Research Internship of MAP595
Report

Simultaneous upper and lower bounds of American option prices

Submitted by

AI Xiaoyun

Internship Supervisor at Société Générale

BENNEHHOU Sanaa
FELLAJI Mohammed

20 March 2023 - 31 July 2023

Contents

List of Figures

Chapter 1

Introduction

Developing arbitrage-free implied volatility surfaces from bid-ask price quotes is a key challenge in the financial industry, especially for pricing exotic options. These surfaces are essential in risk-neutral models, which rely on vanilla option prices for calibration. Without an accurate, arbitrage-free volatility surface, pricing models may suffer from inconsistencies, leading to mispricing and unreliable risk management. To address these challenges, various methods have been proposed in the literature. In this study, we explore a non-parametric calibration method based on the theory of martingale optimal transport (MOT). This method has proven effective in producing arbitrage-free implied volatility surfaces, fitting market quotes within bid-ask spreads, and accurately capturing complex volatility smile patterns, such as Mexican hat-shaped curves.

Despite the limitations of the Black-Scholes-Merton (BSM) model in reflecting real market conditions, it remains widely used due to its practicality in transforming option prices into implied volatility, a single value that provides valuable market insights. Implied volatility helps practitioners assess market sentiment and risk.

When implied volatilities are plotted against strike prices for a given maturity, common patterns like skew or smile often appear. These patterns reflect deviations from normality in the underlying return distribution: a smile indicates fat tails, while a skew signals asymmetry.

Variations in the implied volatility smile across different strikes and maturities provide insights into how the return distribution evolves over time. This makes the collection of vanilla option prices across different strikes and maturities critical for financial applications such as model calibration, pricing, and risk-neutral density (RND) estimation. However, if these price data contains arbitrage, it can result in inaccurate model calibration and erroneous RND estimates. Most derivative pricing models are designed to be arbitrage-free to avoid economically unrealistic outcomes. For example, models like Dupire's local volatility model fail to calibrate correctly when the data contains arbitrage, and arbitrage-free interpolation methods also break down under these conditions.

To improve the accuracy and robustness of pricing models, it is essential to remove arbitrage from the data. Two common techniques for achieving this are smoothing and filtering, and in this study, we focus on smoothing methods. Significant contributions to smoothing have been made by Ait-Sahalia and Duarte, Fengler, Gatheral, and Jacquier. In recent years, stochastic volatility models (SVMs) have gained a lot attention in this context. In practice, smoothing typically involves calibrating models to create a smooth $C^{1,2}$ call price function $(T, K) \mapsto C(T, K)$, or equivalently, an implied volatility function $(T, K) \mapsto \sigma_{\text{imp}}(T, K)$, which results in arbitrage-free surfaces.

While existing calibration methods offer advantages, they also have some limitations. For instance, SVM parameter calibration can be time-consuming, especially without a closed-form solution, and the non-convex optimization involved may not always converge. Additionally, although parametric models like SVI are easy to implement, they do not always ensure arbitrage-free surfaces across both strike prices and maturities.

Instead of using parametric calibration models, this study addresses the arbitrage-free calibration problem with a non-parametric discrete-time model based on martingale optimal transport (MOT) theory. Optimal transport theory has recently attracted significant attention. Initially introduced by Monge [20] in civil engineering, it was later formalized with a mathematical foundation by Kantorovich [19], who introduced the dual problem framework. In recent years, optimal transport theory has been applied to robust hedging and pricing in both discrete and continuous-time models [2, 17, 21, 7]. It was discovered that pricing bounds for path-dependent derivatives using fixed European options could be framed as a martingale optimal transport problem. This led to MOT becoming a field of study in its own right, with the theory further applied to the non-parametric discrete-time model proposed by Guyon [14], particularly to address the VIX/SPX calibration problem.

This approach is equivalent to identifying an element from the non-empty set of probability measures $\mathcal{M}$. To construct this model, we begin by selecting a prior probability $\bar{\mu} \in \mathcal{P}$ based on market information and then search for the model μ within $\mathcal{M}$ that is closest to $\bar{\mu}$ in terms of entropy, minimizing the relative entropy with respect to $\bar{\mu}$, using the Kullback-Leibler divergence in the objective function. Several dual versions of the problem can be formulated, and Guyon demonstrated that one of these has a clear financial interpretation through exponential utility indifference pricing. He also proved there is no duality gap between the primal and dual problems. The minimum entropy model is explicitly characterized by the minimizers of the dual problem.

More specifically, the minimum entropy model is described non parametrically by its Radon-Nikodym derivative with respect to $\bar{\mu}$, expressed as the exponential of a combination of payoffs from the underlying instruments. In the context of equities, this portfolio consists of the same hedging instruments used in the super-replication problem. To numerically solve the dual problem, we applied the Sinkhorn algorithm [81], which was revived by Cuturi [24] for classical optimal transport and later adapted by De March and Henry-Labordère [26] for fast construction of arbitrage-free smile interpolations. The algorithm iteratively optimizes in different directions until it converges at the global optimum, where the gradient of the dual objective function vanishes.

It is worth noting that the dual formulation and another numerical method were explored much earlier in [1] within the context of uncertain volatility calibration using entropy minimization. Additionally, calibration within bid-ask spreads was investigated through the introduction of an L^2 weighted penalization [2]. However, at that time, the connection to optimal transport and the proof of the duality result were not considered.

Our study is largely inspired by the work of De March, Henry-Labordère, and Guyon. However, neither De March nor Henry-Labordère mentioned the property of eliminating arbitrage. Regardless of the input data, the posterior density, if it exists, is truly arbitrage-free due to the positivity ensured by the minimum entropy probability formula. In Guyon's study, he applied strict equity constraints on prices and showed that if there are arbitrages in the price data, the objective function value will approach infinity and the algorithm will not converge. However, similar to the work of Allenaveda, De March, and Henry-Labordère, if we consider bid-ask quotes and apply L^2 penalization, a solution can still be found even if some arbitrage exists in the input price data.

Another difference is that we applied this approach in the realm of interest rate derivatives, rather than equities, where calendar arbitrage is not considered. Towards the end of our study, we explored joint calibration using both vanilla swaption prices and CMS forwards, where the latter product's price depends on the entire market distribution, making the calibration problem more complex.

The rest of this paper is structured as follows: After introducing our discrete-time setting in Section 2, we discuss how to formulate the arbitrage-free calibration problem in Section 3. In Section 4, we derive the Sinkhorn algorithm used to numerically solve these problems, along with some implementation tricks. Section 6 presents the results of several experiments, and Section 7 concludes.

Chapter 2

Setting and notation

In this study, we operate within the multicurve interest rate market, focusing primarily on two key interest rate indices. The Euro Interbank Offered Rate (Euribor, EIB) reflects the 1-month, 3-month, 6-month, and 12-month interest rates at which banks lend unsecured funds to each other within the eurozone. Additionally, the eurozone's risk-free rate, the Euro Short-Term Rate (ESTER, EST), will also be considered.

In terms of the instruments used in this study, the Zero-Coupon bond (ZC) is an instrument that pays 1 unit of currency at maturity. The ZC curve, also known as the discount curve, is currently dependent solely on the risk-free rate, i.e., the EST. We denote $B(t, T)$ as the present value at time t of a ZC bond that pays 1 unit of currency at time T. An interest rate swap is a contract between two parties in which they agree to exchange a predetermined fixed interest rate for a floating interest rate (potentially based on different rate indices) at regular intervals over an agreed period, based on a notional amount. The most common swap product in the market is the plain vanilla swap, typically structured as an exchange of fixed interest rate payments for floating rate payments. For the fixed leg of the swap, all cash flows are determined at the time the swap is initiated. The cash flow at time T_i is calculated by multiplying the accrual factor δ_i^f, which corresponds to the period $[T_{i-1}^f, T_i^f]$, by the agreed fixed rate. For the floating leg, cash flows are determined throughout the life of the swap. On each payment date, starting at T_1^v, the cash flow at T_i^v is calculated at T_{i-1}^v, where the accrual factor δ_i^v is multiplied by the floating index rate $I(T_{i-1}^v, T_{i-1}^v, T_i^v)$, which represents the floating rate for the period $[T_{i-1}^v, T_i^v]$ as observed at time T_{i-1}^v. A forward swap, on the other hand, is an agreement made at time t to enter into a swap starting at T_0 and ending at T_n in the future. The forward swap rate is the fixed rate agreed upon between the counterparties when entering the forward swap, such that the present value of the fixed and floating legs of the forward swap are equal. In other words, it is the fixed swap rate that makes the present value of the swap equal to zero. For a forward swap starting at a future date T_0, the forward swap rate agreed upon at present (time t) is denoted as:

$$R(t, T_0, T_n) = ...$$

If the floating rate I and the Zero-Coupon bond B are based on the same index, i.e., if I is also based on the EST, we can simplify and express the forward swap rate as:

$$R(t, T_0, T_n) = ...$$

A swaption is an option where the underlying security is a forward swap. In line with swap terminology, a call swaption is also referred to as a payer swaption, while a put swaption is called a receiver swaption. Even though multiple yield curves or indices may exist for the same currency, they share the same volatility surface. This volatility surface is calibrated using market prices of swaptions. Using the notion of the swap rate, we can express the present value of a payer swaption with a fixed strike rate K as:

$$\mathbf{PV}(t) = LVL(t, T_0, T_n)(R(t, T_0, T_n) - K)_+$$

where

$$LVL(t, T_0, T_n) = ...$$

represents the level or PVBP (Price Value of a Basis Point) of the swap. According to martingale pricing theory, using LVL as the numeraire, and after performing the appropriate probability change, the price of a payer swaption with an expiry date of T and strike K is given by:

$$C_t(K, T) = LVL(t, T_0, T_n)\mathbb{E}_t^{Q^{LVL}}[(R(T, T_0, T_n) - K)_+],$$

where $LVL(t, T_0, T_n)$ represents the level or Price Value of a Basis Point (PVBP) for the underlying swap, and $\mathbb{E}_t^{Q^{LVL}}$ denotes the expectation under the forward measure Q^{LVL}, which corresponds to the numeraire LVL. Additionally, under this probability measure Q^{LVL}, the swap rate $R(T, T_0, T_n)$ behaves as a martingale.

A Constant Maturity Swap (CMS) is a type of interest rate swap where the floating leg is linked to a swap rate with a fixed maturity, rather than a standard money market rate like Euribor. Building on this concept, a CMS forward is a forward contract based on the future value of a CMS rate. Similar to a traditional forward rate agreement (FRA), a CMS forward locks in the swap rate for a future period. However, instead of focusing on short-term interest rates, a CMS forward specifically targets the constant maturity swap rate, enabling market participants to hedge or speculate on the movement of long-term swap rates over a defined period. The price of a CMS forward encapsulates all relevant information about the payer and receiver swaptions in the market, and this can be understood from two perspectives. First, the pricing formula for CMS instruments can be expressed as an integral over the distribution of vanilla swaptions. Second, the CMS forward rate reflects the market's expectations of future swap rates, which are influenced by the pricing of both payer and receiver swaptions. These swaptions capture the market's view on interest rate volatility and directional movements, making the CMS forward price a comprehensive measure that incorporates implied volatility and risk-neutral probabilities derived from swaption markets. From a probabilistic perspective, the price of the CMS forward can be written as:

$$\mathbb{E}^{Q_T}[R(T, T_0, T_n)],$$

where Q_T is the forward measure associated with the maturity T, and $R(T, T_0, T_n)$ is the forward swap rate at time T. Under the swap measure Q^{LVL}, this expression can be transformed as:

$$\mathbb{E}^{Q_T}[R(T, T_0, T_n)] = \frac{LVL(0)}{B(0,T)}\mathbb{E}^{Q^{LVL}}\left[\frac{R(T, T_0, T_n)}{LVL(T)}\right].$$

Chapter 3

Calibration problem in MOT formulation

3.1 Arbitrage constraints

Arbitrage refers to a costless trading strategy that offers a positive probability of earning a risk-free profit. There are two main types: static and dynamic arbitrage. Static arbitrage involves exploiting opportunities through fixed positions in options and the underlying stock at the initial time, with the option to adjust the underlying position at a finite number of trading times in the future. In contrast, dynamic arbitrage relies on the dynamic and path-dependent properties of tradable assets, allowing continuous adjustments over time. Given that our method is model-independent, we focus solely on static arbitrage, where no modeling of asset dynamics is required.

Initially, we do not incorporate the price of the CMS forward. From the perspective of linear programming, by treating the swaption prices at different strikes as discrete data points, we can establish the arbitrage-free shape constraints for the call price surface. This ensures that the call price surface remains consistent with the principles of no-arbitrage across the range of strike prices. Since there is no concern for calendar arbitrage in our context, we focus on the following conditions to ensure arbitrage-freeness in our case.

A model $\mathbb{M}$ is characterized as a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{T}}, \mathbb{P})$ that includes an adapted diffusion process $\{(R_t, \mathbf{C}_t)\}_{t\in\mathcal{T}}$. In this framework, $\mathbf{C}_t$ denotes the prices of $N = |\mathcal{K}|$ options at time t, and we initially observe $\mathbf{C}_0$. The set $\mathcal{T}$ signifies the discrete trading times for the underlying asset, which in this case is $\mathcal{T} = \{0, T_f\}$. The First Fundamental Theorem of Asset Pricing (FFTAP) establishes a fundamental link between the no-arbitrage principle (both static and dynamic) and the existence of an equivalent martingale measure (EMM). Since the seminal work of Harrison and Kreps [22], different formulations of the FFTAP and broader concepts of no-arbitrage have emerged. In general terms, for a given model $\mathbb{M}$, there is no arbitrage if and only if $\exists \mathbb{Q} \sim \mathbb{P}$, such that

$$\forall (T, K) \in (\mathcal{T} \times \mathcal{K}),\ D(t)C_t(T, K) = D(s)\mathbb{E}^{\mathbb{Q}}[C_s(T, K)|\mathcal{F}_t]$$

for all $t < s \leq T$, where $t, s \in \mathcal{T}$. Consequently, static arbitrage constraints result from the relationships between terminal payoffs, discounted to the present moment.

Recall that, under the swap measure, the price of a payer swaption is given by: ..., and in this context, the forward swap rate behaves as a martingale. Hence, our goal—constructing an arbitrage-free implied volatility model—is equivalent to finding a martingale probability measure $\mu^* \in \mathcal{P}(\{R_0\} \times \mathbb{R})$ that precisely matches these market prices. These probabilistic constraints can be expressed as: ....

Furthermore, the equivalence between the shape constraints on prices and the constraints on the probabilistic measure has been established.

Another criterion for checking arbitrage involves plotting the RND. According to Breeden and Litzenberger, the second partial derivative of the call price with respect to the strike price represents the RND of R_1, up to a discount factor. Specifically, it can be expressed as:

$$\text{pdf}(R_1) = \frac{1}{LVL(0)} \frac{\partial^2 C_t}{\partial K^2},$$

To ensure the absence of arbitrage, the RND must be a true probability density function, meaning it should not take on negative values. We will further present this as existence of arbitrages in the experiments chapter.

3.2 MOT problem formulation

To build such an element belonging to $\mathcal{M}$, we begin by selecting a prior probability measure m_0 on $\mathcal{P}(R_0 \times \mathbb{R})$, which represents the space of all possible measures on the pair (R_0, R_1). We then seek a measure $\mu \in \mathcal{M}$ that minimizes the "distance" to m_0, as illustrated in the accompanying graph.

The "distance", in more rigorous mathematical expression, is the objective function $\mathcal{F}$, so our objective is to solve the optimization problem $inf_{\mu \in \mathcal{M} \mathcal{F}_{m_0}}(\mu) = \mathcal{F}_{m_0}(\mu^*)$.

3.3 CMS forward joint calibration problem

Algorithm 1: Example Algorithm

```
Data: Input data
Result: Output result
Initialization;
for i ← 1 to n do
    if x < i then
        x ← x + i;
    end
end
```

Chapter 4

Numerical implementation details

Chapter 5

Experiments and results

Chapter 6

Conclusion