := \max\left(1 - \frac{|x-u_j|}{\delta}, 0\right)$ is a hat function centered at the j^{th} knot u_j of the input subdivision. Note that the basis functions ϕ_j, $j = 0, \ldots, N$ are increasing as primitives of positive functions. We then define the finite-dimensional approximation of Y on D as the process Y^N such that

$$Y^N(x) = \eta + \sum_{j=0}^{N} \xi_j \phi_j(x), \qquad x \in D, \tag{15}$$

where $\boldsymbol{\xi} = (\eta, \xi_0, \ldots, \xi_N)^\top$ is a zero-mean Gaussian vector. If the Gaussian process Y has almost surely differentiable paths, choosing $\eta = Y(u_0)$ and $\xi_j = Y'(u_j)$, $j = 0, \ldots, N$ guarantees that the finite-dimensional process Y^N uniformly converges on D towards Y almost surely as N tends to infinity (see Maatouk and Bay, 2014b). In that case, the covariance matrix Γ^N of $\boldsymbol{\xi}$ is

given as

$$\Gamma^N = \begin{bmatrix} K(u_0,u_0) & \frac{\partial K}{\partial x'}(u_0,u_j) \\ \\ \frac{\partial K}{\partial x}(u_i,u_0) & \frac{\partial^2 K}{\partial x \partial x'}(u_i,u_j) \end{bmatrix}_{0\leq i,j\leq N}, \tag{16}$$

where K is the covariance matrix of the original GP Y and $u_j,\ j=0,\dots,N$ are the knots of the input subdivision.

Proposition 4.1 (Monotonicity)**.** *The process Y^N defined in Equation (15) is non-increasing (resp. non-decreasing) if and only if all the coefficients $\xi_j,\ j=0,\dots,N$ are nonpositive (resp. nonnegative).*

Proof. If $\xi_j,\ j=0,\dots,N$ are nonpositive then, since ϕ_j are increasing, then Y^N is increasing. For the reverse implication, let us first notice that the derivative of the basis functions $\phi_j,\ j=0,\dots,N$ are such that

$$\phi_j'(u_k) = h_j(u_k) = \mathbb{1}_{j=k} = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if } j\neq k \end{cases}$$

Thus, the derivative of the process Y^N at any knots $u_k,\ k=0,\dots,N$ is

$$\left(Y^N\right)'(u_k) = \sum_{j=0}^{N} \xi_j \phi_j'(u_k) = \xi_k,$$

which concludes the proof. □

The choice of the basis functions ϕ_j and of Γ^N depends on the type of shape-preserving constraints. Other type of basis function can be used for other constraints (for more details, see Maatouk and Bay, 2014b).

In order to construct curves which are compatible with market quotes, linear equality constraints like the one given in Subsection 3.2 have to be imposed on the process Y^N at some points $x^{(1)},\dots,x^{(m)}$ in D. Then, if $Y^N(X) = \left(Y^N\left(x^{(1)}\right),\dots,Y^N\left(x^{(m)}\right)\right)^\top$ denotes the vector of values involved in the curve construction and given Equation (15), the condition $A\cdot Y^N(X) = \boldsymbol{b}$ translates into the following linear equality constraint on the Gaussian vector $\boldsymbol{\xi}$:

$$A\cdot\Phi\cdot\boldsymbol{\xi} = \boldsymbol{b}, \tag{17}$$

where Φ is a $m\times(N+2)$ matrix defined as

$$\Phi_{i,j} := \begin{cases} 1 & \text{for } i=1,\dots,m \text{ and } j=1, \\ \phi_{j-2}\left(x^{(i)}\right) & \text{for } i=1,\dots,m \text{ and } j=2,\dots,N+2. \end{cases}$$

Note that, generally speaking, the linear equality condition (17) on $\boldsymbol{\xi}$ admits solutions only when $N+2\geq n$ as $A\cdot\Phi$ is a matrix of dimension $n\times(N+2)$.

Curve simulation. Conditional GP satisfying both monotonicity and linear equality constraints can be sampled by generating truncated Gaussian vector $\boldsymbol{\xi}$ restricted to :

$$\begin{cases} B \cdot \boldsymbol{\xi} = \boldsymbol{b} & \text{linear equality condition} \\ \boldsymbol{\xi} \in \mathcal{C} & \text{monotonicity constraint} \end{cases}$$

where, $B = A \cdot \Phi$ and for instance $\mathcal{C} = \left\{\boldsymbol{\xi} \in \mathbb{R}^{N+2} \ : \ \xi_j \leq 0, \ j = 0, \ldots, N\right\}$ for non-increasing constraints. Then, simulated paths can be sampled in two steps. First, the conditional distribution of the vector $\boldsymbol{\xi}$ given $B \cdot \boldsymbol{\xi} = \boldsymbol{b}$ is still Gaussian with mean

$$(B\Gamma^N)^\top \left(B\Gamma^N B^\top\right)^{-1} \boldsymbol{b}$$

and covariance matrix

$$\Gamma^N - \left(B\Gamma^N\right)^\top \left(B\Gamma^N B^\top\right)^{-1} B\Gamma^N,$$

so that it can be simulated very efficiently. Then, the simulation of truncated Gaussian vectors restricted to, for instance, negativity of the components (here $\boldsymbol{\xi} \in \mathcal{C}$) can be done by using improved rejection sampling algorithm such as the one described in Maatouk and Bay (2014a) and Robert (1995). By Equation (15) we get sample paths that fulfil both constraints. These simulations can be used to construct confidence intervals for the value of the curve at each point x in D.

Most likely curve. Given a covariance kernel K and its estimated parameters, it is possible to determine the most likely path of the conditional Gaussian process under both linear and monotonicity constraints. This most likely curve corresponds to the mode[7] of the finite-dimensional truncated Gaussian vector $\boldsymbol{\xi}$ (see also Abrahamsen and Benth, 2001, for a discussion about the mode). In bayesian statistics, it is known as the Maximum A Posteriori (MAP) estimator (see Maatouk and Bay, 2014b, for more details). Its expression is given by :

$$M_K^N\left(x \mid A, \boldsymbol{b}\right) = \nu + \sum_{j=0}^{N} \nu_j \phi_j(x), \tag{18}$$

where $\boldsymbol{\nu} = (\nu, \nu_0, \ldots, \nu_N)^\top \in \mathbb{R}^{N+2}$ is the solution of the following convex optimization problem :

$$\boldsymbol{\nu} = \arg \min_{\boldsymbol{c} \in \mathcal{C} \cap \mathcal{I}(A, \boldsymbol{b})} \left(\frac{1}{2} \boldsymbol{c}^\top \left(\Gamma^N\right)^{-1} \boldsymbol{c}\right), \tag{19}$$

and where Γ^N is the covariance matrix of the Gaussian vector $\boldsymbol{\xi}$ defined in (16). The vector $\boldsymbol{\nu}$ can be seen as the mode of the Gaussian vector $\boldsymbol{\xi}$ restricted to $\mathcal{C} \cap \mathcal{I}(A, \boldsymbol{b})$, where $\mathcal{C}$ is the set of vectors satisfying monotonicity constraints and $\mathcal{I}(A, \boldsymbol{b}) = \left\{\boldsymbol{\xi} \in \mathbb{R}^{N+2} \ : \ A \cdot \Phi \cdot \boldsymbol{\xi} = \boldsymbol{b}\right\}$ is the set of vectors which are compatible with the linear equality constraint. Obviously, this curve satisfies both constraints. Additionally, it does not depend on the variance σ^2 of the covariance function K since σ^2 is a multiplicative constant in the matrix Γ^N and then does not affect the arg min in Equation (19). In Bay et al. (2015) and Bay et al. (2016), the convergence of the proposed estimator (18) as N tends to infinity is studied and its limit corresponds to a constrained spline function (which depends on the underlying kernel function). By this methodology, one can thus retrieve classical spline interpolation with the additional possibility of getting confidence intervals.

[7] The maximum of the probability density function.

4.2 Two dimensional case

As explained in Remark 2.1, the term-structure construction can be stated in a two-dimensional setting. The aim is to incorporate in the curve construction process market information at several quotation dates. Contrary to the previous section, the output is now a surface which may represent the evolution of a reference quantity with respect to time-to-maturities and quotation dates. More formally, we consider a two-dimensional input variable $\boldsymbol{x} = (x, t)$ which is assumed to belong to a rectangle $D = [\underline{x}, \overline{x}] \times [\underline{t}, \overline{t}]$ of $\mathbb{R}^2$. The variable x may represent time-to-maturities whereas the variable t may represent evolution of time or quotation dates. For the sake of simplicity, we assume that the unknown bivariate real function f is monotone, say non-increasing, with respect to the first input variable only :

$$x_a \leq x_b \quad \Rightarrow \quad f(x_b, t) \leq f(x_a, t), \qquad \text{for all } t \in [\underline{t}, \overline{t}],\ x_a, x_b \in [\underline{x}, \overline{x}]. \tag{20}$$

More general inequality conditions can be considered in this bivariate setting (see Maatouk and Bay, 2014b, for more details). The aim of this section is to explain how to construct a process that simultaneously satisfies a series of linear equality constraints (one for each considered quotation date) and a monotonicity constraint as described in (20). As in the one-dimensional setting, we start with an original bivariate Gaussian process Y with zero-mean and with a covariance function K.

The idea is the same as the one dimensional case presented in Section 4.1. We begin by discretizing the input rectangle D in a $(N_x + 1) \times (N_t + 1)$ grid which for simplicity is assumed to be regular. The subdivision of the x-axis is $u_0 < \ldots < u_{N_x}$ with $u_0 = \underline{x}$, $u_{N_x} = \overline{x}$ and a constant mesh δ_x, so that $u_i = u_0 + i\delta_x$, $i = 0, \ldots, N_x$. The subdivision of the y-axis is $v_0 < \ldots < v_{N_t}$ with $v_0 = \underline{t}$, $v_{N_t} = \overline{t}$ and a constant mesh δ_t, so that $v_j = v_0 + j\delta_t$, $j = 0, \ldots, N_t$. The following developments can be easily extended to irregular grids. As in the one-dimensional case, we consider an original Gaussian process Y with covariance function K. We then define the finite-dimensional approximation of Y as the process Y^N such that

$$Y^N(x, t) = \sum_{i=0}^{N_x} \sum_{j=0}^{N_t} \xi_{i,j} g_i(x) h_j(t), \qquad \text{for all } (x, t) \in D, \tag{21}$$

where $g_i(x) = \max\left(1 - \frac{|x-u_i|}{\delta_x}, 0\right)$ and $h_j(t) = \max\left(1 - \frac{|t-v_j|}{\delta_t}, 0\right)$ are hat functions centered at the knots u_i and v_j, for $i = 0, \ldots, N_x$ and $j = 0, \ldots, N_t$ and $\boldsymbol{\xi} = (\xi_{0,0}, \xi_{0,1}, \ldots, \xi_{i,j}, \ldots, \xi_{N_x,N_t})^\top$ is a zero-mean Gaussian vector with components $(\boldsymbol{\xi})_{\rho_{ij}} = \xi_{i,j}$ where $\rho_{ij} = (N_t + 1)i + j + 1$, $i = 0, \ldots, N_x$ and $j = 0, \ldots, N_t$. Let $N_{tot} = (N_x + 1)(N_t + 1)$ be the size of the column vector $\boldsymbol{\xi}$. Choosing $\xi_{i,j} = Y(u_i, v_j)$, $i = 0, \ldots, N_x$ and $j = 0, \ldots, N_t$ guarantees that the finite-dimensional process Y^N converges on D towards Y almost surely as N_x and N_t tends to infinity (see Maatouk and Bay, 2014b). In that case, the covariance matrix $\Gamma^N \in \mathbb{R}^{N_{tot}^2}$ of the Gaussian vector $\boldsymbol{\xi}$ can be written as :

$$\Gamma^N_{\rho_{ij},\rho_{i'j'}} = \mathrm{Cov}(\xi_{i,j}, \xi_{i',j'}) = K\left((u_i, v_j), (u_{i'}, v_{j'})\right),$$

where $i, i' = 0, \ldots, N_x$ and $j, j' = 0, \ldots, N_t$. As can be shown in the next proposition, the monotonicity constraint (20) reduces to a linear inequality condition on the vector $\boldsymbol{\xi}$.

Proposition 4.2 (Monotonicity). *The process Y^N defined in Equation (21) is non-increasing (resp. non-decreasing) with respect to the first variable x if and only if all the coefficients ξ_j, $j = 0, \dots, N$ are such that*

$$\xi_{i-1,j} \leq \xi_{i,j}, \ (\textit{resp. } \xi_{i-1,j} \geq \xi_{i,j}) \qquad i = 1, \dots, N_x \textit{ and } j = 0, \dots, N_t. \tag{22}$$

Proof. The proof is similar to the one of Proposition 4.2. □

In order to construct surfaces which are compatible with market quotes observed at times $t = t_1, \dots, t_I$, I different linear equality constraints like the one given in Section 3.2 has to be imposed on the bivariate process Y^N. Let us consider that the market-fit conditions involve m points $x^{(1)}, \dots, x^{(m)}$ for any quotation time $t = t_1, \dots, t_I$. Then, if $Y^N(X,t) = \left(Y^N\left(x^{(1)},t\right), \dots, \ Y^N\left(x^{(m)},t\right)\right)^\top$ denotes the vector of values involves in the curve construction at time t and given Equation (21), the condition $A_t \cdot Y^N(X,t) = \boldsymbol{b}_t$ translates into the following linear equality constraint on the Gaussian vector $\boldsymbol{\xi}$:

$$A_t \cdot H_t \cdot \boldsymbol{\xi} = \boldsymbol{b}_t \tag{23}$$

where the $m \times N_{tot}$ matrix H_t has components $(H_t)_{k,\rho_{ij}} = g_i\left(x^{(k)}\right) h_j(t)$, $k = 1, \dots, m$ and $\rho_{ij} = (N_t+1)i + j + 1$, $i = 0, \dots, N_x$ and $j = 0, \dots, N_t$. The former condition has to hold simultaneously for every time $t = t_1, \dots, t_I$, which can be summarized as one single linear equality constraint

$$B \cdot \boldsymbol{\xi} = \boldsymbol{b}, \tag{24}$$

where the $nI \times N_{tot}$ matrix B (resp. $\boldsymbol{b}$) is formed by vertical concatenation of matrices $A_t \cdot H_t$ (resp. $\boldsymbol{b}_t$) for $t = t_1, \dots, t_I$. Notice that, generally speaking, this linear system admits solutions when $N_{tot} \geq nI$.

Surface simulation. As in the one-dimensional setting, the simulation of the bivariate Gaussian process Y^N conditionally to both linear equality and monotonicity constraints reduces to the simulation of the Gaussian vector $\boldsymbol{\xi}$ restricted to

$$\begin{cases} B \cdot \boldsymbol{\xi} = \boldsymbol{b} & \text{linear equality condition} \\ \boldsymbol{\xi} \in \mathcal{C} & \text{monotonicity constraint} \end{cases}$$

where, for instance, $\mathcal{C} = \left\{\boldsymbol{\xi} \in \mathbb{R}^{N_{tot}} \ : \ \xi_{i-1,j} \leq \xi_{i,j}\right\}$ for a non-increasing constraint. The simulation procedure is the same as the one presented in Section 4.1. These simulations can be used to construct confidence intervals for the surface value at each point (x,t) in D.

Most likely surface. The methodology can be easily extended from Section 4.1.

4.3 Parameters estimation

The most likely path of the constrained process (mode estimator) depends on the choice of the underlying Gaussian process or equivalently on its covariance function K. In this section, we investigate the estimation of its parameter, i.e., the length and the variance hyper parameters. In the literature, estimation of the covariance function hyper parameters is usually done using two types of methods : Maximum Likelihood (ML) estimators as in Santner et al. (2003) and

Cross Validation (CV) methods as in Bachoc (2013), Cressie (1993) and Roustant et al. (2012). Both methods ML and CV are not suited to monotonicity constraints.

Recently, an Adapted Cross-Validation (ACV) technique has be proposed by Maatouk et al. (2015) to estimate covariance hyper-parameters of Gaussian processes in the presence of inequality constraints. The main idea is to consider the mode curve (as opposed to the mean curve) as the estimator to be used in the cross-validation method. The principle of cross-validation is to select the set of parameters that minimizes a distance between observed values and their estimates while successively omitting some set of observations. Thus, cross-validation is similar in spirit to backtesting (see Kerkhof and Melenberg, 2004), but omitted data are not necessarily taken in chronological order.

Length parameters. In classical kriging, the usual cross-validation estimator of the covariance length parameter $\boldsymbol{\theta}$ is constructed from the so-called Leave One Out (LOO) mean square error criterion. Given that the unknown function f takes values $y_1, \ldots, y_n$ at points $x^{(1)}, \ldots, x^{(n)}$ (pure interpolation constraints), the cross-validation estimator $\hat{\boldsymbol{\theta}}_{CV}$ of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_{CV} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \left(y_i - \hat{y}_{i,\boldsymbol{\theta}}(\boldsymbol{y}_{-i})\right)^2, \tag{25}$$

where $\boldsymbol{y}_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)^\top$ and Θ is a compact subset of $\mathbb{R}^d$. The estimator $\hat{y}_{i,\boldsymbol{\theta}}(\boldsymbol{y}_{-i}) = \mathbb{E}\left(Y(x^{(i)}) \,\middle|\, Y(X^{(-i)}) = \boldsymbol{y}_{-i}\right)$ is the kriging mean at point $x^{(i)}$ obtained by removing observation y_i in the estimation process (see Equation (8)). The vector $Y\left(X^{(-i)}\right)$ is the same vector as $Y(X)$ without component $Y\left(x^{(i)}\right)$.

The previous criterion cannot be used in the presence of monotonicity constraints since the classical kriging mean estimate does not respect such kind of constraints. As suggested in Maatouk et al. (2015) under pure interpolation constraints, the most likely curve (mode estimator) defined in Section 4.1 can be used instead of the kriging mean since the former satisfies the monotonicity constraints. Then, a new LOO criterion adapted to monotonicity constraints can be defined as :

$$\hat{\boldsymbol{\theta}}_{ACV} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \left(y_i - M_K^N\left(x^{(i)} \,\middle|\, I_{n-1}, \boldsymbol{y}_{-i}\right)\right)^2, \tag{26}$$

where $M_K^N\left(x^{(i)} \,\middle|\, I_{n-1}, \boldsymbol{y}_{-i}\right)$ is the mode estimator (defined in Equation (18)) of y_i based on all observations but y_i. As explained in Section 4.1, the latter does not depend on the variance σ^2. The matrix I_{n-1} is the identity matrix with dimension $n-1$ (the same as the vertical dimension as the column vector $\boldsymbol{y}_{-i}$). This methodology cannot be applied immediately in our setting since the output values y_i in the LOO criterion (26) are not available under linear equality constraints as defined in (17). To this end, we consider the following formulation of the LOO criterion :

$$\hat{\boldsymbol{\theta}}_{ACV} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \left(b_i - \left(A \cdot M_K^N\left(X \mid A_{-i}, \boldsymbol{b}_{-i}\right)\right)_i\right)^2, \tag{27}$$

where the subscript i refers to the i-th component, A_{-i} and $\boldsymbol{b}_{-i}$ are respectively the matrix and the vector without the i^{th} row.

Variance parameter. In the case of pure interpolation constraints and without monotonicity constraints, several classical criterions can be used for the estimation of σ. For instance, in Bachoc (2013), an estimator $\hat{\sigma}_{CV}$ is defined as the parameter σ such that the following equality holds

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left(y_i-\hat{y}_{i,\hat{\boldsymbol{\theta}}_{CV}}(\boldsymbol{y}_{-i})\right)^2}{\mathbb{E}\left(\left(Y\left(x^{(i)}\right)-\hat{y}_{i,\hat{\boldsymbol{\theta}}_{CV}}(\boldsymbol{y}_{-i})\right)^2\,\middle|\,Y(X^{(-i)})=\boldsymbol{y}_{-i}\right)}=1. \tag{28}$$

Given our linear equality conditions and monotonicity constraints, we instead propose to define $\hat{\sigma}_{ACV}$ as the parameter σ such that the following equality holds

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\left(b_i-\left(A\cdot M_K^N\left(X\mid A_{-i},\boldsymbol{b}_{-i}\right)\right)_i\right)^2}{\mathbb{E}\left(\left(AY\left(X\right)-AM_K^N\left(X\mid A_{-i},\boldsymbol{b}_{-i}\right)\right)_i^2\,\middle|\,\mathcal{D}_i\right)}=1, \tag{29}$$

where $\mathcal{D}_i$ is the set of monotonicity constraints on the whole domain and the additional linear constraints without the i^{th} one, i.e. $A_{-i}Y(X^{(-i)})=\boldsymbol{b}_{-i}$ where A_{-i} is the matrix A without the i-th row. The expectation is estimated by simulation, approximating the process Y by Y^N.

5 Empirical investigation

In this section, the construction method developed in Section 4 is illustrated in different financial applications. Based on market quotes observed at different dates, we construct curves together with confidence intervals for term-structures of OIS discount rates, term-structure of zero-coupon swaps rates and term-structure of CDS implied default probabilities. Using the bivariate setting of Section 4.2, we also build interest-rate and default probability surfaces by considering time (quotation dates) as an additional dimension.

5.1 Interest-rate curves based on Swaps versus Euribor

We apply the kriging procedure to construct zero-coupon swap curves based on market quotes of fixed-vs-floating interest-rate swaps for different standard maturities. Market observations are given as par rates of Swaps vs Euribor 6M. We consider the 10 quotation dates given in Table 2. For each quotation date, the term-structure is built on 14 swap rates $S_1,\ldots,S_{14}$ associated with standard maturities in years belonging to the set $E:=\{1,\ldots,10,15,20,30,40\}^\top$. As explained in Section 2, each observed par rate provides an indirect information on the curve. This information takes the form of a linear relation given by (3). For each standard maturity $T\in E$, this relation involves the value $P(t,k)$ of the curve at time horizon $k=1,\ldots,T$. As a result, the curves are compatible with observed quotes if, for each observation date t, the vector of discount factors $P(t,X):=(P(t,1),\ldots,P(t,40))^\top$ satisfies a linear system of the form

$$A_t\cdot P(t,X)=\boldsymbol{b}_t, \tag{30}$$

where A_t is a 14×40 real matrix and $\boldsymbol{b}_t=(1,\ldots,1)^\top\in\mathbb{R}^{14}$. In this case, we have $n=14$ products whose value depends on $m=40$ points of the curve.

We consider that the associated discount factor curve $P(t,X)$ is one realization of a decreasing spatial process which starts from 1 ($P(t,0)=1$) and satisfies the linear condition (30).

Section 4 explains how to construct and simulate a process with monotonicity and linear equality constraints. This construction involves a finite-dimensional approximation Y^N of Gaussian processes as defined in (15). The latter depends on a $N+2$-dimensional Gaussian vector $\boldsymbol{\xi}$ and a set of basis functions ϕ defined on a subdivision of the input domain $D = [0, 40]$. We consider here a regular subdivision $u_j = u_0 + j\delta$, $j = 0, \ldots, N$, where $u_0 = 0$, $N = 50$ sub-intervals and $\delta = 1$. Since the curve $T \to P(t, T)$ is known to start from 1, i.e., $P(t, 0) = 1$, the linear equality condition on $\boldsymbol{\xi}$ defined in (17) can be reformulated as follows :

$$\begin{pmatrix} 1 & \phi_0(0) & \ldots & \phi_N(0) \\ & & A_t \cdot \Phi & \end{pmatrix} \boldsymbol{\xi} = \begin{pmatrix} 1 \\ \boldsymbol{b}_t \end{pmatrix}. \tag{31}$$

Then the simulation of the GP Y^N conditionally to the linear equality condition and non-increasing constraints is equivalent to generating a truncated zero-mean Gaussian vector $\boldsymbol{\xi}$ restricted to Equation (31) and to non-positive components.

Parameters estimation. We first consider the correlation-length hyper parameter θ. It has been estimated for each quotation dates in Table 2 by using the adapted cross-validation (ACV) method described in Section 4.3. We consider covariance functions of the form $K(x, x') = \sigma^2 C(x - x', \theta)$ and we discuss two alternative kernels C, that is the Gaussian and the Matérn 5/2 kernel (see Table 1). We denote by $\hat{\theta}_G$ (resp. $\hat{\theta}_M$) the estimated length parameter of the Gaussian (resp. Matérn 5/2) covariance function. As shown in Table 2, the estimated length parameter $\hat{\theta}$ remains stable for both Gaussian and Matérn 5/2 covariance functions (value around 25 for $\hat{\theta}_G$ and 30 for $\hat{\theta}_M$). Note also that the minimal value of the objective function in the LOO criterion (27) is slightly smaller when using a Gaussian covariance function. However, as can be seen in Figure 2, the objective function (cross-validation error) reach 0.12 for the Gaussian covariance function whereas it never goes above 0.02 for the Matérn 5/2 covariance function. Additionally, the global minimum is easier to find in the Matérn 5/2 case.

Table 2: Parameters estimation using ACV methods (Swap versus Euribor 6M).

Date	$\hat{\theta}_G$	$\hat{\theta}_M$	Gaussian optimal value	Matérn 5/2 optimal value
02/06/2010	25.8	30.8	4.0e-06	**1.1e-06**
05/07/2010	26.2	25.2	**3.8e-06**	1.1e-05
03/08/2010	26.6	27.0	**4.1e-06**	4.7e-06
29/11/2010	23.5	39.9	**5.4e-07**	2.3e-05
30/12/2010	27.5	28.0	4.5e-06	**1.5e-06**
31/01/2011	29.2	29.2	9.2e-06	**1.2e-06**
10/05/2011	28.8	27.6	2.5e-06	**1.2e-06**
10/06/2011	26.0	30.7	**5.0e-07**	2.8e-06
30/12/2011	20.3	30.0	6.8e-06	**3.6e-06**

Once the length parameter θ has been estimated, the standard deviation parameter σ is estimated using Equation (29). For instance, at quotation date 30/12/2011, we obtain for respective Gaussian and Matérn 5/2 covariance kernel $\hat{\sigma}_G = 2.89$ and $\hat{\sigma}_M = 0.93$.

Curve construction at a single quotation date. We now illustrate the curve construction method described in Section 4.1 in a one-dimensional setting. The construction is based on

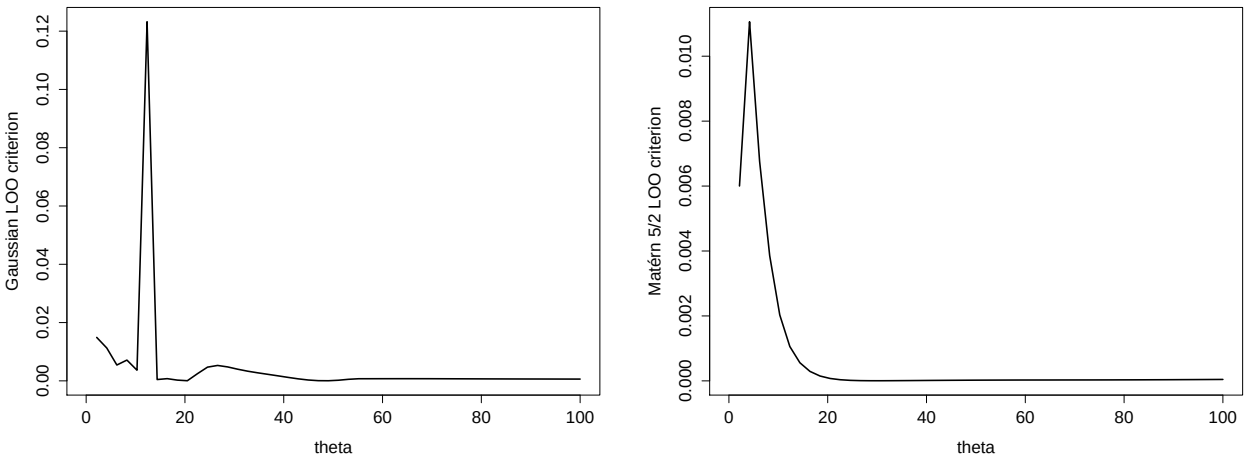

Figure 2: The function to be optimized in the LOO criterion (27) using the Gaussian (left) and the Matérn 5/2 covariance function (right). Swap versus Euribor 6M on 30/12/2011.

market quotes as of 30/12/2011. For this particular date, the estimated length parameter for the Gaussian and the Matérn 5/2 kernel are given in Table 2 ($\hat{\theta}_G = 20.3$ and $\hat{\theta}_M = 30.0$). In that case, using the method described in Section 4.3, the estimated variance parameter is equal to $\hat{\sigma}_G = 2.89$ for the Gaussian kernel and to $\hat{\sigma}_M = 0.93$ for the Matérn 5/2 kernel. Figure 3 compares the sample paths of discount factors for the Gaussian and the Matérn 5/2 covariance function using the corresponding estimated parameters. In both cases, we generate 100 sample paths taken from model (15) conditionally to linear equality constraints (31) and non-increasing constraints. Note that the simulated curves (gray lines) are non-increasing in the entire domain. Additionally, the black solid line represents the most likely curve, i.e., the mode of the conditional GP. Recall that, by construction, this curve satisfies the given constraints. The black dashed-lines represent the 95% point-wise confidence intervals quantified by simulation. Figure 4 and Figure 5 give the corresponding spot rates and instantaneous forward curves.

In order to compare our results with some models commonly used in most central banks, all figures are given together with the associated best-fitted Nelson-Siegel curves (see Nelson and Siegel, 1987) and the associated best-fitted Svensson curves (see Svensson, 1994). Parameters have been estimated by minimizing the sum of squared errors between market and model prices. We use a gradient descent algorithm with randomly chosen starting values as described in Gilli et al. (2010). The optimal parameters are given in Table 3. Discount factors and forward rates have been deduced from Nelson-Siegel and Svensson yield curves.

Table 3: Parameters estimation for Nelson-Siegel and Nelson-Siegel-Svensson model (Swap versus Euribor 6M on 30/12/2011).

	λ_1	λ_2	β_1	β_2	β_3	β_4
Nelson-Siegel	7.4615	-	0.0189	-0.0160	0.0487	-
Nelson-Siegel-Svensson	4.0486	28.4285	0.1719	-0.1590	-0.1101	-0.4093

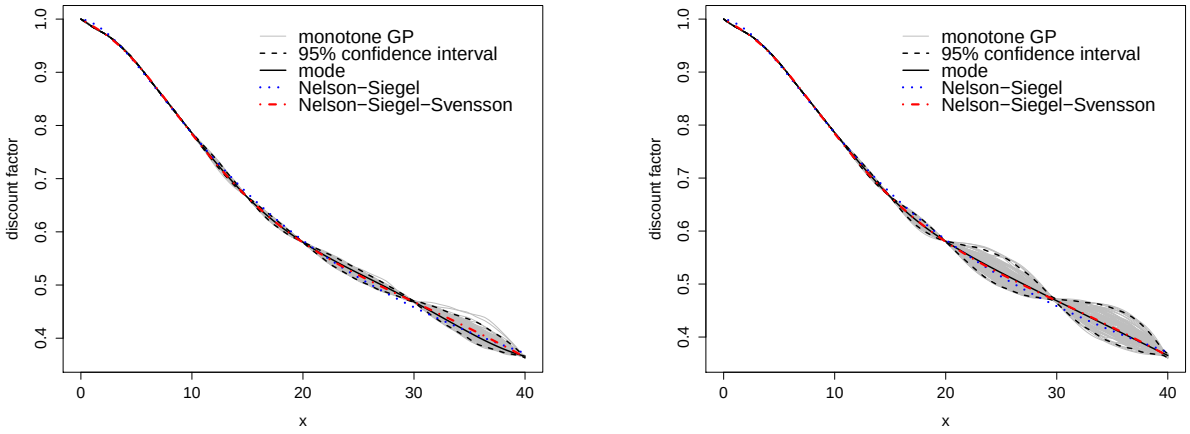

Figure 3: Simulated paths (gray lines) taken from the conditional GP with non-decreasing constraints and market-fit constraints using the Gaussian covariance function with nugget equal to 10^{-5} (left) and the Matérn 5/2 covariance function without nugget (right). Swap vs Euribor 6M market quotes as of 30/12/2011.

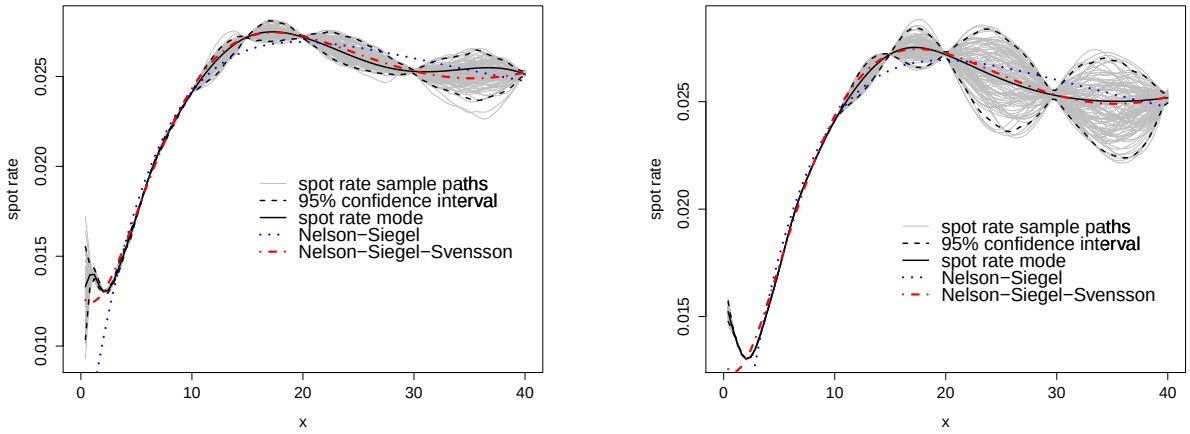

Figure 4: Spot rates obtained from sample paths of Figure 3 with Gaussian covariance function (left) and Matérn 5/2 covariance function (right). Gray lines represent $-\frac{1}{x}\log Y^N(x)$ for each sample path. The black solid line is the most likely spot rate curve $-\frac{1}{x}\log M_K^N\left(x \mid A, \boldsymbol{b}\right)$.

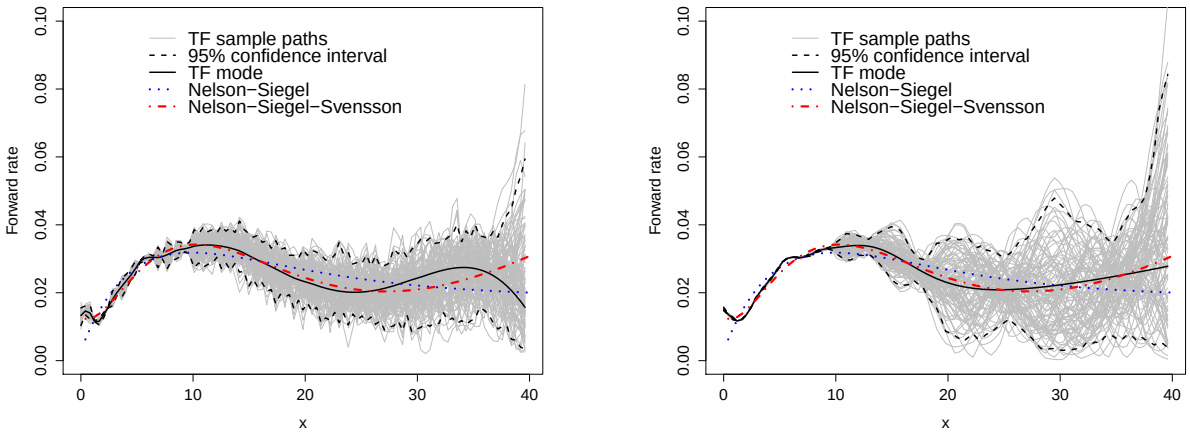

Figure 5: Forward rates obtained from sample paths of Figure 3 with Gaussian covariance function (left) and Matérn 5/2 covariance function (right). Gray lines represent $-\frac{d}{dx}\log Y^N(x)$ for each sample path. The black solid line is the most likely forward rate curve $-\frac{d}{dx}\log M_K^N\left(x \mid A, \boldsymbol{b}\right)$.

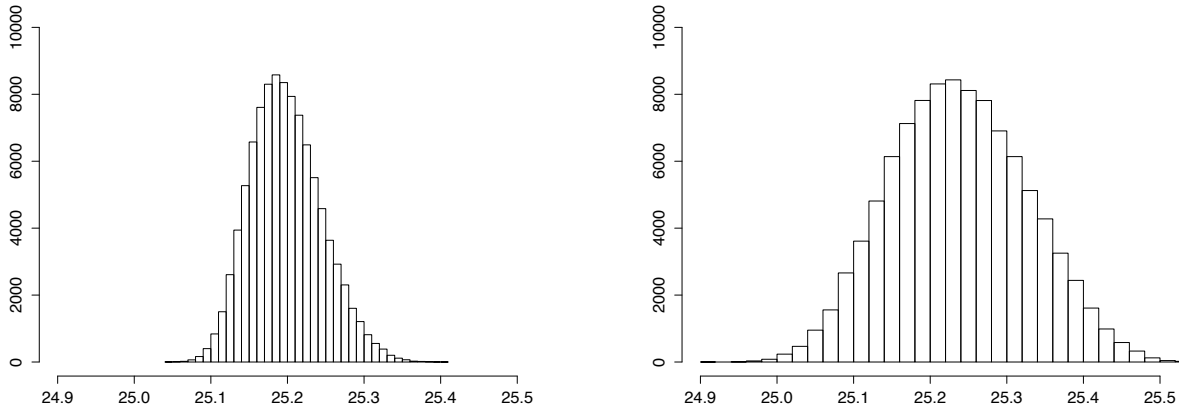

Figure 6: Histogram of a periodic annuity-due present values constructed from 100 000 independent simulations of discount factor curves under a Gaussian covariance function (left) and under a Matérn covariance function (right).

It can be seen on Figures 3, 4 and 5 that the Nelson-Siegel model and the Nelson-Siegel-Svensson model are close to the monotonic kriging mode, this is especially the case for the Svensson extension. However, these models do not always fulfil the given market constraints (especially for the Nelson-Siegel model). Furthermore, they do not provide confidence intervals.

Remark 5.1. *As can be seen on Figure 5, the shape of the most probable forward curves (black solid lines) differs under the two considered covariance kernels. When using a Gaussian covariance kernel, the mode curve features an extraneous hump for maturities greater than 30 years whereas, in the Matérn 5/2 case, it is slightly increasing and follows the fitted Nelson-Siegel-Svensson curve. This difference comes from the fact that $\mathcal{C}^\infty$ forward curves generated by the Gaussian kernel are less flexible to adjust to local tendency of the data than $\mathcal{C}^1$ forward curves generated by the Matérn 5/2 kernel. Note moreover that, in the Gaussian kernel case, inverting the covariance matrix without a nugget effect may involve numerical instabilities due to the restriction of infinite differentiability. This issue has also been reported in Golchi et al. (2015), where the authors state in particular that "this choice of the Matérn covariance function over the commonly used squared exponential family (Sacks et al., 1989) avoids numerical instability, often observed when inverting the covariance matrix, by removing the restriction of infinite differentiability".*

The previous simulations can be used to estimate the distribution of other financial assets whose values depend on the curve. As an example, in Figure 6, we plot an histogram of the present value $\ddot{a}_{\overline{n}|}^{(p)} = \sum_{k=0}^{pn-1} \frac{1}{p} Y^N(k/p)$ of a periodic annuity-due using 100 000 simulations of discount factors sample paths, where $n = 40$ and $p = 12$ (monthly payments). Notice that, despite the variability of the simulated sample paths of the conditional Gaussian process, the present value of the periodic annuity-due remains stable with 95% confidence intervals $[25.12, 25.30]$ using the Gaussian kernel and $[25.07, 25.41]$ using the Matérn kernel.

Several quotation dates. We now illustrate the building procedure in dimension two, when data observed at different quotation dates are incorporated. We then construct a surface representing the evolution of discount factors as a function of time-to-maturities and quotation dates. To do this, we use the approach described in Section 4.2. In Figure 7, the surface represents the mode estimator of the conditional GP in dimension two. The construction relies on the swap quotations at the 9 dates given in Table 2. We choose $N_x = 40$ and $N_t = 20$ and we consider a two-dimensional Gaussian kernel written as

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{(x-x')^2}{2\theta_1} - \frac{(t-t')^2}{2\theta_2}\right),$$

where $\boldsymbol{x} = (x, t)$ and $\boldsymbol{x}' = (x', t')$. For each vector $\boldsymbol{x} = (x, t)$, the first component x represents a time-to-maturity and the second component t represents a quotation date. Without loss of generality, the distance $t' - t$ between two quotation dates has been expressed in percentage of the length between the two extreme dates of the sample. The parameters θ_1 and θ_2 are fixed respectively to 25 and 0.5. Notice that the constructed discount factor surface is non-increasing with respect to time-to-maturities.

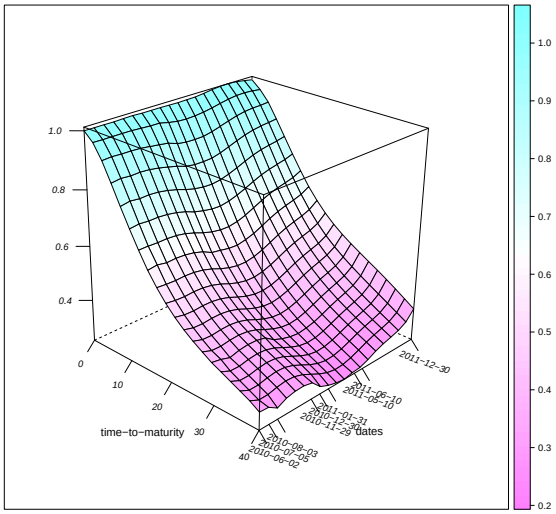

Figure 7: Swap-vs-Euribor discount factors as a function of time-to-maturities and quotation dates.

5.2 OIS discount curves

We now apply the kriging method to construct OIS discount curves. The aim is to construct, at different quotation dates t, discount curves $T \to P^D(t,T)$ based on market quotes of Overnight Indexed Swaps associated with different standard maturities. We consider par rates of OIS at the 10 quotation dates given in Table 4. For each of these quotation dates, the term-structure is built from 14 swap rates $S_1, \ldots, S_{14}$ associated with standard maturities in the set $E := \{1, \ldots, 10, 15, 20, 30, 40\}^\top$. For each standard maturity $T \in E$, the value $P(t,k)$ of the curve at time horizons $k = 1, \ldots, T$ are linked through the linear relation (3). Then, the curve is compatible with market quotes if the vector of discount factors $P^D(t,X) := (P^D(t,1), \ldots, P^D(t,40))^\top$ satisfies a linear system of the form

$$A_t \cdot P(t,X) = \boldsymbol{b}_t, \tag{32}$$

where A_t is a 14×40 real matrix and $\boldsymbol{b}_t = (1, \ldots, 1)^\top \in \mathbb{R}^{14}$. In this case, we have $n = 14$ observations which depends on $m = 40$ points of the curve. Note that the market fit condition has exactly the same form as for the previous example in Subsection 5.1.

Parameters estimation. In Table 4, we estimate the length hyper parameter $\hat{\theta}_G$ and $\hat{\theta}_M$ associated respectively to Gaussian and Matérn 5/2 covariance functions (see Table 1). The optimal values in the last two columns of Table 4 correspond to the values of the LOO criterion defined in (26) at the global optimum. Note that the estimated parameters $\hat{\theta}_G$ and $\hat{\theta}_M$ remain stable across the considered quotation dates. In addition, the obtained optimal values are close for the two covariance functions, even if they are slightly smaller for the Gaussian covariance function. Figure 8 represents the function to be optimized in criterion (26) using the OIS data on 03/06/2010. Given the shape of the functions, the estimation procedure has turned to be much more straightforward using a Matérn 5/2 covariance function.

As previously, once the length parameter θ has been estimated, the standard deviation parameter σ is estimated using Equation (29).

Table 4: Parameters estimation using ACV methods (OIS data).

Date	$\hat{\theta}_G$	$\hat{\theta}_M$	Optimal value Gaussian	Optimal value Matérn 5/2
03/06/2010	26.2	19.1	**2.5e-05**	9.5e-05
04/10/2010	27.8	20.6	**1.2e-05**	7.4e-05
31/12/2010	26.1	18.7	**2.6e-05**	8.4e-05
04/03/2011	28.2	19.0	**1.5e-05**	4.7e-05
15/06/2011	27.3	18.2	**1.2e-05**	6.9e-05
10/10/2011	26.5	23.8	**1.3e-05**	5.3e-05
14/11/2011	26.1	23.8	**1.4e-05**	7.4e-05
15/12/2011	25.8	24.2	**1.9e-05**	7.6e-05

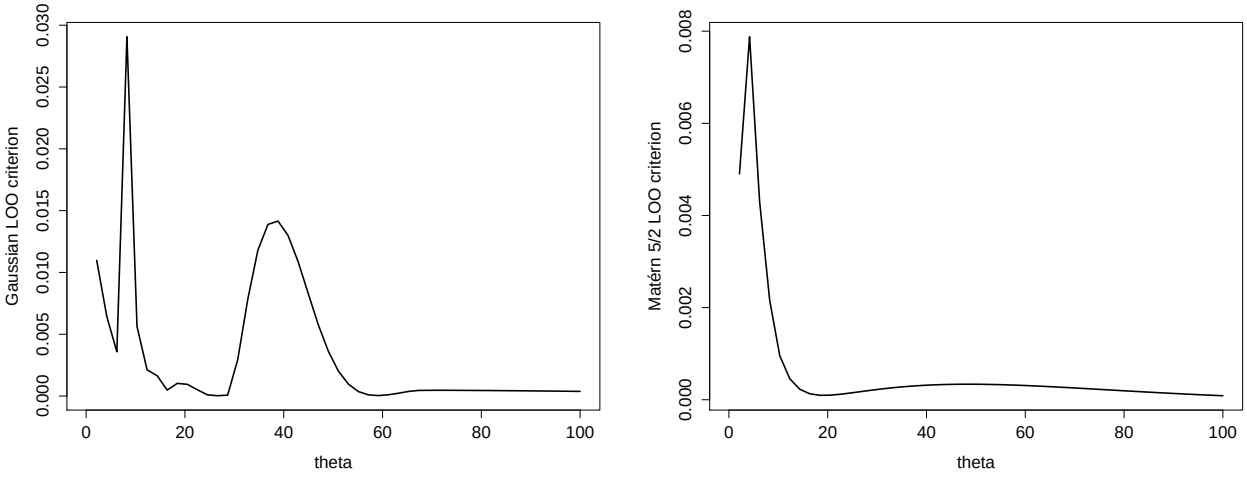

Figure 8: The function to be optimized in LOO criterion (27) using the Gaussian covariance function (left) and the Matérn 5/2 covariance function (right). OIS data on 03/06/2010.

One single quotation date. In Figure 9, we choose $N = 50$ and generate 100 sample paths of discount factors constructed from model (15) when using a Gaussian covariance function (left graph) and a Matérn 5/2 covariance function (right graph). All the curves are non-increasing with respect to time-to-maturities. In addition, they all are perfectly compatible with OIS data as of 03/06/2010. The Gaussian process hyper-parameters have been estimated by the ACV method described in Section 4.3. The estimated hyper-parameters are given by $(\hat{\theta}_G, \hat{\sigma}_G) = (26.2, 4.24)$ when using a Gaussian covariance function and by $(\hat{\theta}_M, \hat{\sigma}_M) = (19.1, 0.24)$ when using a Matérn 5/2 covariance function. The black solid line represents the most likely curve, i.e., the mode of the conditional GP. Recall that, by construction, this curve satisfies the given constraints. The black dashed-lines represent the 95% point-wise confidence intervals quantified by simulation. Figure 10 and Figure 11 gives the corresponding spot rate and forward curves.

Again, all figures 8, 10 and 11 are given together with the associated best-fitted Nelson-Siegel curves (see Nelson and Siegel, 1987) and the associated best-fitted Svensson curves (see Svensson, 1994). Parameters have been estimated by minimizing the sum of squared errors between market and model prices. We use a gradient descent algorithm with randomly chosen starting values as described in Gilli et al. (2010). The optimal parameters are given in Table 5.

Table 5: Parameters estimation for Nelson-Siegel and Nelson-Siegel-Svensson model (OIS data, 03/06/2010).

	λ_1	λ_2	β_1	β_2	β_3	β_4
Nelson-Siegel	1.0890	-	0.0341	-0.0171	-0.0601	-
Nelson-Siegel-Svensson	1.0938	15.1891	0.0502	-0.0164	-0.1074	-0.0494

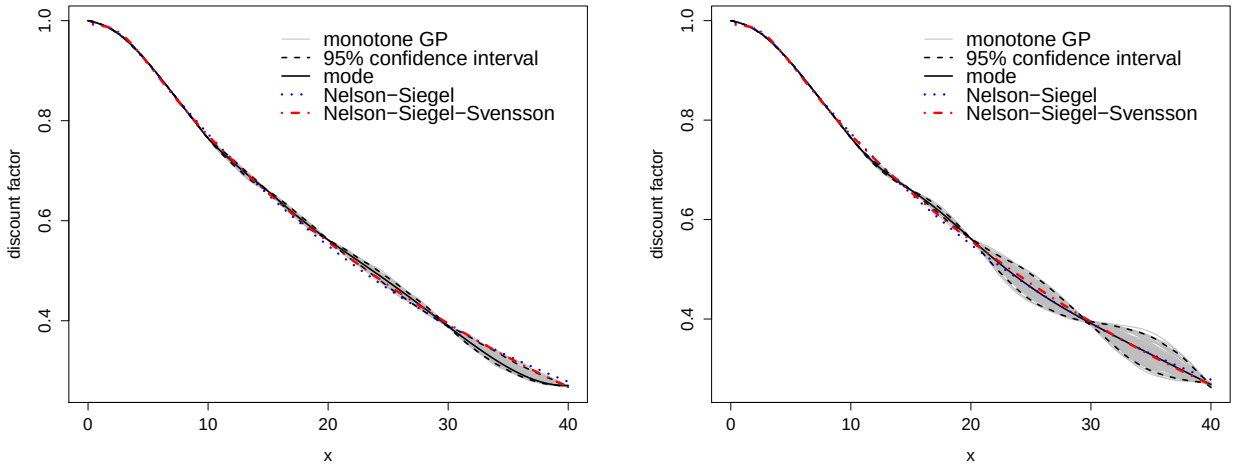

Figure 9: OIS discount factor curves (gray lines) given as simulated paths of a conditional GP with non-increasing constraints using a Gaussian covariance function with nugget equal to 10^{-5} (left) and a Matérn 5/2 covariance function without nugget (right). OIS data of 03/06/2010.

Regarding comparison of curves constructed by the two considered covariance kernels, Remark 5.1 also applies here.

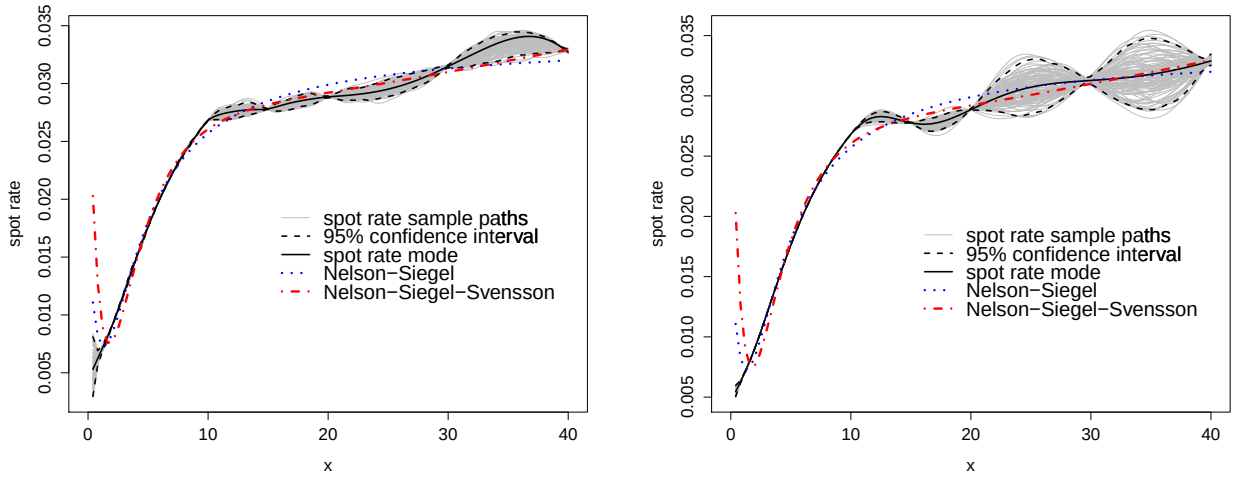

Figure 10: Spot rates obtained from sample paths of Figure 9 with Gaussian covariance function (left) and Matérn 5/2 covariance function (right). Gray lines represent $-\frac{1}{x}\log Y^N(x)$ for each sample path. The black solid line is the most likely spot rate curve $-\frac{1}{x}\log M_K^N\left(x \mid A, \boldsymbol{b}\right)$.

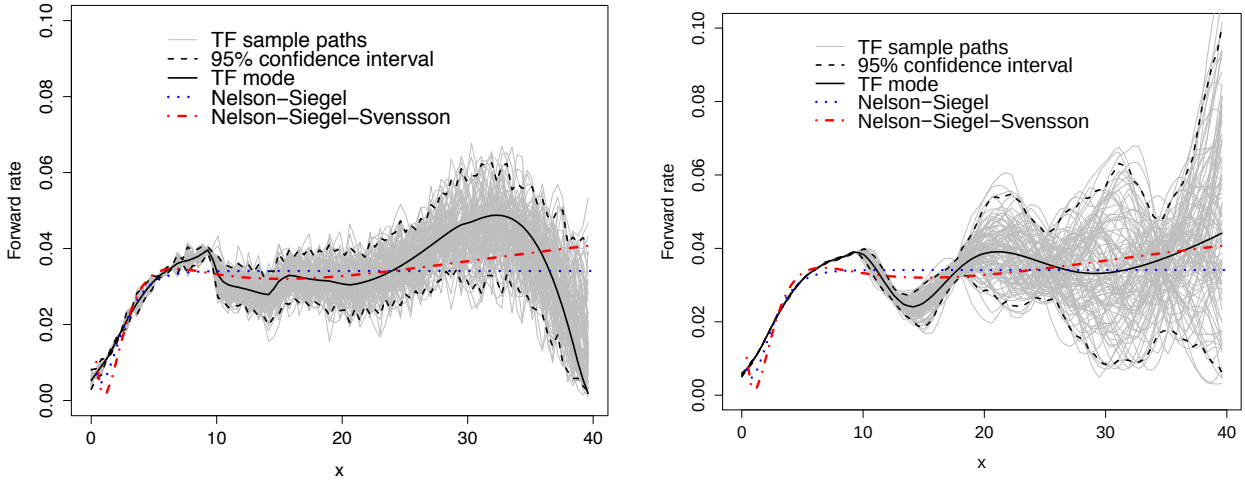

Figure 11: Forward rates obtained from sample paths of Figure 9 with Gaussian covariance function (left) and Matérn 5/2 covariance function (right). Gray lines represent $-\frac{d}{dx}\log Y^N(x)$ for each sample path. The black solid line is the most likely forward rate curve $-\frac{d}{dx}\log M_K^N\left(x \mid A, \boldsymbol{b}\right)$.

It appears that Nelson-Siegel and Nelson-Siegel-Svensson models do not fulfil all market constraints, contrary to the proposed methodology, which also gives confidence intervals.

Several quotation dates. Using the two dimensional approach described in Section 4.2, we build in Figure 12 a surface representing OIS discount factors with respect to time-to-maturities and quotation dates. The construction relies on OIS quotations at the 8 dates given in Table 4. The surface corresponds to the mode of the conditional GP given market-fit equality constraints and non-increasing constraints in the direction of time-to-maturities. We choose $N_x = 40$ and $N_t = 20$ and we consider a two-dimensional Gaussian kernel written as

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{(x-x')^2}{2\theta_1} - \frac{(t-t')^2}{2\theta_2}\right),$$

where $\boldsymbol{x} = (x, t)$ and $\boldsymbol{x}' = (x', t')$. For each vector $\boldsymbol{x} = (x, t)$, the first component x represents a time-to-maturity and the second component t represents a quotation date. Without loss of generality, the distance $t - t'$ between two quotation dates has been expressed in percentage of the length between the two extreme dates of the sample. The parameters θ_1 and θ_2 are fixed respectively to 25 and 0.5. Observe that the constructed discount factor surface is non-increasing with respect to time-to-maturities.

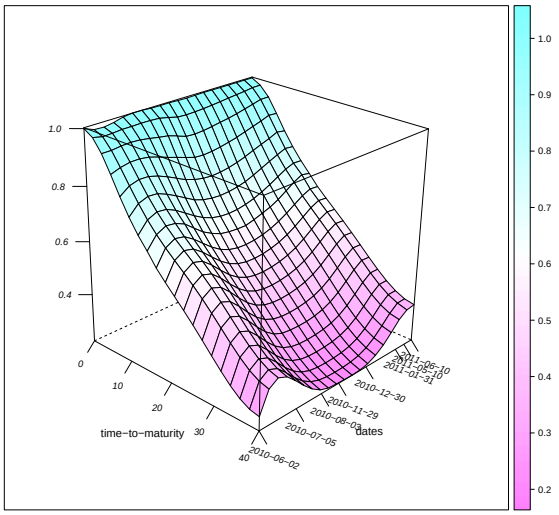

Figure 12: OIS discount factors as a function of time-to-maturities and quotation dates.

5.3 CDS implied default distributions

We now apply the kriging method to build CDS implied default distributions. Indeed, CDS can be seen as an insurance product that covers the loss of a particular debt issuer if the latter defaults over a certain protection period or maturity. The market quotes (CDS spreads) of these products provide information on the current cost for protection and, in turn, on how the market assesses default probabilities of the underlying entity at different time horizons.

Our aim is to construct, at different quotation dates t, implied survival functions $T \to Q(t, T)$ of a particular debt issuer by observing the corresponding term-structure of CDS spreads

(CDS spreads at increasing protection horizons). The quantity $Q(t,T)$ gives the probability at time t that the underlying debt issuer does not default before the time horizon T. In this numerical illustration, we consider CDS on the Russian sovereign debt for the 10 quotation dates given in Table 6. For each of these quotation dates, this corresponds to 7 CDS spreads $S_1, \ldots, S_7$ associated with protection maturities (in years) in the set $E := \{1, 2, 3, 4, 5, 7, 10\}^\top$. For each standard maturity $T \in E$, the values of the survival probabilities at each premium payment dates $\tau_1 < \cdots < \tau_n = T$ are linked through the linear relation (7). Note that the premium payment dates are separated by a quarter period and these dates coincide for all quoted CDS maturities. In our numerical illustrations, the expected recovery rate R is fixed to 40% and the discount factors $P^D(t, \tau_k)$ are constructed by linear interpolation of treasury constant maturity rates given (for all considered quotation dates) by the Federal Reserve Bank of St. Louis. Then, if $\delta = 1/4$ represents a quarter period, the implied default distribution is compatible with market quotes if the vector of survival probabilities $Q(t,X) := (Q(t,\delta), Q(t,2\delta), \ldots, Q(t,10))^\top$ satisfies a linear system of the form

$$A_t \cdot Q(t,X) = \boldsymbol{b}_t \tag{33}$$

where A_t is a 7×40 real matrix and $\boldsymbol{b}_t = (1-R, \ldots, 1-R)^\top \in \mathbb{R}^7$. In this case, we have $n = 7$ observations which depends on $m = 40$ points of the curve.

Parameters estimation. In Table 6, we compare the estimation of the length hyper parameter for a Gaussian kernel ($\hat{\theta}_G$) and for a Matérn 5/2 kernel ($\hat{\theta}_M$). The estimation has been performed using the ACV method described in Section 4.3. As can be seen on Figure 13, the LOO objective function looks similar for these two covariance functions.

Table 6: Length parameters estimation using ACV methods (CDS data).

Date	$\hat{\theta}_G$	$\hat{\theta}_M$	Optimal value Gaussian	Optimal value Matérn 5/2
06/01/2005	6.6	10.5	6.3e-06	**4.9e-07**
02/02/2006	4.9	15.5	2.7e-07	**1.8e-07**
20/03/2007	4.5	19.8	7.6e-06	**7.1e-07**
04/04/2008	9.7	11.4	5.0e-06	**2.6e-06**
11/05/2009	4.9	10.1	1.3e-04	**4.2e-05**
21/06/2010	11.1	10.1	1.1-05	**3.4e-06**
14/07/2011	14.5	17.1	**1.2e-06**	1.8e-06
23/08/2012	4.1	10.8	8.2e-06	**1.5e-06**

As previously, once the length parameter θ has been estimated, the standard deviation parameter σ is estimated using Equation (29).

One single quotation date. In Figure 14, we choose $N = 50$ and generate 100 sample paths of CDS implied survival curves, constructed from model (15) when using a Gaussian covariance function (left graph) and a Matérn 5/2 covariance function (right graph). All the curves are non-increasing with respect to time horizons. In addition, they all are perfectly compatible with CDS data as of 06/01/2005. The Gaussian process hyper-parameters have been estimated by the ACV method described in Section 4.3. The estimated hyper-parameters are given by $(\hat{\theta}_G, \hat{\sigma}_G) = (4.9,\ 0.09)$ when using a Gaussian covariance function and by $(\hat{\theta}_M, \hat{\sigma}_M) = (10.5,\ 0.22)$ when using a Matérn 5/2 covariance function. The black solid line represents the

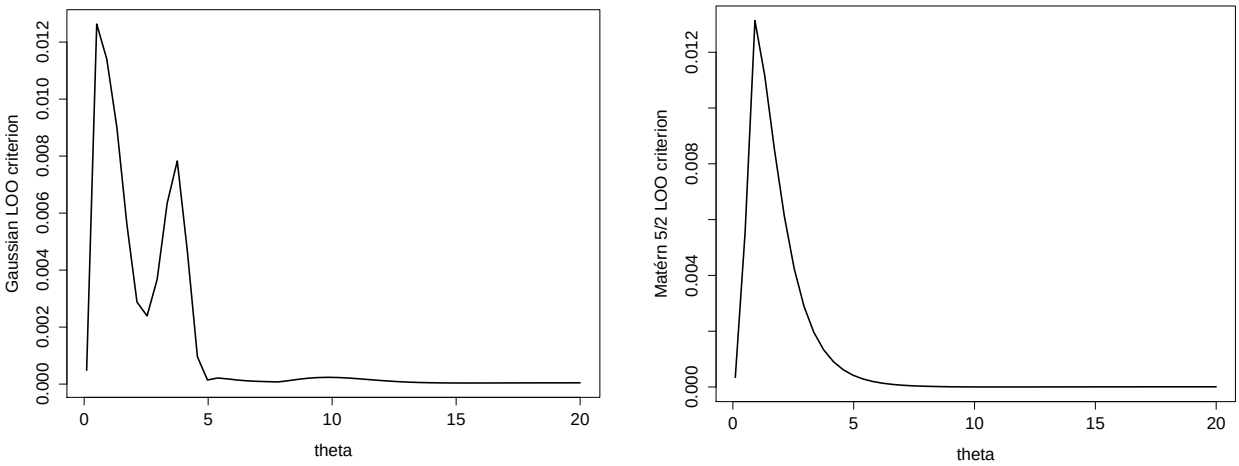

Figure 13: The function to be optimized in the LOO criterion (27) using the Gaussian (left) and the Matérn 5/2 (right) covariance function. CDS quotes as of 06/01/2005.

most likely curve, i.e., the mode of the conditional GP. Recall that, by construction, this curve satisfies the given constraints. The black dashed-lines represent the 95% point-wise confidence intervals quantified by simulation.

Several quotation dates. Using the two dimensional approach described in Section 4.2, we build in Figure 15 a surface representing CDS implied survival curves as a function of time horizons and quotation dates. The construction relies on CDS quoted spreads at the 8 dates given in Table 6. The surface corresponds to the mode of the conditional GP given market-fit equality constraints and non-increasing constraints in the direction of time-to-maturities. We choose $N_x = 40$ and $N_t = 20$ and we consider a two-dimensional Gaussian kernel written as

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{(x-x')^2}{2\theta_1} - \frac{(t-t')^2}{2\theta_2}\right),$$

where $\boldsymbol{x} = (x, t)$ and $\boldsymbol{x}' = (x', t')$. For each vector $\boldsymbol{x} = (x, t)$, the first component x represents a time-to-maturity and the second component t represents a quotation date. Without loss of generality, the distance $t - t'$ between two quotation dates has been expressed in percentage of the length between the two extreme dates of the sample. The parameters θ_1 and θ_2 are fixed respectively to 8 and 1.7. Notice that the constructed discount factor surface is non-increasing with respect to time-to-maturities. Considering several quotation dates simultaneously offers the advantage of increasing the data set for a better estimation of hyper-parameters and of creating a consistent interpolation procedure across two directions (time horizons and quotation dates).

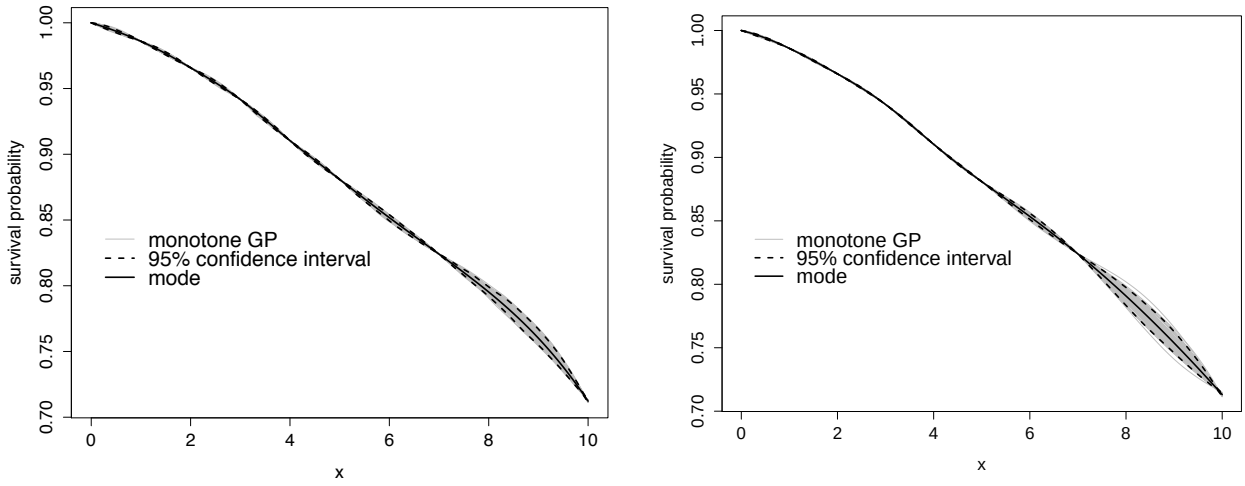

Figure 14: CDS implied survival curves (gray lines) given as simulated paths of a conditional GP with non-increasing constraints using a Gaussian covariance function (left) or a Matérn 5/2 covariance function (right).

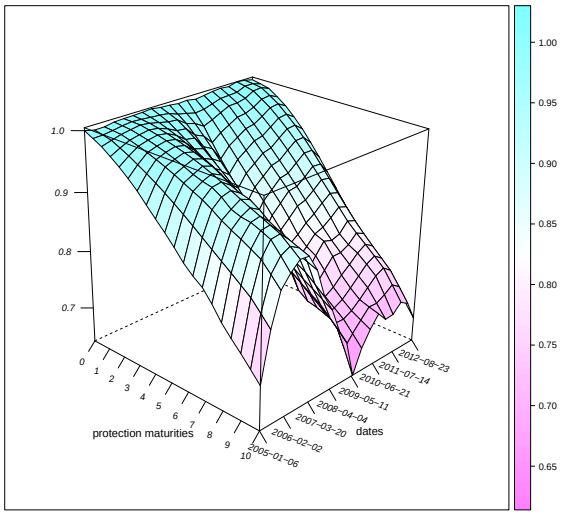

Figure 15: CDS implied survival probabilities as a function of time-to-maturities and quotation dates.

5.4 Need for monotonic kriging techniques

At the beginning of the Section 4, we justified monotonic techniques developed in this paper by some no-arbitrage constraints. A natural question is whether a monotonic condition is required

to construct realistic term-structures. Let us stress that the construction problems we have considered rely on a market information summarized by an ill-posed system of equations of the form

$$A\mathbf{Y}(\mathbf{X}) = \mathbf{b}\,,$$

where A is not necessarily a square matrix. Obviously, these constraints do not lead to a unique set of possible values for $Y(x^{(i)})$, $i = 1, \dots, m$, so that spot rates or implied default probabilities are not directly available at point $x^{(1)}, \dots, x^{(m)}$. In the absence of arbitrage opportunities, the process Y representing default-free zero-coupon bonds, discount factors or implied survival probabilities shall be monotone. It is thus natural to consider a constrained interpolation technique to construct curves based on such quantities.

Consider here the previous swap curve construction problem but relax the monotonicity constraint on the default-free zero-coupon bond process Y. As can be seen on Figure 16 (left panel), the resulting *kriging mean is not a decreasing function*. In addition, even when the average curve is monotonic, some *sample curves are clearly not monotonic*, leading to wide and *unrealistic confidence intervals*. On the right panel of Figure 16 one can see that corresponding spot rates may seem reasonable in average, but omitting the monotonicity constraints leads to wide confidence intervals. Notice that Figure 16 does just aim at illustrating some problems that can be encountered when using unconstrained techniques, so that other choices of kernel or parameters in an unconstrained setting are not discussed here.

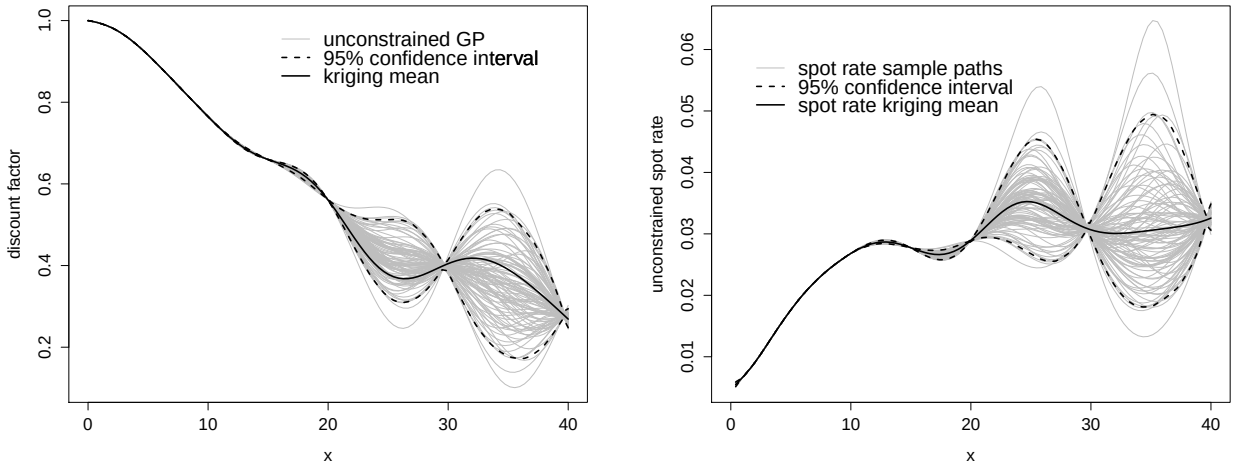

Figure 16: Interpolation of some discount factors (left panel) using standard kriging techniques without monotonicity constraints. OIS Data, 03/06/2010, Gaussian kernel with nugget 10^{-7}. Corresponding spot rates are given in the right panel. In this figure only, linear constraints $A\mathbf{Y}(\mathbf{X}) = \mathbf{b}$ has been approximated by point-wise constraints $Y(x^{(i)}) = M_K^N(x^{(i)}|A, \mathbf{b})$, $i = 1, \dots, m$, in order to use standard unconstrained R package *DiceKriging* (see Roustant et al., 2012). Estimated parameters from this package are $(\theta, \sigma^2) = (5.8, 0.22)$.

If one wants to avoid using monotone interpolation techniques, a natural idea is to interpolate some quantities that are deduced from $Y(x)$ and that are not necessarily monotonic. As an example, one can work on a function $\zeta(x)$ deduced in a bijective way from each path of

$Y(x)$. For example in a financial setting, if $Y(x)$ are discount factors, the function $\zeta(x)$ could be spot rates, $\zeta(x) = -\frac{1}{x}\log Y(x)$, or forward rates, $\zeta(x) = -\frac{d}{dx}\log Y(x)$. Obviously these functions does not suffer from monotonicity constraints, so that this could avoid using monotone interpolation. In the literature, some studies interpolate interest rates without considering any monotonic constraints. Steeley (2008) concludes that it is better to directly interpolates spot rates rather than discount factors: "*better yield curves estimates are obtained by fitting to the yield curve directly rather than fitting first to the discount function*". Some authors also propose to use kriging techniques to some non-monotonic curves, as Benth (2015) who applies kriging to futures curves from energy futures prices. In Kanevski et al. (2008), tools from spatial statistics and machine learning are applied to produce some interest rate mapping in a two dimensional feature space (maturity, time). However, in these studies, rates or prices to be interpolated are assumed to be directly observed. In our setting, we do not necessarily observe spot rates or even discount factors, and one aims at fitting market data in a non-parametric setting. We can make general objections to the interpolation of non-monotonic deduced quantities :

- The first objection is that even if $\zeta(x)$ is not monotone, it is *still constrained*: as an example, quantities such as discount rates or forward rates are expected to be positive in absence of arbitrage opportunity. As for the discount rates, interpolating spot rates could lead to locally negative average spot rates, and confidence intervals obtained by kriging can reach the threshold 0, which is not desirable: the constraints are translated from monotonicity constraints to positivity constraints, and classical kriging techniques cannot handle these constraints.

- A second objection is that in our setting, observations are $A\mathbf{Y}(\mathbf{X}) = \mathbf{b}$, where A is not necessarily a square and invertible matrix. Thus *we cannot directly observe*, with our data, spot rates or forward rates. Even in the very special case where A is a square and invertible matrix, spot rates can be deduced from $\mathbf{Y}(\mathbf{X})$, but not forward rates: knowing a function at some abscissas does not give straightforward constraints on its derivative.

- At last, using kriging interpolation, even if omitting the positivity of $\zeta(x)$, would lead to *non linear conditions* in terms of $\zeta(X)$. One has seen that conditionally to $A\mathbf{Y}(\mathbf{X}) = \mathbf{b}$, $Y(x)$ is still a Gaussian Process, but it would not be the case any more for the process $\zeta(x)$. Even in the simple case where A is invertible, the process $\zeta(x)$ must be positive and given $A\mathbf{Y}(\mathbf{X}) = \mathbf{b}$, it is not Gaussian any more, so that suited interpolation techniques still have to be introduced.

Conclusion

In this paper, we show how suitable kriging techniques can be used to quantify model uncertainty embedded in the construction of financial term-structures. We consider that the curve under construction is an unobservable path of a conditional (spatial) Gaussian process satisfying some linear equality constraints and monotone properties. A suitable cross-validation method is proposed to estimate the Gaussian process covariance parameters that control the level of uncertainty. We then investigate the efficiency of the proposed approach on some illustrative examples in one and two dimensions. The generated curves are all compatible with market quotes and respect non-arbitrage conditions. The conditional Gaussian process also allows to derive confidence bands for financial term-structures and related quantities. We compare the

Gaussian and the Matérn 5/2 covariance kernels over different data sets : swaps vs Euribor, overnight indexed swaps, credit default swaps. We conclude that, for these applications, the Matérn 5/2 covariance kernel seems to be the more appropriate since it generates more realistic forward curves in comparison with fitted Nelson-Siegel or Svensson models. In this work, we do not fully investigate the impact of curve uncertainty on the assessment of related products and their hedging strategies. Moreover, in the illustrative example we have presented, we consider that market information is observed without uncertainty. It may be the case that, due to lack of liquidity, market quotes cannot be considered to be reliable. The kriging techniques we have proposed could then be adapted to the presence of noisy observations (see section 3.3). Improvements on the rejection sampling algorithm could also be useful, especially for the estimation of hyper-parameters in dimension 2 where a large number of quotation dates has to be considered. These points are leaved for future research.

Acknowledgement

We would like to thank the three anonymous reviewers and the editor for the time spent on this article and for useful suggestions. The authors also thank Xavier Bay (EMSE) and Nicolas Durrande (EMSE) for helpful discussions. This work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. It also benefits from the support of the GRI in Financial Services and the Louis Bachelier Institute. One author also thanks the ANR Lolita research project and DAMI research project.

References

Abrahamsen, P. (1997). *A review of Gaussian random fields and correlation functions*. Norsk Regnesentral/Norwegian Computing Center.

Abrahamsen, P. and Benth, F. E. (2001). Kriging with inequality constraints. *Mathematical Geology*, 33(6):719–744.

Ametrano, F. and Bianchetti, M. (2009). *Bootstrapping the illiquidity: Multiple yield curves construction for market coherent forward rates estimation*, chapter 1. Risk Books.

Andersen, L. (2007). Discount curve construction with tension splines. *Review of Derivatives Research*, 10(3):227–267.

Asgharian, H., Hess, W., and Liu, L. (2013). A spatial analysis of international stock market linkages. *Journal of Banking and Finance*, 37(12):4738 – 4754.

Bachoc, F. (2013). Cross Validation and Maximum Likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66(0):55 – 69.

Barzanti, L. and Corradi, C. (1999). A note on direct term structure estimation using monotonic splines. *Rivista di matematica per le scienze economiche e sociali*, 22:101–108.

Bay, X., Grammont, L., and Maatouk, H. (2015). A New Method for Interpolating in a Convex Subset of a Hilbert Space. hal-01136466.

Bay, X., Grammont, L., and Maatouk, H. (2016). Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *ArXiv e-prints*.

Baysal, R. E., Nelson, B. L., and Staum, J. (2008). Response surface methodology for simulating hedging and trading strategies. In *Simulation Conference, 2008. WSC 2008. Winter*, pages 629–637. IEEE.

Benth, F. E. (2015). Kriging smooth futures curves. *Energy Risk*, February:64–69.

Branger, N. and Schlag, C. (2004). Model risk: A conceptual framework for risk measurement and hedging. In *EFMA 2004 Basel Meetings Paper*.

CFO Forum and CRO Forum (2010). QIS 5 Technical Specification; Risk-free interest rates. *Technical report, CFO Forum and CRO Forum.*

Chibane, M., Selvaraj, J., and Sheldon, G. (2009). Building curves on a good basis. working paper.

Chiu, N.-C., Fang, S.-C., Lavery, J. E., Lin, J.-Y., and Wang, Y. (2008). Approximating term structure of interest rates using cubic L_1 splines. *European Journal of Operational Research*, 184(3):990–1004.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical finance*, 16(3):519–547.

Cressie, N. (1990). The origins of kriging. *Mathematical geology*, 22(3):239–252.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, revised edition*. John Wiley & Sons, New York.

Davis, M. H. and Hobson, D. G. (2007). The range of traded option prices. *Mathematical Finance*, 17(1):1–14.

de Andrés Sánchez, J. and Gómez, A. T. (2004). Estimating a fuzzy term structure of interest rates using fuzzy regression techniques. *European Journal of Operational Research*, 154(3):804–818.

Debón, A., Martínez-Ruiz, F., and Montes, F. (2010). A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics and Economics*, 47(3):327 – 336.

Derman, E. (April 1996). Model Risk. *Quantitative Strategies Research Notes, Goldman Sachs*.

Eberlein, E. and Jacod, J. (1997). On the range of options prices. *Finance and Stochastics*, 1(2):131–140.

El Karoui, N., Jeanblanc-Picquè, M., and Shreve, S. E. (1998). Robustness of the Black and Scholes formula. *Mathematical finance*, 8(2):93–126.

Fengler, M. R. and Hin, L.-Y. (2015). A simple and general approach to fitting the discount curve under no-arbitrage constraints. *Finance Research Letters*, 15:78–84.

Fernandes, G. B. and Artes, R. (2015). Spatial dependence in credit risk and its improvement in credit scoring. forthcoming in European Journal of Operational Research.

Fries, C. P. (2013). Curves and term structure models: Definition, calibration and application of rate curves and term structure models. DZ Bank AG; LMU Munich, Department of Mathematics.

Fujii, M., Shimada, Y., and Takahashi, A. (2010). Collateral posting and choice of collateral currency. *Working Paper*.

Gan, G. and Sheldon Lin, X. (2015). Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics*, 62(0):138 – 150.

Gilli, M., Große, S., and Schumann, E. (2010). Calibrating the Nelson-Siegel-Svensson model. *COMISEF working papers series*, (31).

Golchi, S., Bingham, D. R., Chipman, H., and Campbell, D. A. (2015). Monotone Emulation of Computer Experiments. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):370–392.

Green, T. C. and Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4):1465–1499.

Guojun, G. (2013). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, 53(3):795 – 801.

Hagan, P. S. and West, G. (2006). Interpolation methods for curve construction. *Applied Mathematical Finance*, 13(2):89–129.

Henaff, P. (2010). A normalized measure of model risk. working paper.

Hull, J. and White, A. (2013). LIBOR vs. OIS: The derivatives discounting dilemma. *Journal of Investment Management*, 11(3):14–27.

Iwashita, Y. (2013). Piecewise Polynomial Interpolations. Open Gamma Technical Report.

Jones, D. R., Schonlau, M., and Welch, W. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492.

Kanevski, M., Maignan, M., Pozdnoukhov, A., and Timonin, V. (2008). Interest rates mapping. *Physica A: Statistical Mechanics and its Applications*, 387(15):3897–3903.

Kenyon, C. and Stamm, R. (2012). *Discounting, Libor, CVA and Funding: Interest Rate and Credit Pricing*. Palgrave Macmillan.

Kerkhof, J. and Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking & Finance*, 28(8):1845–1865.

Kimeldorf, G. S. and Wahba, G. (1970). A correspondence between bayesian estimation on stochastic processes and smooting by splines. *The Annals of Mathematical Statistics*, 41(2):495–502.

Kleijnen, J. P. and Van Beers, W. C. (2012). Monotonicity-preserving bootstrapped kriging metamodels for expensive simulations. *Journal of the Operational Research Society*, 64(5):708–717.

Krige, D. (1951). *A statistical approach to some mine valuation and allied problems on the Witwatersrand: By DG Krige*. PhD thesis, University of the Witwatersrand.

Laurini, M. P. and Moura, M. (2010). Constrained smoothing b-splines for the term structure of interest rates. *Insurance: Mathematics and Economics*, 46(2):339–350.

Laurini, M. P. and Ohashi, A. (2015). A noisy principal component analysis for forward rate curves. *European Journal of Operational Research*, 246(1):140–153.

Liu, M. and Staum, J. (2010). Stochastic kriging for efficient nested simulation of expected shortfall. *Journal of Risk*, 12(3):3.

Ludkovski, M. (2015). Kriging metamodels for Bermudan option pricing. working paper.

Maatouk, H. and Bay, X. (2014a). A New Rejection Sampling Method for Truncated Multivariate Gaussian Random Variables Restricted to Convex Sets. To appear in Monte Carlo and Quasi-Monte Carlo Methods 2014. Springer-Verlag, Berlin 2016.

Maatouk, H. and Bay, X. (2014b). Gaussian Process Emulators for Computer Experiments with Inequality Constraints. in revision.

Maatouk, H., Roustant, O., and Richet, Y. (2015). Cross-Validation Estimations of Hyper-Parameters of Gaussian Processes with Inequality Constraints. *Procedia Environmental Sciences*, 27:38 – 44. Spatial Statistics conference 2015.

Mardia, K. V., Kent, J. T., Goodall, C. R., and Little, J. A. (1996). Kriging and splines with derivative information. *Biometrika*, 83(1):207–221.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

Morini, M. (2011). *Understanding and Managing Model Risk: A practical guide for quants, traders and validators*. Wiley.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):pp. 473–489.

Pallavicini, A. and Tarenghi, M. (2010). Interest-rate modeling with multiple yield curves. *Available at SSRN 1629688*.

Paulson, N. D. and Hart, C. E. (2006). A spatial approach to addressing weather derivative basis risk: A drought insurance example. In *2006 Annual Meeting of American Agricultural Economics Association*.

Ramponi, A. (2003). Adaptive and monotone spline estimation of the cross-sectional term structure. *International Journal of Theoretical and Applied Finance*, 6(02):195–212.

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

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5(2).

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software*, 51(1):1–55.

Santner, T. J., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.

Smith, A. and Wilson, T. (2001). Fitting yield curves with long term constraints. Technical report, Technical report, Bacon and Woodrow.

Sousa, J. B., Esquível, M. L., and Gaspar, R. M. (2012). Machine learning Vasicek model calibration with Gaussian processes. *Communications in Statistics-Simulation and Computation*, 41(6):776–786.

Steeley, J. M. (2008). Testing term structure estimation methods: Evidence from the UK STRIPS market. *Journal of Money, Credit and Banking*, 40(7):1489–1512.

Stutvoet, E. (2007). Fitting Financial Models to Market Data Using Kriging. Master's thesis, Delft university of Technology.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. Technical report, National Bureau of Economic Research.