Kriging of financial term-structures

Areski Cousin
ISFA, Université Lyon 1

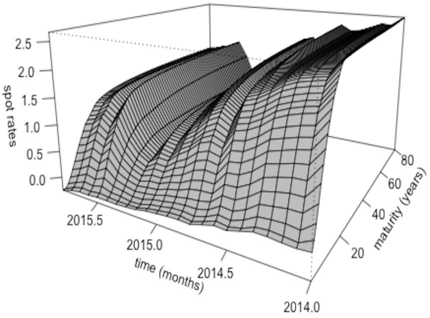
Joint work with Hassan Maatouk and Didier Rullière

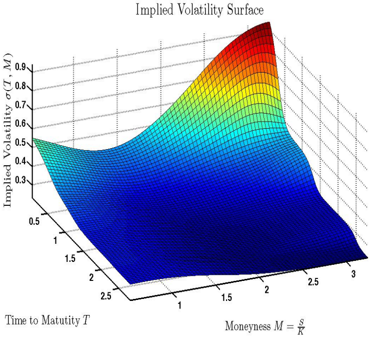
Séminaire de probabilité et statistiques, Université de Nice Sophia Antipolis

Nice, May 24, 2016

Motivation

- Financial term-structures describes the evolution of some financial or economic quantities as a function of time horizon.
- **Examples** : term-structure of interest-rates, bond yields, credit spreads, implied default probabilities, stock return implied volatilities.
- **Applications** : valuation of financial and insurance products, risk management



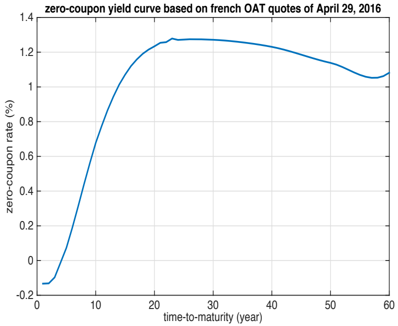


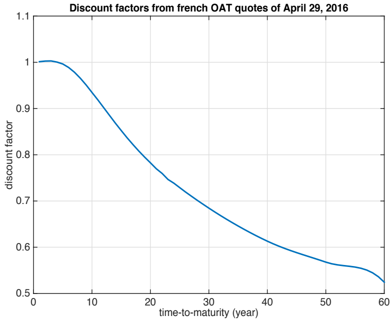
The term-structure construction problem

Several constraints have to be considered

- Compatibility with market information : at a given date t_0, the curve under construction $T \to P(t_0, T)$ shall be compatible with observed prices of some reference products. $\Longrightarrow$ Static problem.
- Arbitrage-free construction : this translates into some specific shape properties such as positivity, monotonicity, convexity or bounds on the curve values
- Additional conditions can be required : minimum degree of smoothness, control of local convexity

Example 1 : zero-coupon yield curve





Curves extracted from quotes of french OAT bonds as of April 29, 2016.
Source : "Comité de Normalisation Obligataire" (CNO).

Example 1 : zero-coupon yield curve

- The zero-coupon yield for time horizon T is defined as $Y(t_0,T) = -\frac{1}{T-t_0}\ln\left(P(t_0,T)\right)$ where $P(t_0,T)$ is the price at time t_0 of a (default-free) zero-coupon bond with maturity T
- However, $T \to P(t_0,T)$ is not directly observed : we only known information on this curve through market prices of some coupon-bearing bonds.
- Let $S_1,\ldots,S_n$ be observed prices at time t_0 of the issuer traded coupon bonds with maturity $T_1,\ldots,T_n$
- Under no-arbitrage condition, the observations S_i, $i=1,\ldots,n$ provides an information on $T \to P(t_0,T)$ in the form of a linear system.

The term-structure construction problem

Example 1 : zero-coupon yield curve (cont.)

- S_i : market price (in percentage of nominal) at time t_0 of a bond with maturity T_i
- c_i : coupon rate
- $t_1 < \ldots < t_{p_i} = T_i$: coupon payment dates, δ_k : year fraction of period (t_{k-1}, t_k)
- The no-arbitrage assumption gives the following linear relation

$$c_i \sum_{k=1}^{p_i} \delta_k P(t_0, t_k) + P(t_0, T_i) = S_i$$

In addition, the arbitrage-free curve $T \to P(t_0, T)$ is a decreasing function such that $P(t_0, t_0) = 1$

Example 2 : Discount curve based on overnight-indexed-swaps (OIS)

- S_i : par rate at time t_0 of an OIS with maturity T_i
- $t_1 < \cdots < t_{p_i} = T_i$: fixed-leg payment dates (annual time grid)
- δ_k : year fraction of period (t_{k-1}, t_k)

$$S_i \sum_{k=1}^{p_i-1} \delta_k P(t_0, t_k) + (S_i \delta_{p_i} + 1) P(t_0, T_i) = 1, \quad i = 1, \ldots, n$$

where $P(t_0, T)$ is the OIS discount factor with maturity T

In addition, the arbitrage-free curve $T \to P(t_0, T)$ is a decreasing function such that $P(t_0, t_0) = 1$

The term-structure construction problem

Example 3 : Default time distribution implied from CDS spreads

- S_i : fair spread at time t_0 of a credit default swap with maturity T_i
- $t_1 < \cdots < t_p = T_i$: trimestrial premium payment dates, δ_k : year fraction of period (t_{k-1}, t_k)
- $D(t_0, T)$ is the discount factor associated with maturity date T
- R : expected recovery rate of the reference entity

$$S_i \sum_{k=1}^{p_i} \delta_k D(t_0, t_k) P(t_0, t_k) = -(1 - R) \int_{t_0}^{T_i} D(t_0, u) dP(t_0, u)$$

where $T \to P(t_0, T)$ is the $\mathcal{F}_{t_0}$-conditional (risk-neutral) survival distribution of the reference entity.

The term-structure construction problem

Example 3 : Default time distribution implied from CDS spreads (cont.)

Using an integration by parts, the survival function $u \to P(t_0, u)$ satisfies a linear relation :

$$S_i \sum_{k=1}^{p_i} \delta_k D(t_0, t_k) P(t_0, t_k) + (1-R) D(t_0, T_i) P(t_0, T_i)$$
$$+ (1-R) \int_{t_0}^{T_i} f(t_0, u) D(t_0, u) P(t_0, u) du = 1 - R$$

where $f(t_0, u)$ is the instantaneous forward (discount) rate associated with maturity date u.

As a survival function, $T \to P(t_0, T)$ **shall be decreasing and such that** $P(t_0, t_0) = 1$

The term-structure construction problem

1) Compatibility with market information :

- At time t_0, we observe the market quotes $S_1, \ldots, S_n$ of n liquidly traded instruments
- The values of these products depend on the value of the curve at points $\tau_1, \ldots, \tau_m$

The vector of curve values $P(t_0, X) := (P(t_0, \tau_1), \ldots, P(t_0, \tau_m))^\top$ satisfies a linear system of the form

$$A \cdot P(t_0, X) = \boldsymbol{b}, \tag{1}$$

where

- A is a $n \times m$ real-valued matrix
- $\boldsymbol{b}$ is a n-dimensional column vector

$\Longrightarrow$ Indirect and partial information on the curve at points $\tau_1, \ldots, \tau_m$

2) No-arbitrage assumption :

In the previous examples, $T \to P(t_0, T)$ shall be a non-increasing function

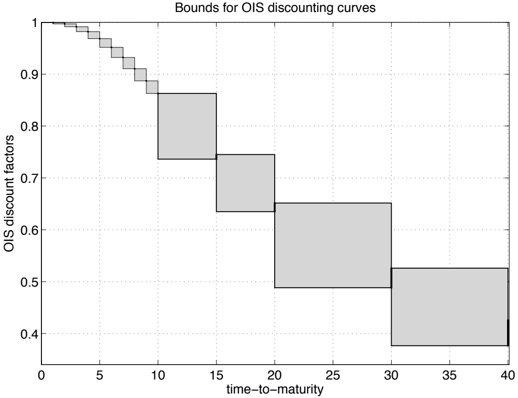
The term-structure construction problem

Uncertainty embedded in the construction of term-structure functions

- The curve $T \to P(t_0, T)$ is an infinite-dimensional mathematical object
- Partial information on P : in many applications, the number n of observations is quite small
- Uncertainty in the data : due to market illiquidity, quotes may not be fully reliable
- The no-arbitrage condition limits (to some extent) the uncertainty to the space of monotonic functions

Range of arbitrage-free OIS discount curves

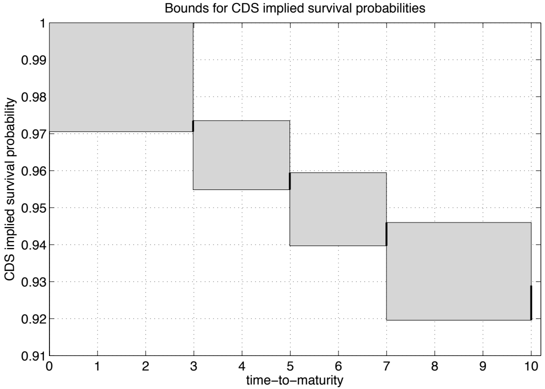
- $n = 14$ liquidly traded maturities. The associated OIS values depend on $m = 40$ points of the curve
- Cousin and Niang (2014) : No-arbitrage bounds on OIS discount factors



Input data : OIS swap rates as of May, 31st 2013.

Range of arbitrage-free CDS-implied survival functions

- $n = 4$ liquidly traded maturities. CDS fair spreads depend on $m = 40$ points of the curve
- Cousin and Niang (2014) : No-arbitrage bounds on the issuer implied survival distribution function



Input data : CDS spreads of AIG as of December 17, 2007, $R = 40\%$,
$$D(t, T) = \exp(-3\%(T - t))$$

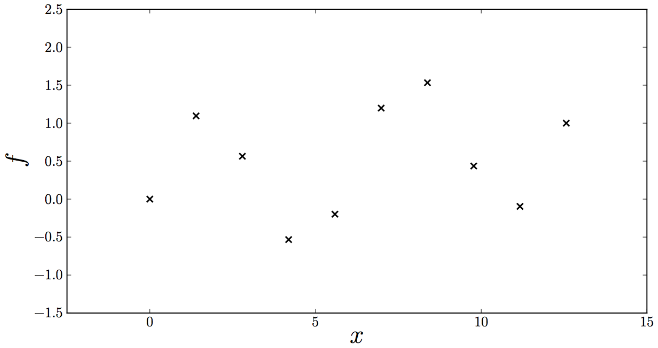
From spline interpolation to kriging

In practice, financial term-structures are constructed using deterministic interpolation techniques.

- Parametric approaches : Nelson-Siegel or Svensson models (used by most central banks)
- Non-parametric interpolation methods : shape-preserving spline techniques (lack of interpretability but better ability to fit the data).

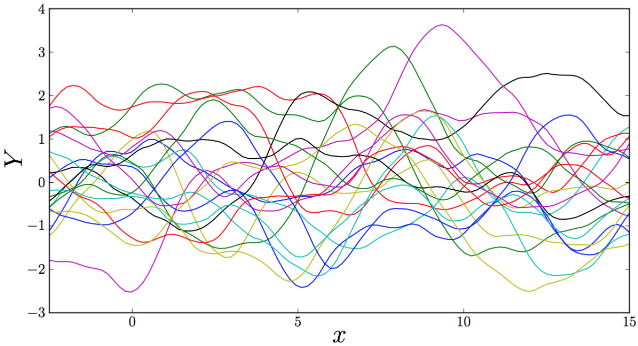
Could we propose an arbitrage-free interpolation method that additionally allows for quantification of uncertainty ?

A function f is only known at a limited number of points $x_1, \ldots, x_n$



Classical kriging

The (unknown) function f is assumed to be a sample path of a Gaussian process Y



Classical kriging

Definition : Gaussian process (GP) or Gaussian random field

A Gaussian process is a collection of random variables, any finite number of which have (consistent) joint Gaussian distributions.

A Gaussian process $\left(Y(x), x \in \mathbb{R}^d\right)$ is characterized by its mean function

$$\mu : x \in \mathbb{R}^d \longrightarrow \mathbb{E}(Y(x)) \in \mathbb{R}.$$

and its covariance function

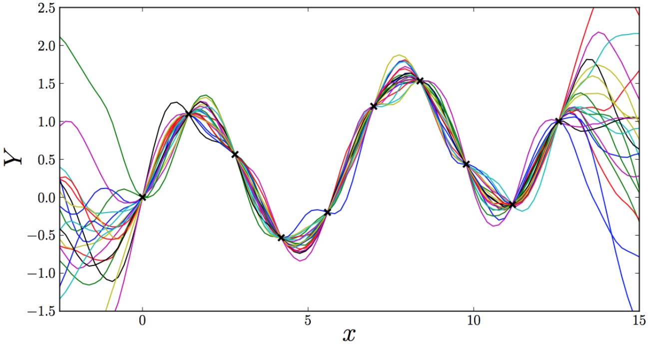
$$K : (x, x') \in \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathrm{Cov}(Y(x), Y(x')) \in \mathbb{R}.$$

Table: Some popular covariance functions $K(x, x')$ used in 1D kriging methods.

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

Classical kriging

The estimation of f relies on the conditional distribution of Y given the observed values $y_i = f(x_i)$ at points x_i, $i = 1, \dots, n$.



Classical kriging

- $\boldsymbol{X} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^{n \times d}$: some design points
- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$: observed values of f at these points
- $Y(\boldsymbol{X}) = (Y(x_1), \ldots, Y(x_n))^\top$: vector composed of Y at point $\boldsymbol{X}$

The conditional process is still a Gaussian Process

Let Y be a GP with mean μ and covariance function K. The conditional process $Y \mid Y(\boldsymbol{X}) = \boldsymbol{y}$ is a GP with mean function

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

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

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

where $\boldsymbol{\mu} = \mu(X) = (\mu(x_1), \ldots, \mu(x_n))^\top$, $\mathbb{K}$ is the covariance matrix of $Y(\boldsymbol{X})$ and $\boldsymbol{k}(x) = (K(x, x_1), \ldots, K(x, x_n))^\top$

Extension to linear equality constraints

Recall that, in our term-structure construction problem, the (unknown) real function f satisfies some linear equality constraints of the form

$$A \cdot f(X) = \boldsymbol{b}, \tag{2}$$

where

- A is a given matrix of dimension $n \times m$
- $X = (x_1, \ldots, x_m)^\top \in \mathbb{R}^{m \times d}$
- $f(X) = (f(x_1), \ldots, f(x_m))^\top \in \mathbb{R}^m$
- $\boldsymbol{b} \in \mathbb{R}^n$

Extension to linear equality constraints

- $X = (x_1, \ldots, x_m)^\top \in \mathbb{R}^{m \times d}$: some design points
- $\boldsymbol{b} = (b_1, \ldots, b_n)^\top \in \mathbb{R}^n$: right-hand side of the linear system
- $Y(X) = (Y(x_1), \ldots, Y(x_m))$: vector composed of Y at point $\boldsymbol{X}$

The conditional process is still a Gaussian Process

Let Y be a GP with mean μ and covariance function K. The conditional process $Y \mid AY(X) = \boldsymbol{b}$ is a GP with mean function

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

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

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

where $\boldsymbol{\mu} = \mu(X) = (\mu(x_1), \ldots, \mu(x_m))^\top$, $\mathbb{K}$ is the covariance matrix of $Y(X)$, $\boldsymbol{k}(x) = (K(x, x_1), \ldots, K(x, x_m))^\top$

Extension to monotonicity constraints

New formulation of the problem : estimation of an unknown function f given that

$$\begin{cases} A \cdot f(X) = \boldsymbol{b} \\ f \in \mathcal{M} \end{cases}$$

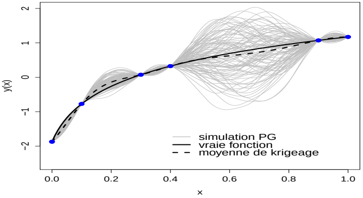
where $\mathcal{M}$ is the set of (say) non-increasing functions.

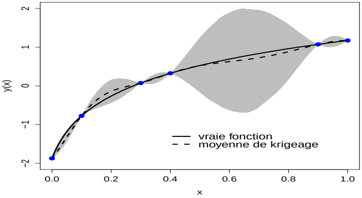
Problem : The conditional process is not a Gaussian process anymore.

- Which estimator could we propose for the term-structure ?
- How could we recover the distribution of the conditional process ?
- How to cope with the infinite-dimensional monotonicity constraints

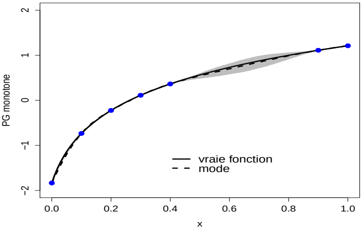
Extension to monotonicity constraints

What happens if a monotonic function is estimated using classical kriging (i.e., with no constraints) ?





Whereas kriging with monotonicity constraints gives

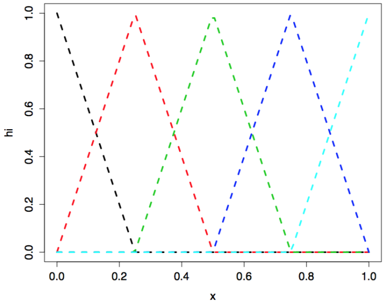


Extension to monotonicity constraints (1D case)

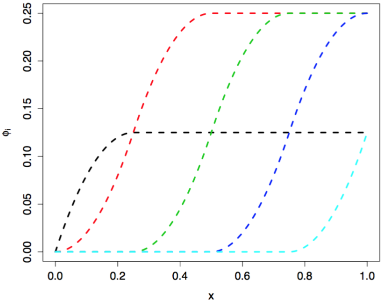
Proposed methodology : On an interval $D = [\underline{x}, \overline{x}]$ of $\mathbb{R}$, we construct a finite-dimensional approximation of Y for which the monotonicity constraint is easy to check.

- Regular subdivision $u_0 < \ldots < u_N$ of D with a constant mesh δ
- Set of increasing basis functions $(\phi_i)_{i=0,\ldots,N}$ defined on this subdivision

$$h_i(x) := \max\left(1 - \frac{|x-u_i|}{\delta}, 0\right)$$



$$\phi_i(x) = \int_{\underline{x}}^{x} h_i(u)du$$



Extension to monotonicity constraints (1D case)

Proposition (Maatouk and Bay, 2014b)

Let Y be a zero-mean GP with covariance function K and with almost surely differentiable paths.

- The finite-dimensional process Y^N defined on D by

$$Y^N(x) = Y(u_0) + \sum_{j=0}^{N} Y'(u_j)\phi_j(x)$$

  uniformly converges to Y, almost surely.

- Y^N is non-decreasing (resp. non-increasing) on D if and only if $Y'(u_j) \geq 0$ (resp. $Y'(u_j) \leq 0$) for all $j = 0, \ldots, N$.

- Let $\xi := (Y(u_0), Y'(u_0), \ldots, Y'(u_N))^\top$, then $\xi \sim \mathcal{N}(0, \Gamma^N)$ where

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

Extension to monotonicity constraints (1D case)

For a given covariance function K, we assume that the unknown function f is a sample path of the GP

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

where $\boldsymbol{\xi} := (\eta, \xi_0, \ldots, \xi_N)^\top \sim \mathcal{N}(0, \Gamma^N)$.

Kriging f is equivalent to find the conditional distribution of Y^N given

$$\begin{cases} A \cdot Y^N(X) = \boldsymbol{b} & \text{linear equality condition} \\ \xi_j \leq 0, \ j = 0, \ldots, N & \text{monotonicity constraint} \end{cases}$$

Extension to monotonicity constraints (1D case)

Or equivalently, to find the distribution of the truncated Gaussian vector $\boldsymbol{\xi} \sim \mathcal{N}(0, \Gamma^N)$ given

$$\begin{cases} A \cdot \Phi \cdot \boldsymbol{\xi} = \boldsymbol{b} & \text{linear equality condition} \\ \xi_j \leq 0, \ j = 0, \ldots, N & \text{monotonicity constraint} \end{cases}$$

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

$$\Phi_{i,j} := \begin{cases} 1 & \text{for } i = 1, \ldots, m \text{ and } j = 1, \\ \phi_{j-2}(x_i) & \text{for } i = 1, \ldots, m \text{ and } j = 2, \ldots, N+2. \end{cases}$$

Extension to monotonicity constraints (1D case)

Which estimator could we use for f ?

We consider the mode of the truncated gaussian process (most probable path) :

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

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

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

with

- $\mathcal{C} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^{N+2} \ : \ \xi_j \leq 0, \ j = 0, \ldots, N \right\}$
- $\mathcal{I}(A, \boldsymbol{b}) = \left\{ \boldsymbol{\xi} \in \mathbb{R}^{N+2} \ : \ A \cdot \Phi \cdot \boldsymbol{\xi} = \boldsymbol{b} \right\}$

Extension to monotonicity constraints (1D case)

The mode estimator has several advantages (over alternative estimators) :

- It satisfies the constraints on the entire domain D
- It is easy to compute as the solution of a quadratic optimisation problem
- It corresponds to the maximum a posteriori estimator in the sense of Bayesian statistics
- The mode estimator does not depend on the hyper-parameter σ
- As N tends to infinity, the limit of M_K^N corresponds to a constrained spline function that depends on K (Bay et al., 2016)

Efficient simulation of the truncated Gaussian vector

1) Simulate a truncated vector $\boldsymbol{\xi}$ given the linear equality constraint :

$$Z \sim \{\boldsymbol{\xi} \mid B \cdot \boldsymbol{\xi} = \boldsymbol{b}\} \sim \mathcal{N}\left((B\Gamma^N)^\top \left(B\Gamma^N B^\top\right)^{-1} \boldsymbol{b}, \Gamma^N - \left(B\Gamma^N\right)^\top \left(B\Gamma^N B^\top\right)^{-1} B\Gamma^N \right)$$

where $B = A \cdot \Phi$.

2) Simulate

$$\{Z \mid \xi_j \leq 0, j = 0, \ldots, N\} \sim \{\boldsymbol{\xi} \mid B \cdot \boldsymbol{\xi} = \boldsymbol{b} \text{ and } \xi_j \leq 0, j = 0, \ldots, N\}$$

by an accelerated rejection sampling algorithm (we use the method proposed in Maatouk and Bay, 2014a)

3) The corresponding sample curves $Y^N(\cdot) = \eta + \sum_{j=0}^{N} \xi_j \phi_j(\cdot)$ satisfies the constraints on the entire domain D.

Estimation of the covariance hyper-parameters

Classical covariance functions K depend on two parameters θ and σ. **How could we estimate these hyper-parameters ?**

- θ is estimated by minimizing a LOO cross-validation criterion

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

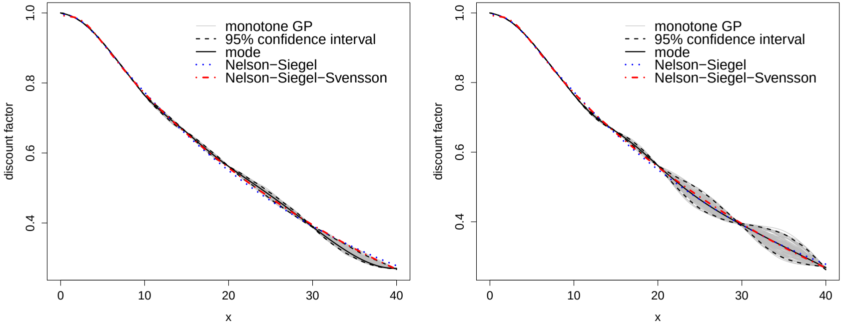
- Following Bachoc (2013), given $\widehat{\boldsymbol{\theta}}_{ACV}$, σ is estimated as the solution of

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

where $\mathcal{D}_i$ is the set of monotonicity and linear equality constraints without the i^{th} component

Kriging of OIS discount curves

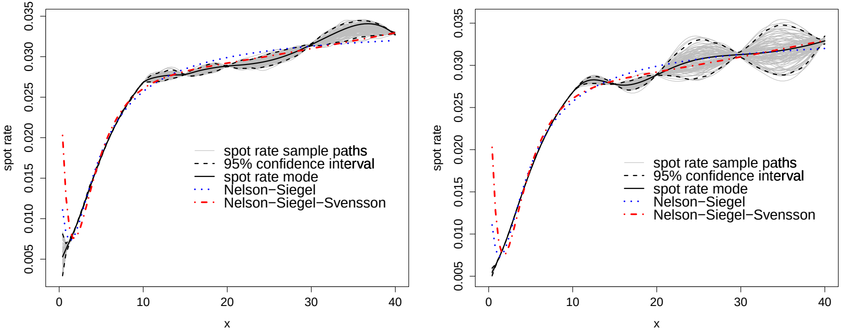
- We compare two covariance functions : Gaussian and Matérn 5/2
- Hyper-parameters θ and σ are estimated using cross-validation
- Comparison with Nelson-Siegel and Svensson curve fitting



Discount curves. $N = 50$, 100 sample paths. Left : Gaussian covariance function. Right : Matérn 5/2 covariance function. OIS data of 03/06/2010.

Kriging of OIS discount curves

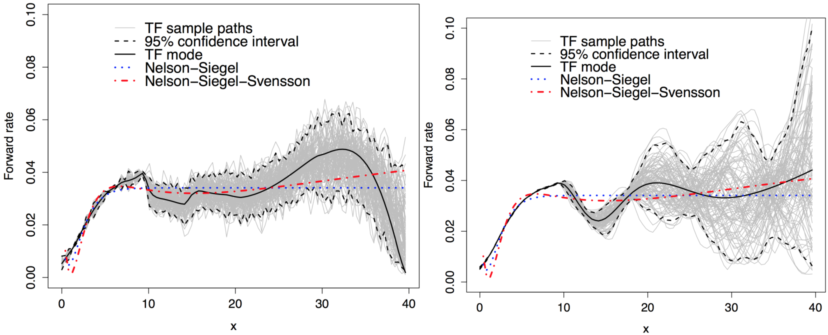
Corresponding spot rate curves : $-\frac{1}{x}\log P(x)$



Spot rate curves. Left : Gaussian covariance function. Right : Matérn 5/2 covariance function. OIS data of 03/06/2010. The black solid line is the most likely spot rate curve $-\frac{1}{x}\log M_K^N(x \mid A, \boldsymbol{b})$.

Kriging of OIS discount curves

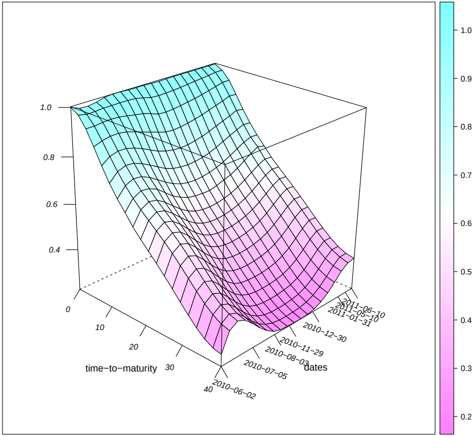
Corresponding forward rate curves : $-\frac{d}{dx}\log P(x)$



Spot rate curves. Left : Gaussian covariance function. Right : Matérn 5/2 covariance function. OIS data of 03/06/2010. The black solid line is the most likely forward rate curve $-\frac{d}{dx}\log M_K^N(x \mid A, \boldsymbol{b})$.

Kriging of OIS discount curves (2D)

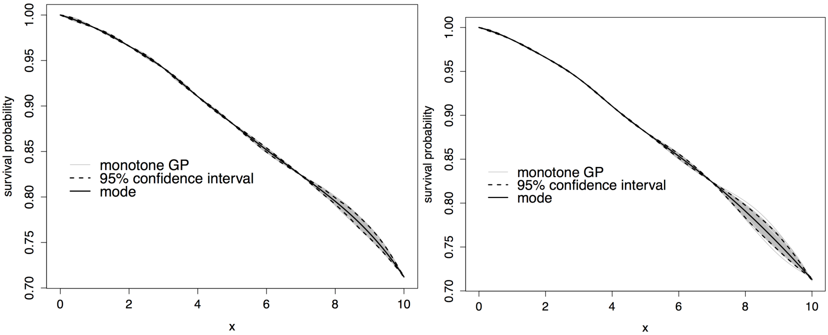
The previous approach can be extended in dimension 2.



Dicount curves. OIS discount factors as a function of time-to-maturities and quotation dates.

Kriging of CDS-implied default distribution

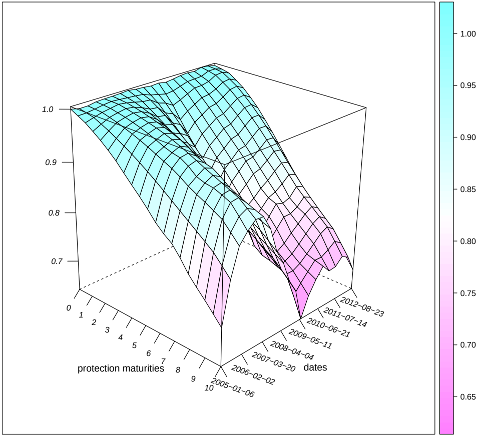
Implied survival function of the Russian sovereign debt



CDS implied survival curves. $N = 50$, 100 sample paths. Left : Gaussian covariance function. Right : Matérn 5/2 covariance function. CDS spreads as of 06/01/2005.

Kriging of CDS-implied default distribution (2D)

The previous approach can be extended in dimension 2.



Survival curves. CDS implied survival probabilities as a function of time-to-maturities and quotation dates.

Perspectives

- Impact of curve uncertainty on the assessment of related products and their associated hedging strategies
- Kriging of other type of financial term-structures such as arbitrage-free volatility surfaces
- What if the underlying market quotes are not reliable due to e.g. market illiquidity (data observed with a noise) ?

Thanks for your attention.

References

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.

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

Cousin, A., Maatouk, H., and Rullière, D. (2015).
Kriging of Financial Term-Structures.

Cousin, A. and Niang, I. (2014).
On the Range of Admissible Term-Structures.

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

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 SIAM/ASA J. Uncertainty Quantification.