Article

Gaussian Process Regression for Swaption Cube Construction under No-Arbitrage Constraints

Areski Cousin [1,2,*], Adrien Deleplace [1] and Adrien Misko [1]

[1]   Nexialog Consulting, 75011 Paris, France
[2]   Institut de Recherche en Mathématique Avancée, Université de Strasbourg, 7 rue René Descartes, CEDEX, 67084 Strasbourg, France
[*]   Correspondence: acousin@nexialog.com or a.cousin@unistra.fr

Abstract: In this paper, we introduce a 3D finite dimensional Gaussian process (GP) regression approach for learning arbitrage-free swaption cubes. Based on the possibly noisy observations of swaption prices, the proposed 'constrained' GP regression approach is proven to be arbitrage-free along the strike direction (butterfly and call-spread arbitrages are precluded on the entire 3D input domain). The cube is free from static arbitrage along the tenor and maturity directions if swaption prices satisfy an infinite set of in-plane triangular inequalities. We empirically demonstrate that considering a finite-dimensional weaker form of this condition is enough for the GP to generate swaption cubes with a negligible proportion of violation points, even for a small training set. In addition, we compare the performance of the GP approach with the SABR model, which is applied to a data set of payer and receiver out-of-the-money (OTM) swaptions. The constrained GP approach provides better prediction results compared to the SABR approach. In addition, we show that SABR calibration is better when using the GP cube output as new observations (in terms of predictive error and absence of arbitrage). Finally, the GP approach is able to quantify in- and out-of-sample uncertainty through Hamiltonian Monte Carlo simulations, allowing for the computation of model risk Additional Valuation Adjustment (AVA).

Keywords: swaption cube; constrained kriging; no arbitrage; option pricing; Hamiltonian Monte Carlo

1. Introduction

The swaption cube is an important financial object as it is used as an input of many interest-rate derivative pricing engines. It consists in a swaption implied volatility surface, which is indexed by maturity, strike and tenor. The swaption market is one of the biggest in the interest rate derivatives market. Swaptions are usually used to hedge, and while at-the-money (ATM) swaptions are observed to be very liquid, out-of-the-money swaption prices are often missing or not reliable. The major issue with the swaption cube is that along its three axes, missing quotes must be interpolated in an arbitrage-free way in order to be used as the input of some pricing or hedging models.

For equity options, the volatility surface, indexed by the maturity and the strike of the derivative, is a topic that has been thoroughly studied. A number of machine learning methods for learning pricing functions have emerged (see, e.g., Ruf and Wang (2019) for an in-depth literature review). For these volatility surfaces, absence of arbitrage is guaranteed if there is no call spread, butterfly spread and calendar spread arbitrage opportunities. A Gaussian process regression methodology, constrained on the no-arbitrage conditions

for equity options, has been introduced in Chataigner et al. (2021). Using this approach, they achieved to produce arbitrage-free interpolated prices, and local volatility surfaces, with their uncertainties. In the literature, several ways of interpolating the missing prices or volatilities of the swaption cube have been introduced. A distinction needs to be made between functional interpolation, which is for example polynomial interpolation, and stochastic interpolation, where stochastic models such as SABR and CEV models are used. The most common method for interpolating in an arbitrage-free way between quoted strikes, as well as extrapolating beyond, is to use the SABR model. However, this construction method requires several observed quotes at different strikes for each considered pair of tenors and maturities. When there is only one quoted strike observation, for a given tenor and maturity, three methods are usually employed in the industry to interpolate the swaption smile. Two of them are based on the cap market. The first approach consists of using caplet SABR parameters as a basis for the calibration of the swaption SABR parameters (see, e.g., Hagan and Konikov (2004)). The second, discussed in Skantzos and Garston (2019), consists of mapping underlying caplet volatilities to the swaption volatility by making an assumption in relating the moneyness of the swaption to the moneyness of the underlying caplets. The validity of both approaches decreases when the tenor of the considered swaptions increases. When cap market data are not available, parameters of the SABR model are taken as the closest SABR parameters that are calibrated with quotes. In order to interpolate with respect to maturity and tenor, it is possible to use functional interpolation directly on swaptions prices, or in a more elaborate manner, to use functional interpolation on the parameters of stochastic models. As explained in Johnson and Nonas (2009), in addition to the no call spread and butterfly spread arbitrage, the interpolated swaption prices must also respect an additional arbitrage condition, named "in-plane triangular".

Inspired by the local volatility construction proposed by Chataigner et al. (2021), the main goal of this paper is to present and study a 3D-constrained GP construction method for interpolating arbitrage-free swaption price cubes at a given quotation date. Compared to the equity local volatility construction method proposed in Chataigner et al. (2021), several improvements have been added:

- The constrained GP regression approach has been extended to an input space of dimension 3. Numerical implementation of the 3D 'constrained' kriging problem is much more involved than in the 2D case considered in Chataigner et al. (2021). We show that considering an anisotropic stationary kernel allows us to benefit from the tensorization of the GP covariance matrix, and it significantly reduces numerical complexity.
- Compared to the equity price surface, the no-arbitrage constraints on swaption prices are no more directional. Indeed, the in-plane triangular inequality involves prices of swaptions with different maturities and tenors. This constraint is by far more complex to handle than the no calendar spreads arbitrage constraint in the equity option case. It can be viewed as a weaker form of the no calendar spreads arbitrage condition, and it requires an infinite number of checks in order to be imposed everywhere on the domain. In our methodology, we consider a weaker version of the in-plane inequality constraint, and we empirically check in Section 4 that the weaker form is enough to ensure no arbitrage almost everywhere.
- Computation of the GP prior hyperparameters by Maximum Likelihood Estimation (MLE) is much longer in the swaption case than in the equity case (see Section 3.3).

Search for the maximal likelihood has been considerably improved by explicitly computing the Jacobian of the log-likelihood function.

The paper is organized as follows. In Section 2, we present the conditions on European swaption prices that preclude from any static arbitrage opportunity at a particular quotation date. These arbitrage-free conditions on the swaption cube translate into a convexity and a monotonicity constraints in strike, and into an "in-plane triangular" constraint in maturity and tenor. In Section 3, we detail the constrained GP regression methodology used to construct swaption price cubes that do not lead to arbitrage opportunities. Finally, in Section 4, we apply the presented GP methodology to a data set of payer and receiver OTM swaptions, and we compare the result to a SABR construction approach, which is the standard market practice to parameterize rate volatilities.

2. Absence of Static Arbitrages for the Swaption Cube

Before establishing the arbitrage-free constraints for the swaption cube, we set some notation.

Let $S_w(T_s, T_e, K)$ be the price, at a given quotation date, of a European payer (receiver) swaption, with a payer (receiver) swap $S(T_s, T_e, K)$ as underlying asset, where T_s and T_e are, respectively, the first and last payment date of the swap with fixed interest rate K. For equity options, in order to construct an arbitrage-free price or volatility surface, we must verify that the price of call spreads, butterfly spreads and calendar spreads are positive. In the strike direction, the butterfly and call-spread arbitrage conditions also apply to the swaption cube. However, the calendar spread condition does not hold here, as explained in Johnson and Nonas (2009). A non-trivial no-arbitrage constraint between European swaptions with the same strike and different option expiries was introduced in this article. Since this condition translates into an inequality between the price of three swaptions with the same strikes, it is called an "in-plane triangular condition".

Proposition 1. *The European payer (receiver) swaption price cube* $(T, t, K) \to S_w(T, T + t, K)$ *is free of static arbitrage if and only if the three following conditions hold:*

(i)   *For all* $T, t, K \geq 0$, $S_w(T, T + t, K) \geq 0$;
(ii) $K \to S_w(T, T + t, K)$ *is a convex, decreasing (increasing) function;*
(iii)   *The in-plane triangular condition holds:*
    For all $T, t, h, K \geq 0$, $S_w(T, T + t, K) + S_w(T + t, T + t + h, K) \geq S_w(T, T + t + h, K)$.

Item (ii) comes from the need for call spreads and butterfly spreads to be positive. A proof of (iii) is given in Johnson and Nonas (2009).

3. GP Regression to Construct an Arbitrage Free Swaption Cube

In this section, we present a GP-based methodology to build an arbitrage-free European payer (receiver) swaption cube.

For an input domain $\mathcal{D}$ of maturities, tenors and strikes, we build at quotation date t_0, a payer (receiver) swaption price cube

$$S_w : \mathcal{D} \to \mathbb{R}^+$$
$$(T, t, K) \mapsto S_w(T, T + t, K),$$

satisfying the no arbitrage conditions given in Proposition 1, from n noisy observations $\mathbf{y} = [y_1, ..., y_n]^\top$ of the function S_w at input points $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]$ where $\mathbf{x}_i = (T_i, t_i, K_i)$,

$i = 1, \ldots, n$. We assume that the price function S_w is represented as a Gaussian process. The market fit condition is then written as

$$y = S_w(\mathbf{X}) + \varepsilon, \tag{1}$$

where $S_w(\mathbf{X}) := [S_w(\mathbf{x}_1), \ldots, S_w(\mathbf{x}_n)]^\top$ is the vector of the values of the GP at the input points $\mathbf{X}$. The additive noise term $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]^T$ is assumed to be a zero-mean Gaussian vector, independent from S_w, and with a homoscedastic covariance matrix given as $\varsigma^2 I_n$, where I_n is the identity matrix of size n.

3.1. Classical 3-Dimensional Gaussian Regression

We consider a zero-mean Gaussian process prior on the mapping $S_w = S_w(\mathbf{x})_{\mathbf{x} \in D}$ with covariance function c.

Then, the output vector $S_w(\mathbf{X})$ has a normal distribution with zero mean and covariance matrix $\mathbf{C}$ with components $\text{cov}(S_w(T_i, T_i + t_i, K_i), S_w(T_j, T_j + t_j, K_j)) = c((T_i, t_i, K_i), (T_j, t_j, K_j)) = c(\mathbf{x_i}, \mathbf{x_j})$. We consider a 3-dimensional anisotropic, stationary covariance kernel given as, for $\mathbf{x}$ and $\mathbf{x}' \in \mathcal{D}$,

$$c(\mathbf{x}, \mathbf{x}') = \sigma^2 R_T(T - T', \theta_T) R_t(t - t', \theta_t) R_K(K - K', \theta_K),$$

where $\theta_T, \theta_t, \theta_K$ are the length scale parameters associated with directions T, t, K, and σ^2 is the marginal variance of the GP prior. The functions R_T, R_t and R_K are kernel correlation functions. As explained in Rasmussen and Williams (2005), $S_w \mid y = S_w(\mathbf{X}) + \varepsilon$ is again a GP with mean function η and covariance function $c^\star$ given respectively by

$$\eta(\mathbf{x}) = \mathbf{c}(\mathbf{x})^\top (\mathbf{C} + \varsigma^2 I_n)^{-1} y, \quad \mathbf{x} \in \mathcal{D} \tag{2}$$

and

$$c^\star(\mathbf{x}, \mathbf{x}') = c(\mathbf{x}, \mathbf{x}') - \mathbf{c}(\mathbf{x})^\top (\mathbf{C} + \varsigma^2 I_n)^{-1} \mathbf{c}(\mathbf{x}'), \quad \mathbf{x}, \mathbf{x}' \in \mathcal{D}, \tag{3}$$

where $\mathbf{c}(\mathbf{x}) = [c(\mathbf{x}, \mathbf{x}_1), \ldots, c(\mathbf{x}, \mathbf{x}_n)]^\top$.

Prediction and uncertainty quantification are made using the conditional distribution $S_w \mid y = S_w(\mathbf{X}) + \varepsilon$. The best linear unbiased estimator of S_w is given as the conditional mean function (2). The conditional covariance function c^* can be used to obtain confidence bounds around the predicted price surface. The hyperparameters of the kernel function c, as well as the variance of the noise, can be estimated using a maximum likelihood approach or a cross-validation approach (see Rasmussen and Williams (2005)). The model described in this subsection corresponds to unconstrained Gaussian process regression, which therefore does not take into account the arbitrage-free conditions as described in Proposition 1.

3.2. Imposing No Arbitrage Constraints to GP Regression

Conditional on the inequality constraints given in Proposition 1, the process S_w is no longer Gaussian, and this leads to two difficulties. The first is that we depart from the classical framework of GP regression, where the posterior distribution remains Gaussian. The second is that testing the inequality constraints on the entire input domain would require an infinite number of checks. We will use the solution proposed by Cousin et al. (2016) which consists of constructing a finite-dimensional approximation S_w^N of the Gaussian prior S_w, which is designed in such a way that the constraints can be imposed for the entire domain $\mathcal{D}$ with a finite number of checks. We then recover the constrained posterior

distribution by sampling a truncated Gaussian vector. Let us describe the methodology with more details.

We first rescale the input domain $\mathcal{D}$ to $[0,1]^3$. Without loss of generality, we then consider a discretized version of this rescaled domain as a 3-dimensional regular grid with nodes (T_i, t_j, K_k), $i = 1, ..., N_T$, $j = 1, ..., N_t$, $k = 1, ..., N_K$, where N_T, N_t, N_K are, respectively, the number of maturities, tenors and strikes we chose for our grid, and $(\delta_T, \delta_t, \delta_K)$ the corresponding constant steps[1]. Then, $N = N_T \times N_t \times N_K$ is the total number of nodes of the 3-dimensional grid. Now, each node of the grid is associated with a hat basis function defined as:

$$\phi_{i,j,k}(T,t,K) := \max(1 - \frac{|T - T_i|}{\delta_T}, 0) \max(1 - \frac{|t - t_j|}{\delta_t}, 0) \max(1 - \frac{|K - K_k|}{\delta_K}, 0) .$$

Then, the GP prior S_w is approximated on $\mathcal{D}$ by the following finite-dimensional process:

$$S_w^N(T, T + t, K) = \sum_{i=1}^{N_T} \sum_{j=1}^{N_t} \sum_{k=1}^{N_K} S_w(T_i, T_i + t_j, K_k) \phi_{i,j,k}(T,t,K), \quad \text{for all } (T,t,K) \in \mathcal{D}. \quad (4)$$

Note that the process S_w^N corresponds to a piecewise linear interpolation of S_w at nodes (T_i, t_j, K_k), for $i = 1, \ldots, N_T$, $j = 1, \ldots, N_t$, and $k = 1, \ldots, N_K$.

If we denote $S_w(T_i, T_i + t_j, K_k)$ as $\xi_{i,j,k}$, then the vector $\boldsymbol{\xi} = [\xi_{1,1,1}, \ldots, \xi_{i,j,k} \ldots, \xi_{N_T, N_t, N_K}]^\top$ is a zero-mean Gaussian vector with $N \times N$ covariance matrix Γ^N with components $\Gamma^N_{\mathbf{x},\mathbf{x}'} = c(\mathbf{x}, \mathbf{x}')$ for $\mathbf{x}$ and $\mathbf{x}'$ two nodes of the grid. Let us define $\boldsymbol{\phi}(T,t,K)$ as the N dimensional vector given as:

$$\boldsymbol{\phi}(T,t,K) = [\phi_{1,1,1}(T,t,K), ..., \phi_{i,j,k}(T,t,K), ..., \phi_{N_T, N_t, N_K}(T,t,K)].$$

Then, the finite-dimensional GP prior defined in (4) can be restated in matrix form: Equality (4) can then be written as:

$$S_w^N(T, T + t, K) = \boldsymbol{\phi}(T,t,K) \cdot \boldsymbol{\xi}.$$

Then, if we define $\boldsymbol{\Phi}(\mathbf{X})$ as the $n \times N$ matrix of basis functions such that each row i corresponds to the vector $\boldsymbol{\phi}(\mathbf{x}_i)$, we have that $S_w^N(\mathbf{X}) = \boldsymbol{\Phi}(\mathbf{X}) \cdot \boldsymbol{\xi}$, with $S_w^N(\mathbf{X}) := [S_w^N(\mathbf{x}_1), ..., S_w^N(\mathbf{x}_n)]^\top$.

From now on, we will refer to $\boldsymbol{\Phi}(\mathbf{X})$ as $\boldsymbol{\Phi}$.

Proposition 2. *The following statements hold for European payer swaptions.*

(i)   *The finite-dimensional process S_w^N uniformly converges to S_w on $\mathcal{D}$ as $N_T \to \infty$, $N_t \to \infty$ and $N_K \to \infty$, almost surely;*

(ii) *S_w^N is a decreasing function of K on $\mathcal{D}$ if and only if $\xi_{i,j,k} \geq \xi_{i,j,k+1}$;*

(iii) *S_w^N is a convex function of K on $\mathcal{D}$ if and only if $\xi_{i,j,k+2} - \xi_{i,j,k+1} \geq \xi_{i,j,k+1} - \xi_{i,j,k}$.*

These properties can be proved using the same methodology as in Maatouk and Bay (2017) for proving monotonicity and convexity. Note that these inequality constraints are linear in the components of $\boldsymbol{\xi}$.

Remark 1. *For a European receiver swaption, the above properties are the same for (i) and (iii). However, (ii) has to be replaced by : S_w^N is an increasing function of K if and only if $\xi_{i,j,k} \leq \xi_{i,j,k+1}$.*

However, it is not possible to impose the in-plane triangular condition everywhere on the domain $\mathcal{D}$ with a finite number of checks. In empirical Section 4, we consider a weaker version of this constraint by only imposing its validity for time steps of size 1 year.

This constraint is set using a "fictitious grid" of maturities $T^* = [T_1^*, T_2^*, \ldots, T_{N_T^*}^*]$, where $T_1^* = T_1$, $T_{N_T^*}^* = T_{N_T}$ and the step $\delta_T^* = \frac{1}{T_1 - T_{N_T}}$. Therefore, it is a rescaled 1 year step.

Formally, we consider the following weaker version of the in-plane triangular condition:

$$S_w^N(T_l^*, T_{l+1}^*, K) + S_w^N(T_{l+1}^*, T_{l+2}^*, K) - S_w^N(T_l^*, T_{l+2}^*, K) \geq 0, \quad l = 1, \ldots, N_T^*. \tag{5}$$

This is equivalent to:

$$\boldsymbol{\phi}(T_l^*, T_{l+1}^*, K) \cdot \boldsymbol{\xi} + \boldsymbol{\phi}(T_{l+1}^*, T_{l+2}^*, K) \cdot \boldsymbol{\xi} - \boldsymbol{\phi}(T_l^*, T_{l+2}^*, K) \cdot \boldsymbol{\xi} \geq 0. \tag{6}$$

Then, (6) is linear in $\boldsymbol{\xi}$ and can be added to the constraints defined in Proposition 2.

Given property (i) of Proposition 2, if we define $\mathcal{M}$ as the set of 3-dimensional continuous functions which are increasing and monotonic with respect to K, and respecting the in-plane triangular inequality (for steps of size 1 year, defined as (5)), then our construction problem consists in finding the conditional distribution of S_w^N such that

$$\begin{cases} \boldsymbol{y} = S_w^N(\mathbf{X}) + \boldsymbol{\varepsilon} \\ S_w^N \in \mathcal{M}. \end{cases}$$

Given properties (ii) and (iii) of Proposition 2, our process S_w^N satisfies the no-arbitrage conditions in the strike direction on the entire domain $\mathcal{D}$ when these constraints are satisfied at the nodes. The problem above can then be restated as follows

$$\begin{cases} \boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} \\ \boldsymbol{\xi} \in \mathcal{C}_{ineq}. \end{cases}$$

Indeed, $S_w^N \in \mathcal{M} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}_{ineq}$, where $\mathcal{C}_{ineq}$ is a set of linear inequality constraints as defined in (ii), (iii) of Proposition 2 and the relaxed in-plane triangular inequality constraint (6).

3.3. Maximum Likelihood Estimation

The parameters of the kernel function c and the noise parameter ς can be given as an input or estimated. Let $\Theta = [\theta_T, \theta_t, \theta_K, \sigma, \varsigma^2]^\top$. In our approach, we estimate Θ as the hyperparameter vector that maximizes the marginal log likelihood $\mathcal{L}(\Theta)$ for the process S_w^N.

The unconstrained Gaussian Likelihood is written as $\mathbb{P}(\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon}|\Theta)$. Under the finite dimensional approximation, the marginal log likelihood can be expressed as:

$$\mathcal{L}(\Theta) = -\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{\Phi}\Gamma^N\boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1}\boldsymbol{y} - \frac{1}{2}\log(\det(\boldsymbol{\Phi}\Gamma^N\boldsymbol{\Phi}^\top + \varsigma^2 I_n)) - \frac{n}{2}\log(2\pi). \tag{7}$$

This marginal log likelihood also has a closed form derivative with respect to each parameter Θ_i (see Rasmussen and Williams (2005)), which is given as:

$$\frac{\partial \mathcal{L}(\Theta)}{\partial \Theta_i} = \frac{1}{2}\boldsymbol{y}^\top A^{-1}\frac{\partial A}{\partial \Theta_i}\boldsymbol{y} - \frac{1}{2}tr(A^{-1}\frac{\partial A}{\partial \Theta_i}). \tag{8}$$

Here, $A = (\boldsymbol{\Phi}\Gamma^N\boldsymbol{\Phi}^\top + \varsigma^2 I_n)$ and $\frac{\partial A}{\partial \Theta_i}$ is the element-wise derivative of A with respect to Θ_i.

This closed form can be used in order to shorten the computation time of the MLE in practice. This is useful because the computation of the MLE is longer for the swaption cube than for the equity surface, since the product $\mathbf{\Phi}\Gamma^N\mathbf{\Phi}^\top$ depends on the dimension of the problem.

Remark 2. *The reason for the choice of maximizing the unconstrained likelihood instead of maximizing the constrained likelihood $\mathbb{P}(\mathbf{y} = \mathbf{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} | \boldsymbol{\xi} \in \mathcal{C}_{ineq}, \Theta)$ is discussed in Bachoc et al. (2019). In this paper, it is explained that constraining the GP increases the computational burden of the maximization, with a negligible impact on the resulting MLE.*

3.4. The Most Probable Response Cube and Measurement Noises

In the constrained GP regression case, we consider the mode of the truncated Gaussian process as an estimator for the cube instead of the mean. For the unconstrained GP regression case, mean and mode coincide because of the Gaussian profile of the conditional distribution, but the mode is easier to compute.

The maximum a posteriori (MAP) of S_w^N given the constraints satisfies the constraints on the entire domain. In the sense of Bayesian statistics, it coincides with the mode of the truncated Gaussian process. Its expression is given in Cousin et al. (2016),

$$\mathbf{M}_{S_w^N}(T, t, K) := \sum_{i=1}^{N_T}\sum_{j=1}^{N_t}\sum_{k=1}^{N_K} \hat{\boldsymbol{\xi}}^{i,j,k}\phi_{i,j,k}(T, t, K), \tag{9}$$

where $\hat{\boldsymbol{\xi}} = \left(\hat{\boldsymbol{\xi}}^{(1,1,1)}, \cdots \hat{\boldsymbol{\xi}}^{(i,j,k)}, \cdots, \hat{\boldsymbol{\xi}}^{(N_T,N_t,N_K)}\right)^\top$ is the MAP of the Gaussian coefficients $\boldsymbol{\xi}$, which satisfies the inequality constraints. In order to locate the largest noise terms, we proceed as in Chataigner et al. (2021). We compute the joint MAP$(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\varepsilon}})$ of the truncated Gaussian vector $(\boldsymbol{\xi}, \boldsymbol{\varepsilon})$. This joint MAP $(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\varepsilon}})$ is solution of:

$$\underset{\boldsymbol{\vartheta}, \boldsymbol{e}}{\operatorname{argmax}}\ \mathbb{P}\big(\boldsymbol{\xi} \in [\boldsymbol{\vartheta}, \boldsymbol{\vartheta} + d\boldsymbol{\vartheta}], \boldsymbol{\varepsilon} \in [\boldsymbol{e}, \boldsymbol{e} + d\boldsymbol{e}] \mid \mathbf{\Phi} \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{y}, \boldsymbol{\xi} \in \mathcal{C}_{ineq}\big).$$

The fact that $(\boldsymbol{\xi}, \boldsymbol{\varepsilon})$ is Gaussian with a mean 0 with the covariance matrix and the block-diagonal matrix

$$\begin{pmatrix} \Gamma^N & 0 \\ 0 & \varsigma^2 I_n \end{pmatrix},$$

implies that $(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\varepsilon}})$ is a solution to the quadratic problem:

$$\underset{\mathbf{\Phi}\cdot\boldsymbol{\vartheta}+\boldsymbol{e}=\boldsymbol{y},\,\boldsymbol{\vartheta}\in\mathcal{C}_{ineq}}{\operatorname{argmin}} \left(\boldsymbol{\vartheta}^\top(\Gamma^N)^{-1}\boldsymbol{\vartheta} + \boldsymbol{e}^\top(\varsigma^2 I_n)^{-1}\boldsymbol{e}\right). \tag{10}$$

The most probable response cube is then $M_{S_w^N}(\mathbf{x}) = \mathbf{\Phi}(\mathbf{x}) \cdot \hat{\boldsymbol{\xi}}$, for $\mathbf{x} \in \mathcal{D}$, and the most probable measurement noise is $\hat{\boldsymbol{\varepsilon}}$.

3.5. Sampling

Contrary to constrained spline interpolation approaches, the GP regression is a truly probabilistic approach. Sampling of the posterior distribution can be used for uncertainty quantification. The sampling paths of the swaption price cube consist of sampling $\boldsymbol{\xi}$

truncated on $\mathcal{C}_{ineq}$, such that $\boldsymbol{\Phi} \cdot \boldsymbol{\zeta} + \boldsymbol{\varepsilon} = \boldsymbol{y}$. The conditional distribution of $\boldsymbol{\zeta} \mid \boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{\zeta} + \boldsymbol{\varepsilon}$ is multivariate normal with mean vector μ_{cond} and covariance matrix Σ_{cond} defined as follows:

$$\mu_{\mathrm{cond}} = \Gamma^N \boldsymbol{\Phi}^\top (\boldsymbol{\Phi} \Gamma^N \boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1} \boldsymbol{y}, \tag{11}$$

and

$$\Sigma_{\mathrm{cond}} = \Gamma^N \boldsymbol{\Phi}^\top (\boldsymbol{\Phi} \Gamma^N \boldsymbol{\Phi}^\top + \varsigma^2 I_n)^{-1} \boldsymbol{\Phi} \Gamma^N. \tag{12}$$

In practice, we will sample this truncated distribution by using Hamiltonian Monte Carlo, as presented in Pakman and Paninski (2012). The starting vector we must choose has to satisfy the constraints. Thus, $\hat{\boldsymbol{\zeta}}$, the MAP of $\boldsymbol{\zeta}$ defined in Section 3.4, is an ideal candidate.

4. Empirical Results

This section compares the performance of the "constrained" GP regression with a calibrated SABR approach.

4.1. Data

For our empirical study, our data set consists of, at a quotation date t_0, 112 OTM payer swaptions, 28 ATM swaptions and 112 OTM receiver swaptions with maturities $T = [5, 10, 15, 20]$ and tenors $t = [1, 2, 5, 10, 15, 20, 30]$. There is then $N = 252$ input prices. For OTM payer swaptions, quotes are provided for the relative strikes $K_{rel_{Payer}} = [0.0025, 0.005, 0.01, 0.02]$. For the OTM receiver swaptions, we have $K_{rel_{Receiver}} = [-0.02, -0.01, -0.005, -0.0025]$. At the money, the prices of payer and receiver swaptions coincide. Each option of our data set is listed with market price, market implied normal volatility, forward swap rate, absolute strike and annuity. The annuities are computed from a yield curve, and they are built from the observed swap rates.

Absence of Arbitrages in the Data Set

The first step of our analysis is to check if the data set itself is exempted from arbitrages, as defined in Proposition 1. The in-plane triangular condition compares swaptions with the same absolute strike and different options expiries. Since the options in our data set are quoted for relatives strikes, we do not have multiple observations for the same absolute strike; then, we lack the data to test the in-plane triangular constraint. We can, however, test the monotonicity and convexity constraint in the strike direction for both payer and receiver swaptions and for each pair (Maturity, Tenor) := (T, t). In this data set, there are 28 pairs of maturities and tenors.

As stated in Table 1, there are no violations of the monotonicity constraint; therefore, the data set is exempted from call spreads arbitrage. The data set is not exempted from butterfly spreads arbitrage because, for the payer swaptions, only 67.86% of the pairs (T, t) respect the convexity constraint. Nevertheless, arbitrable observations are not removed from the data set, since the magnitude of constraint violation is relatively low[2], and we consider that these data bring valuable information to the swaption cube construction problem. In our GP methodology, we assume that violations of the no-arbitrage constraints are due to measurement noises. The idea is that even if observations may slightly violate the convexity constraints, due to, e.g., a lack of market liquidity, the predicted GP surface is guaranteed to respect this constraint everywhere on the input domain.

Table 1. Percent of pairs (T, t) respecting the absence of arbitrage in strike constraints defined in Proposition 1.

Pairs (T, t) Respecting the AOA Constraints Input Data	Monotonicity Constraint in K	Convexity Constraint in K
Payer	100%	67.86%
Receiver	100%	100%

4.2. SABR Model Benchmark

In all our study, we will use the SABR model as a benchmark regarding the prediction performance for our Gaussian process model. The SABR model has become a market standard for interpolation of swaption volatilities in the strike direction.

In this section, we briefly explain the calibration of the SABR model and how we estimate the parameters of the model.

4.2.1. SABR Model

SABR is a dynamic model in which the forward process F and the volatility process $\hat{\sigma}$ satisfy the following system of SDE:

$$\begin{cases} d\hat{F}(t) = \hat{\sigma}(t)\hat{F}(t)^{\beta}dW(t), & \hat{F}(0) = f \\ d\hat{\sigma}(t) = \alpha\hat{\sigma}(t)dZ(t), & \hat{\sigma}(0) = \sigma \\ dW(t)dZ(t) = \rho dt \end{cases}$$

where W, Z are Wiener processes with correlation $\rho \in [-1, 1]$, $\alpha \geq 0$, $0 \leq \beta \leq 1$ and $\sigma > 0$. In our study, we fix $\beta = 1$. In order to construct the swaption price cube via the SABR model, we consider as many SABR models as underlying swap, i.e., pair (T, t).

For a pair (T, t), the calibration procedure is explained in Section 4.3.4. To accommodate with negative rates, we use the so-called Shifted SABR model defined as follows:

$$\begin{cases} d\hat{F}(t) = \hat{\sigma}(t)(\hat{F}(t) + s)^{\beta}dW(t), & \hat{F}(0) = f \\ d\hat{\sigma}(t) = \alpha\hat{\sigma}(t)dZ(t), & \hat{\sigma}(0) = \sigma \\ dW(t)dZ(t) = \rho dt \end{cases}$$

where s is the shift parameter.

4.2.2. Model Calibration and Pricing

As explained in Hagan et al. (2002), an approximated analytical solution for the implied volatility exists. The volatility of a swaption with maturity T_{ex} is given as the following asymptotic approximation, where we consider $\beta = 1$:

$$\sigma_N(K) = \sigma\frac{f - K}{\log\frac{f}{K}} \cdot \left(\frac{z}{x(z)}\right) \cdot \left[1 + (-\frac{1}{24}\sigma^2 + \frac{1}{4}\rho\sigma\alpha + \frac{1}{24}(2 - 3\rho^2)\alpha^2)T_{ex}\right],$$

where

$$z = \frac{\alpha}{\sigma}\log\frac{f}{K}, \qquad x(z) = \log\left(\frac{\sqrt{1 - 2\rho z + z^2} - \rho + z}{1 - \rho}\right).$$

This model is free of arbitrage except for very low strikes and large maturities T_{ex}, for which the smile $\sigma_N(K)$ leads to negative probabilities and therefore arbitrages. This behavior of the smile for such strikes and maturities has been empirically shown in Hagan et al. (2014). The normal volatilities of the shifted SABR model are obtained by replacing

in the formula above f by $f + s$ and K by $K + s$, where s is the shift parameter, which is chosen according to the data set. In our study, we consider $s = 0.025$.

In this approach, a set of parameters $(\beta, \sigma, \alpha, \rho)$ is mapped to each distinct pair (T, t) observed. Once we fixed $\beta = 1$, the other parameters (σ, α, ρ) are calibrated by minimizing the following across the m strikes observed for the observed pair (T, t):

$$\sum_{k=1}^{m} [\sigma_N(K_k; \sigma, \alpha, \rho) - \sigma^{MARKET}(K_k)]^2, \quad m > 0. \tag{13}$$

Then, once the parameters are calibrated for each observed pair (T, t), the price of European payer and receiver swaptions $S_w(T, t, K)$ is obtained using the normal SABR volatility $\sigma_N(K)$ as input in Bachelier's pricing formula.

4.2.3. Unobserved Parameter Estimation

The previous section explained how to calibrate the parameters of the shifted SABR model for each pair (T, t), where we have at least one observation of swaptions prices. In order to reconstruct the cube, a set of parameters $(\beta, \sigma, \alpha, \rho)$ needs to be mapped to each unobserved couple (T, t). As previously, β is set to 1. Then, for each unobserved pairs (T, t), each parameter of the set (σ, α, ρ) is chosen as the weighted average of the corresponding parameter of the four nearest calibrated pairs (T, t). The weight is the inverse of the distance between the unobserved pair and the four nearest calibrated pairs. This method of interpolation and extrapolation has been chosen because it offered better results than a cubic spline approach, especially when extrapolating in-sample. From now on, this method will be referred to as the "barycentric mean" approach. One problem of such a method of interpolation and extrapolation is that the SABR model parameters $(\beta, \sigma, \alpha, \rho)$, for each unobserved pair (T, t), are estimated independently in contrary to the classical calibration. For an unobserved pair (T, t), there is then no guarantee that the resulting smile will be free of arbitrage. The absence of arbitrage of the swaption cube obtained using this method is investigated in Section 4.3.3.

4.3. *Absence of Arbitrage, Calibration and Prediction Performance*

In this section, we compare the violation of AOA, calibration and prediction performance for the cubes constructed from the constrained GP approach and the SABR benchmark approach introduced in Section 4.2.

4.3.1. Implementation of the GP Regression Approach

Since OTM payer and receiver swaptions are different contracts, with different domains of strike (absolute and relative), they need to be considered separately. For the OTM payer swaptions, the limits of the 3-dimensional grid are defined by $\mathcal{D}_{Payer}$ where $T : [5, 20]$, $t : [1, 30]$ and $K : K_{Payer}$, and the limits of the 3-dimensional grid for OTM receiver swaptions are defined by $\mathcal{D}_{Receiver}$ where $T : [5, 20]$, $t : [1, 30]$ and $K : K_{Receiver}$. For both payer and receiver swaptions, we consider a GP method with matern 5/2 kernels[3] over a $[0, 1]^3$ rescaled domain. The parameters of the kernels $[\theta_T, \theta_t, \theta_K, \sigma]$ and the homoscedastic noise variance are fitted through MLE as described in Section 3.3. Each different set of observed data will lead to a different metamodel. The MLE is maximized with the function 'minimize', with the method 'L-BFGS-B', from the Python library 'Scipy.optimize'. With the purpose of not ending up with a local minimum, the algorithm is restarted 20 timed with random starting points. In order to speed up the computation time, the Jacobian of the log likelihood is given as the input of the optimizer routine.

The grid of basis functions for constructing the finite-dimensional process S_w^N consists of *six* nodes in the absolute strike direction, *seven* nodes in the tenor direction and *six* nodes in the maturity direction.

In order to solve the quadratic problem (10), and thereby find the MAP, the Python library 'quadprog' is used. Sampling of the posterior swaption cubes distribution is conducted by Hamiltonian Monte Carlo (see Section 3.5). Its implementation is inspired by the algorithm described in Pakman and Paninski (2012).

4.3.2. Calibration and Testing Methodology

Let us now explain the methodology used for calibration and testing in practice for our GP regression model and for the SABR model.

GP regression models and the SABR model are fitted on the same randomly generated subsets of the whole data set. Each random subset is divided in two equally sized subsets: one with prices of OTM payer swaptions and ATM swaptions and the other with prices of OTM receiver swaptions and ATM swaptions. For the prediction performance, the whole dataset is chosen as the testing dataset.

For the SABR model to calibrate correctly, each pair (T, t) needs at least one ATM swaption observation, one OTM payer swaption observation, and one OTM receiver swaption observation. When a small random training set is chosen, it is possible that some pairs (T, t) end up with one or two observations; then, the calibration could lead to absurd results. To address this issue, we decided that each random training set will always contain one observation of the ATM price and at least one observation of a payer and receiver OTM price for the pairs $(5, 5)$, $(5, 10)$, $(10, 5)$ and $(10, 10)$. Those points are very liquid in the market, and quotes are then expected in practice. The details of the SABR calibration procedure are explained in Appendix A.

For the sake of comparability, the observed data for the (T, t) pairs $(5, 5)$, $(5, 10)$, $(10, 5)$ and $(10, 10)$ are also included in the training data for the GP.

All results below are produced from the following experimental procedure. We first choose a percentage p. Then, the training set is constituted of two random subsets S_1 and S_2. S_1 includes a random $p\%$ of the ATM swaptions and OTM payer swaptions, and S_2 includes a random $p\%$ of the ATM swaptions and OTM receiver swaptions. The SABR model is fitted on the union of S_1 and S_2 (there is one set of parameters $(\beta, \sigma, \alpha, \rho)$ for each pair (T, t), which is obtained through calibration or interpolation and extrapolation). The GP regression model is fitted two times, once on S_1 and once on S_2, and then, the results for the payer and receiver swaptions are gathered together. For each ATM value, the value kept is the average of the values obtained for the swaptions payer and receiver ATM if there is zero or two observations of this value. If there is only one observation, then the predicted value is taken from the output of the GP which was fitted on this observation. For each percentage of training data p we consider, random sampling, fitting, and testing is repeated 200 times. Then, the performance results are averaged and presented in the tables of Sections 4.3.3–4.3.5.

4.3.3. Absence of Arbitrage in the Predicted Swaption Cube

In this subsection, we compare the performances of the constrained GP regression model and the SABR model with respect to the arbitrage-free constraints defined in Proposition 1. The grid we use for testing is a grid of maturities, tenors such that, $T = [5, 6, \ldots, 19, 20]$, $t = [1, 2, \ldots, 29, 30]$ and of *nine* strikes K evenly spaced between the minimum and the maximum of K_{Payer} when testing the payer swaption cube, and between the minimum and maximum of $K_{Receiver}$ when testing the receiver swaption cube.

More precisely, for the monotonicity and convexity in strike constraints, we check how many pairs (T, t) among the 480 pairs of the testing grid respect the constraints. For the in-plane triangular constraint, we check the constraints on 5000 randomly selected T_1, T_2, T_3, K values in the the testing grid. Tables 2 and 3 (respectively, Tables 4 and 5) show the results for payer (respectively, receiver) swaptions.

Table 2. Respect of the no-arbitrage constraints for the predicted OTM payer swaptions priced by constrained and unconstrained GP regression, for increasing sizes of training set.

Respect of the AOA Constraints GP Payer	Decrease in Strike (Constrained)	Convexity in Strike (Constrained)	In-Plane Triangular Inequality (Constrained)	Decrease in Strike (Unconstrained)	Convexity in Strike (Unconstrained)	In-Plane Triangular Inequality (Unconstrained)
$p = 15\%$	100%	100%	100%	87.84%	8.18%	99.58%
$p = 30\%$	100%	100%	100%	93.79%	13.22%	99.97%
$p = 50\%$	100%	100%	100%	94.37%	10.24%	99.98%
$p = 70\%$	100%	100%	100%	99.55%	5.98%	99.97%
$p = 90\%$	100%	100%	100%	92.17%	1.51%	99.92%

Table 3. Respect of the no-arbitrage constraints for the predicted OTM payer swaptions priced by the chosen SABR model with barycentric mean parameter interpolation, for increasing sizes of training set.

Respect of the AOA Constraints SABR Payer	Decrease in Strike	Convexity in Strike	In-Plane Triangular Inequality
$p = 15\%$	99.00%	99.70%	98.09%
$p = 30\%$	99.49%	99.82%	97.96%
$p = 50\%$	99.82%	99.91%	99.186%
$p = 70\%$	99.96%	99.99%	99.90%
$p = 90\%$	100%	100%	100%

Table 4. Respect of the no-arbitrage constraints for the predicted OTM receiver swaptions priced by constrained and unconstrained GP regression, for increasing sizes of training set.

Respect of the AOA Constraints GP Payer	Increase in Strike (Constrained)	Convexity in Strike (Constrained)	In-Plane Triangular Inequality (Constrained)	Increase in Strike (Unconstrained)	Convexity in Strike (Unconstrained)	In-Plane Triangular Inequality (Unconstrained)
$p = 15\%$	100%	100%	99.51%	31.46%	82.67%	97.10%
$p = 30\%$	100%	100%	99.08%	49.29%	82.08%	97.37%
$p = 50\%$	100%	100%	99.53%	56.49%	86.41%	98.00%
$p = 70\%$	100%	100%	99.78%	60.03%	93.69%	97.95%
$p = 90\%$	100%	100%	99.94%	56.36%	91.74%	97.87%

Table 5. Respect of the no-arbitrage constraints for the predicted OTM receiver swaptions priced by the SABR model with barycentric mean parameter interpolation, for increasing sizes of training set.

| Respect of the AOA Constraints SABR Receiver	Increase in Strike	Convexity in Strike	In-Plane Triangular Inequality
$p = 15\%$	99.77%	98.73%	97.67%
$p = 30\%$	99.82%	99.22%	97.62%
$p = 50\%$	99.94%	99.67%	99.05%
$p = 70\%$	100%	99.94%	99.88%
$p = 90\%$	100%	100%	100%

We can observe that the prices produced by the SABR model, for a training set of size $p = 50\%$ of the total input data size or lower, do not respect the monotonicity and convexity constraints. Still, the number of pairs (T, t) violating the constraints compared to

the total number of pairs is very low. When we check for which pairs (T, t) the constraints are not respected, we see that our method for avoiding absurd parameters values during the calibration of the SABR model does not prevent arbitrage for the observed pairs. For instance, pairs (T, t) with less than three observations or no observations of one side of the smile can produce parameters leading to arbitrage. Arbitrage is also violated for unobserved pairs, for which the parameters are interpolated or extrapolated by barycentric mean. For training sets of size higher than 50% of the total input size, we can consider that the SABR model produces free of arbitrages prices.

For payer and receiver swaptions, the monotonicity and convexity constraints in strike are always ensured for prices estimated through constrained Gaussian process regression (CGPR), unlike prices predicted through unconstrained Gaussian process regression (UGPR). We can remark that receiver swaptions prices estimated through UGPR achieve 91.74% on the convexity in strike test, while payer swaptions achieve only 1.51% even when the 90% of the data set is given as training data. This difference in performance could be explained by the fact that the payer swaption database is not convex everywhere in strike.

In Section 3.2, we explain that we implement the in-plane triangular inequality constraints only for time steps of length 1 year. For payer swaptions, we see that the in-plane triangular constraint is always respected by the CGPR, and we can consider that it is also respected by the UGPR. The receiver swaptions table (Table 4) shows the impact of the 1-year step constraint. While the UGPR does not respect the in-plane triangular constraint everywhere, the CGPR produces a negligible error. Although this error is negligible, it is proof that this 1-year step constraint for the GPR is not enough to ensure that the in-plane triangular inequality is respected everywhere in the estimated price cube. In Johnson and Nonas (2009), it is mentioned that practitioners only consider the monotonicity and convexity of Proposition 1. Our empirical results support their decision. Even though our 1-year step constraint does not totally prevent in-plane triangular arbitrage, it could still be helpful for extrapolation.

Figure 1 below presents a few graphical examples of price curves for fixed tenors and maturities produced by constrained Gaussian process regression (left side) and unconstrained Gaussian process regression (right side). We can easily observe the impact of the monotonicity and convexity along the strike direction constraint. In the bottom row of Figure 1, we can notice that the mean relative error of the unconstrained GP regression is lower than the one of the constrained GP regression. If we observe those two graphs, we can see that the unconstrained GP price crosses the option data price with the higher relative strike while the constrained GP price does not. The constrained GP regression forces the path of the mode of the GP to be convex and increasing/decreasing in strike. Since the payer data set is not convex everywhere, it is not surprising that the convexity constraint increases the testing error.
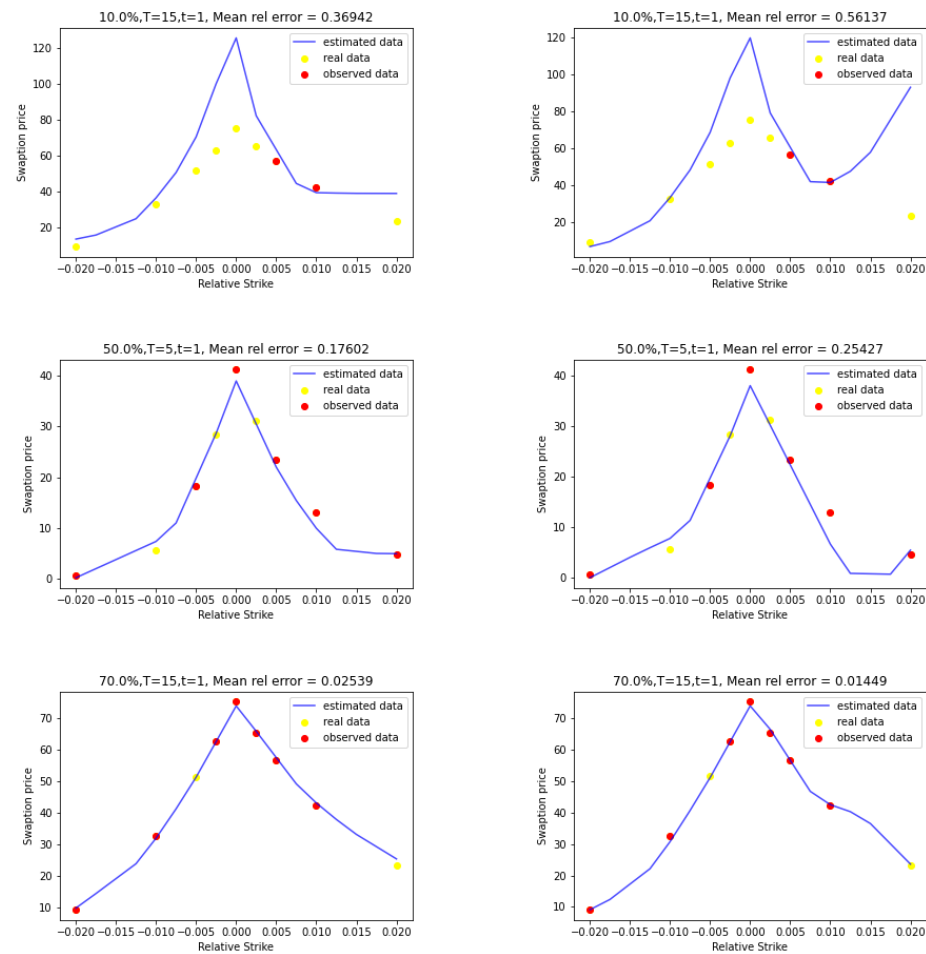
Figure 1. Constrained GP (**left**) and unconstrained GP (**right**) price curves for $p = 10\%$ (**top**), $p = 50\%$ (**middle**) and $p = 70\%$ (**bottom**).

4.3.4. Calibration Performance

Table 6 shows the calibration performance of our models.

Table 6. Mean relative calibration error and calibration RMSE of the constrained GP regression and SABR model with barycentric mean parameter interpolation, for increasing sizes of training set.

Average Calibration Error	Constrained GP Mean Relative Error	SABR Mean Relative Error	Constrained GP RMSE	SABR RMSE
$p = 15\%$	1.77%	1.61%	2.68	12.65
$p = 30\%$	2.29%	1.47%	2.78	10.56
$p = 50\%$	2.67%	2.21%	3.43	11.67
$p = 70\%$	3.02%	2.87%	3.88	13.85
$p = 90\%$	3.25%	3.22%	4.11	14.59

Based on the mean relative error metric, the SABR model calibrates better than the constrained GP model. That was to be expected, because the constrained GP regression assumes that the observations are noisy in order to avoid over-fitting and to respect the constraints of no arbitrage. Since the SABR model is not constrained, it gives better mean relative error. We can observe that the mean relative error increases with the size of the training set, and that the error of the GP regression model converges to the error of the SABR model. However, the constrained GP model average RMSE is lower than the average

RMSE of the SABR model on all percentages p tested. This means that the estimated standard deviation of the distribution of the absolute errors of the SABR model is on average higher than the one of the constrained GP model.

4.3.5. Prediction Performance

Table 7 shows the prediction performance of our models.

Table 7. RMSE of the constrained GP regression and SABR model with barycentric mean parameter interpolation, for increasing sizes of training set.

Average Prediction Error	Constrained GP RMSE	SABR RMSE
$p = 15\%$	29.01	172.42
$p = 30\%$	27.83	58.55
$p = 50\%$	9.98	33.25
$p = 70\%$	6.30	20.57
$p = 90\%$	4.71	16.54

Based on the mean RMSE metric, the constrained GP regression performs better than the SABR model for all percentages. For larger and larger training sets, the two models perform better and better in prediction with respect to the RMSE.

The complexity of the GP regression is dominated by the inversion of the matrix of size (N_{obs}, N_{obs}), where N_{obs} is the number of observations in S_1 when we fit the payer swaption cube or the the number of observations in S_2 otherwise. The total complexity when producing the cube is then a $O((N^3_{obs_{payer}} + N^3_{obs_{receiver}}) \times L_1)$, where L_1 is the computation cost of the function 'minimize' with 20 restarts and with the input of the Jacobian matrix of the MLE, defined in (8), as it significantly speeds up the minimization. The complexity of the calibration of the SABR model is $O((N_{obs_{payer}} + N_{obs_{receiver}}) \times L_2)$ where N_{obs} is the total number of observed data and L_2 is the computation cost of the function 'minimize' with 20 restarts and without inputing a Jacobian[4]. Thus, one of the main drawbacks of GP-based interpolation is the computer time that can be quite substantial for a large dataset.

4.3.6. SABR Model Calibrated on the GP Predicted Cube

Instead of calibrating the SABR model directly on the observations, it is possible to use the GP-predicted cube as input. Table 8 shows a significant improvement with respect to the RMSE between the two methods, especially for low percentage of input data. Indeed, for these ones, the SABR model cannot be calibrated properly because the smile is not well enough described. In fact, we need for each tenor and maturity at least one receiver OTM and one payer OTM to define the skew rho and the convexity alpha, plus a point near the ATM to define the model volatility sigma. As the percentage decreases, this condition is less satisfied. Thus, by using the GP cube, we provide an idea of the whole smile and get around the issue. We can see that the gain for $p = 15\%$ is lower than for $p = 30\%$. This can be explained by the fact that the performance of the GP model deteriorates when there are too few observations. Tables 9 and 10 provide a comparison of the performance of the SABR model and GP-SABR model with respect to the no arbitrage constraints.

Table 8. Average gain in RMSE when calibrating the SABR model with the output of the GP regression model.

Percentage of Data Used as Input	*p* = 15%	*p* = 30%	*p* = 50%	*p* = 70%	*p* = 90%
RMSE relative gain	63%	68%	59%	35%	19%

Table 9. Respect of the no-arbitrage constraints for the predicted OTM payer swaptions priced by the SABR model calibrated on observations or on the output of the GP model, for increasing sizes of training set.

Respect of the AOA Constraints Payer	Decrease in Strike (SABR)	Convexity in Strike (SABR)	In-Plane Triangular Inequality (SABR)	Decrease in Strike (GP-SABR)	Convexity in Strike (GP-SABR)	In-Plane Triangular Inequality (GP-SABR)
p = 15%	99.45%	99.80%	98.51%	99.90%	99.98%	100%
p = 30%	99.68%	99.98%	96.63%	100%	100%	100%
p = 50%	99.75%	99.90%	99.37%	100%	100%	100%
p = 70%	100%	100%	100%	100%	100%	100%
p = 90%	100%	100%	100%	100%	100%	100%

Table 10. Respect of the no-arbitrage constraints for the predicted OTM receiver swaptions priced by the SABR model calibrated on observations or on the output of the GP model, for increasing sizes of training set.

Respect of the AOA Constraints Receiver	Increase in Strike (SABR)	Convexity in Strike (SABR)	In-Plane Triangular Inequality (SABR)	Increase in Strike (GP-SABR)	Convexity in Strike (GP-SABR)	In-Plane Triangular Inequality (GP-SABR)
p = 15%	99.87%	99.20%	98.22%	99.99%	99.53%	99.52%
p = 30%	99.88%	99.40%	96.43%	100%	99.76%	99.97%
p = 50%	99.88%	99.70%	99.10%	100%	99.80%	100%
p = 70%	100%	100%	100%	100%	99.80%	100%
p = 90%	100%	100%	100%	100%	99.80%	100%

4.4. Uncertainty Quantification

By sampling the truncated Gaussian distribution described in Section 3.5, it is possible to obtain uncertainties for all the prices estimated through GP regression. The sampling is performed by Hamiltonian Monte Carlo, using the mode $\hat{\xi}$ as the starting vector. In practice, we sample the distribution 5000 times. Therefore, we obtain 5000 GP paths of the swaption price cube, all respecting the constraints. These samplings could be used to generate quantiles and calculate model error due to uncertainty. Uncertainties due to interpolation/extrapolation could also be propagated into rate models to see the impact on exotic products pricing. For example, we could calibrate an HJM model with "a percentile 90" cube swaption and "a percentile 10" cube swaption; then, we price a basket of bermuda options with these two models and study the differences (AVA model).

Figure 2 shows examples of price curves produced by constrained and unconstrained GP regression alongside their empirical pointwise quantiles. We can observe that adding observations reduces the size of the pointwise confidence intervals. We can also observe that unconstrained GP regression leads to wider pointwise confidence intervals. In Figure 2(right-top), the 5% empirical quantile of the posterior of the unconstrained GP is even negative, which is not allowed for prices.
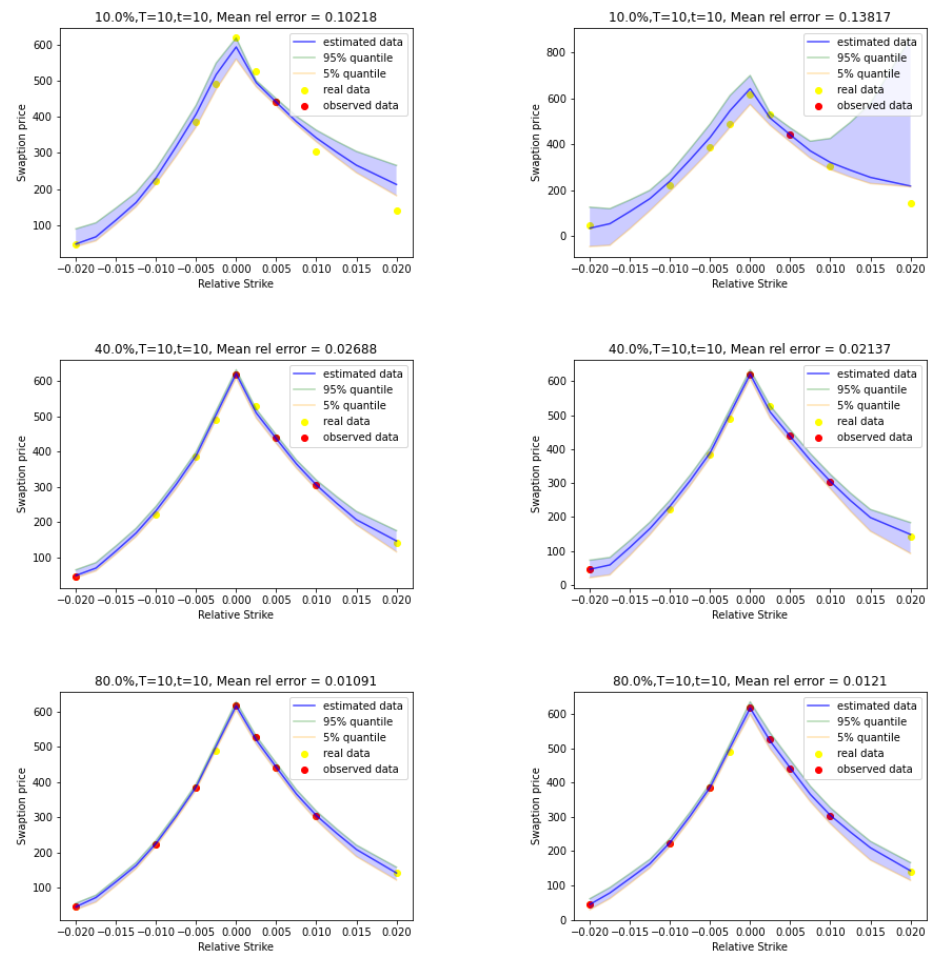
Figure 2. Constrained GP (**left**) and unconstrained GP (**right**) price curves with uncertainties for $p = 10\%$ (**top**), $p = 40\%$ (**middle**) and $p = 80\%$ (**bottom**).

5. Conclusions

In this paper, we introduce a methodology to construct a free of arbitrage swaption price cube based on GP regression. We apply this methodology on a data set of ATM, OTM payer and OTM receiver swaptions. We show that the constrained GP regression leads to arbitrage-free swaption cubes along the strike direction for any size of the training data. Even though the practitioners were usually satisfied by the fact that there are no call spreads and butterfly spreads, we go further by considering the in-plane triangular inequality introduced by Johnson and Nonas (2009). We empirically demonstrate that imposing a finite-dimensional weaker form of this condition is enough for the GP to generate swaption cubes with a negligible proportion of violation points, even for a small training set. We then compare the constrained GP model to a state-of-the-art SABR model fitted on the same data set. We show that the constrained GP performs better in prediction than the SABR model. The predictive performance of SABR seems to improves significantly when the model is calibrated on the GP-predicted cube. Finally, the GP approach is able to quantify in- and out-of-sample uncertainty through Hamiltonian Monte Carlo simulations, allowing for computation of model risk Additional Valuation Adjustment.

6. Further Research

A topic left to be addressed is the update of GP hyperparameters when a new quote streams in. A straightforward solution could be to use the last estimation of the maximum log-likelihood as the new starting point of the optimizer routine. However, since the computation cost of the GP regression increases when the number of observation increases, this solution could be intractable in practice. For this problem, Bayesian Optimization could be considered. As explained in Shahriari et al. (2016), Bayesian Optimization consists of selecting the $n + 1$-th set of hyperparameters, based on an acquisition function computed with the current posterior distribution of the likelihood function. Yin et al. (2019) present another way of selecting the $n + 1$-th point, based on a mean square error criteria.

Appendix A. Calibration of the SABR Model

The calibration procedure we applied is as follows:

1. Calibration of the parameters of the SABR model for the pairs $(5, 5)$, $(5, 10)$, $(10, 5)$ and $(10, 10)$.
2. For all the other pairs, take as a starting value for the parameters α, ρ and σ the value of the calibrated parameters of the nearest neighbor among the pairs $(5, 5)$, $(5, 10)$, $(10, 5)$ and $(10, 10)$.
3. Then, for the pairs (T, t), calibrate:

    (a) α, ρ and σ if there is at least three observations.
    (b) α only if there is one or two observations of payer swaptions OTM (or receiver swaptions OTM).
    (c) σ only if the only observation is the price of a swaption ATM.
    (d) ρ and α if there is an observation of a payer swaption OTM and an observation of a receiver swaption OTM.
    (e) α and σ if we observed one swaption ATM and one payer or receiver swaption OTM.

The calibration of the parameters of the model is made through the function 'minimize' introduced in Section 4.3.1, with the same number of random restarts.

Notes

1   This choice has been made for the sake of clarity. It is also possible to consider a grid with heterogeneous steps.
2   For the observed pairs (T, t) that do not respect the convexity constraint, the violation errors range from 0.4% to 4% of the average swaption price involved in the convexity constraint.
3   A 1-dimensional Matern 5/2 kernel is given as $c(\mathbf{x}, \mathbf{x}') = \sigma^2(1 + \frac{\sqrt{5}|\mathbf{x}-\mathbf{x}'|}{\theta} + \frac{5(\mathbf{x}-\mathbf{x}')}{3\theta^2})\exp(-\frac{\sqrt{5}|\mathbf{x}-\mathbf{x}'|}{\theta})$.
4   To give an idea, for $p = 90\%$, the computation time for both methods takes less than 30 seconds on a laptop.

References

Bachoc, François, Agnès Lagnoux, and Andrés F. López-Lopera. 2019. Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic Journal of Statistics* 13: 2921–69. [CrossRef]

Chataigner, Marc, Areski Cousin, Stéphane Crépey, Matthew Dixon, and Djibril Gueye. 2021. Beyond surrogate modeling: Learning the local volatility via shape constraints. *SIAM Journal on Financial Mathematics* 12: SC58–SC69. [CrossRef]

Cousin, Areski, Hassan Maatouk, and Didier Rullière. 2016. Kriging of financial term-structures. *European Journal of Operational Research* 255: 631–48. [CrossRef]

Hagan, Pat, and Michael Konikov. 2004. Interest Rate Volatility Cube: Construction and Use. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=16e6054b918a1facbb1fed303d889865e8b0cfba (accessed on 1 February 2022).

Hagan, Patrick S., Deep Kumar, Andrew Lesniewski, and Diana Woodward. 2014. Arbitrage-free SABR. *Wilmott* 69: 60–75. [CrossRef]

Hagan, Patrick S., Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. 2002. Managing smile risk. *The Best of Wilmott* 1: 249–96.

Johnson, Simon, and Bereshad Nonas. 2009. Arbitrage-free construction of the swaption cube. *Wilmott Journal* 1: 137–43. [CrossRef]

Maatouk, Hassan, and Xavier Bay. 2017. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences* 49: 557–82. [CrossRef]

Pakman, Ari, and Liam Paninski. 2012. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics* 23: 518–42. [CrossRef]

Rasmussen, Carl Edward, and Christopher K. I. Williams. 2005. *Gaussian Processes for Machine Learning*. Cambridge: The MIT Press, vol. 11.

Ruf, Johannes, and Weiguan Wang. 2019. Neural networks for option pricing and hedging: A literature review. *arXiv*, arXiv:1911.05620.

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE* 104: 148–75. [CrossRef]

Skantzos, Nikos, and George Garston. 2019. The Perfect Smile Filling the Gaps in the Swaption Volatility Cube. Available online: https://www2.deloitte.com/content/dam/Deloitte/be/Documents/risk/deloitte-be-the-perfect-smile.pdf (accessed on 1 February 2022).

Yin, Hong, Jingjing Ma, Kangli Dong, Zhenrui Peng, Pan Cui, and Chenghao Yang. 2019. Model updating method based on kriging model for structural dynamics. *Shock and Vibration* 2019: 8086024. [CrossRef]