Stochastics and Financial Mathematics MSc
MASTER THESIS

Conic Swaption Pricing with Displaced SABR

Author: Loizos Raounas

Supervisor: prof. dr. Asma Khedher

Examination date: October 17, 2018

Daily supervisor: dr. Jovanka Lukič
Abstract

Conic Finance is a recently formulated financial theory, in which bid and ask prices are modelled and priced directly, as distorted expected values through some concave distortion function. In this study, we apply this theory in order to price the bid and ask values of USD and EUR swaptions, using the stochastic volatility model SABR. Our results support the notion of Conic Finance as a solid financial framework, which not only does not contradict the classical risk-neutral framework, but in fact fully incorporates it as a special case.

**Keywords:** Conic Finance, Swaptions, bid and ask pricing, distortion function
Στην Μητέρα.
Introduction

The traditional “one-price” framework in finance dictates that the arbitrage-free price, calculated using a risk-neutral measure, is the fair price used for all transactions, independent of direction (buying or selling). However, this is not true in general, as the bid-ask spread i.e. the difference between the highest available purchasing price and the lowest available selling price is not always negligible. Some of the factors that affect the bid-ask spread upwards, are market illiquidity and high volatility [11].

The Conic Finance theory introduced by Cherny and Madan [30] in 2010, operates under different assumptions. Conic Finance accepts that in a modern economy not all risks can always be eliminated and as such the “set of acceptable risks must be defined as a financial primitive of the financial economy” [29]. To this end, it departs from the risk-neutral “one-price” framework and instead considers two values of importance, the bid price and the ask price, which are valuated separately. This is done by distorting the risk-neutral measure in a way that it reflects the market direction, applying different weights to different outcomes. Hence, the Conic Finance Framework does not contradict the classical framework, but in fact extends and contains it as a special case, while being able to provide a solid mathematical explanation for the spread between bid and ask values observed in the market, as well as providing a more conservative pricing approach.

This thesis deals with the bid and ask value pricing of cash settled (EUR) and swap settled (USD) swaptions, using the displaced SABR model, under Conic Finance assumptions.

Swap settled swaptions are derivatives that give their holder the right (but not the obligation) to enter an interest rate swap with another party, i.e. to exchange payments based on a fixed and a floating interest rate on a predetermined notional for a predetermined amount of time. On the other hand, cash settled swaptions replace these multiple cash flows with a single payment, which is based on the forward value of the swap rate. For more on interest rate swaps and swaptions, we refer to [24]. SABR is a stochastic volatility model, introduced by Hagan et al. [19]. Among its advantages is that it correctly predicts the market observed dynamics of the volatility smile (implied volatility with respect to strikes) and that there is an analytic formula for the implied volatility.

The thesis is structured as such: in the first chapter, we give an introduction to Conic Finance. To do so, we start by starting a liquid complete market, in which case Conic Finance coincides with the classical framework and then move on to assuming an illiquid market. Then, the rule of one price fails, as the bid-ask spread becomes non-negligible and we introduce the rule of “two-prices” through the use of coherent risk measures functionals, as seen in [29]. We derive an analytic formula for both the bid and ask price, by expressing them as Choquet integrals (see [10], [13] and [17]), which may be interpreted as distorted expectations. We then present some examples of parametric functions that may be used in order to distort the expectation. We conclude by presenting the Conic Finance adaptation of two widely used pricing methods, namely the Binomial Trees method and Monte Carlo.
In the second chapter we describe the products that are priced in this thesis (EUR and USD swaptions), as well as the model used to simulate their underlying value, i.e. the Displaced SABR model. This model was chosen because of its ability to handle negative interest rates, an event previously considered impossible, yet a reality in the current, post-crisis, low interest rate environment. In the third chapter, we describe our given data and the methodology we have followed, for the parameter calibration and underlying simulation. Namely, for the parameter calibration, we used (deterministic) curve fitting and the (stochastic) Differential Evolution algorithm, introduced by Storn & Price [37]. For the underlying simulation, we employed the Euler Discretization and the “(Semi) Exact SABR” approach, as suggested by Cai et al. [7].

Continuing, in the fourth chapter we present our results. We begin with a comparison of the calibration methods applied, followed by a comparison of the efficiency of several distortion functions. Afterwards, we briefly discuss about the possibility of interpolating the distortion parameter $\lambda$ for swaptions with no market quoted bid/ask values and conclude that the Conic Finance approach may predict missing bid/ask prices with acceptable accuracy, even when there are missing bid/ask market quoted prices. Then, we present residuals plots of multiple swaption bid/ask values and compare the two underlying simulation models employed. Finally, in the Discussion chapter we wrap up our findings and suggest topics for further research.
1 Conic Finance

In this chapter we briefly describe the Conic Finance framework, which was first introduced by Cherny & Madan [30], highlighting its differences with the classic (“one-price”) model used by market participants. We start by defining a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), which we assume to be atomless\(^1\) or, equivalently, to support a random variable with a continuous distribution [9]. In this context, \(\mathbb{P}\) is called the “historical” or “real-world” probability measure. To avoid potential technical difficulties, we restrict ourselves to the space of bounded random variables \(L^\infty(\Omega, \mathcal{F}, \mathbb{P})\). Let \(0 \leq X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})\) represent a random payoff at a future time \(T\). From here onwards, we call \(X\) a risk. We say that a set \(C\) of risks is convex, if for every \(X, Y \in C\) and for every \(0 \leq \alpha \leq 1\), it holds \(\alpha X + (1 - \alpha)Y \in C\). Furthermore, we call \(C\) a cone if for every \(X \in C\) and \(c > 0\), it holds that \(cX \in C\). Central to what follows is the notion of acceptability sets, as defined in Artzner et al. [3], of which we give the (axiomatic) definition.

Definition 1.1. A set \(A \subset L^\infty(\Omega, \mathcal{F}, \mathbb{P})\) is called an acceptability set if:

1. \(L^+ \subset A\), where \(L^+ := \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : X(\omega) \geq 0, \forall \omega \in \Omega\}\).
2. \(A \cap L^- = \emptyset\), where \(L^- := \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : X(\omega) < 0, \forall \omega \in \Omega\}\).
3. \(A\) is convex.
4. \(A\) is a positively homogeneous cone.

1.1 Liquid Markets

In this chapter, we mainly follow Madan & Schoutens [29]. Our goal is to define the acceptability set \(A\) of acceptable cash flows containing \(X\). For simplicity, we consider only two times, namely \(t = 0\) and \(t = T\). We start with the classic framework, so let us consider a liquid and complete market with a unique risk-free measure \(Q^*\), equivalent to \(\mathbb{P}\). Under the traditional financial framework, the market value of \(X\) given a constant risk-free rate \(r\) equals \(V(X) = e^{-rT}E_{Q^*}[X]\). Hence, from the point of view of the market and if we denote a generic cash amount as \(w \geq 0\), at \(t = 0\) we consider the cash flow

\[
Z := X - e^{rT}w,
\]

which corresponds to the market buying the risk \(X\) for an amount of cash \(w\). Since \(V(X)\) is the arbitrage-free, fair price of the payoff \(X\), we have that \(V(Z) = 0\) for \(w = V(X)\). Hence, the market would accept this transaction for any \(w \leq V(X)\), since that means that the market buys \(X\) for a value less than or equal to its risk-neutral price, at zero cost. Indeed, we have that

\(^1\)Given a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), a set \(A \in \mathcal{F}\) is called an atom of \((\Omega, \mathcal{F}, \mathbb{P})\), if \(\mathbb{P}(A) > 0\) and for every measurable \(B \subset A\), either \(\mathbb{P}(B) = 0\) or \(\mathbb{P}(B) = \mathbb{P}(A)\).
\[ V(Z) = e^{-rT}E_{Q^*}[X - e^{rT}w] = e^{-rT}E_{Q^*}[X] - w = V(X) - w, \]

thus \( V(Z) \geq 0 \) for \( w \leq V(X) \). In the same manner, we represent by \( Z' := e^{rT}w - X \) the cash flow of the market selling the risk \( X \) at zero cost. Then we have

\[ V(Z') = e^{-rT}E_{Q^*}[e^{rT}w - X] = w - e^{-rT}E_{Q^*}[X] = w - V(X) \]

with \( V(Z') \geq 0 \) for \( w \geq V(X) \). Similarly to what was observed above, this means that the market will accept to sell the risk \( X \) for a value greater than or equal to its risk-neutral price. Consequently, the set of acceptable, zero-cost cash flows defined by the risk neutral measure \( Q^* \) is:

\[ A^* \equiv \{ Z \in L^\infty(\Omega, \mathcal{F}, P) : V(Z) = e^{-rT}E_{Q^*}[Z] \geq 0 \}. \]

Hence, under a one-price, liquid market framework, we have that \( A^* \) is the set of acceptable cash-flows, i.e. \( A \equiv A^* \). Furthermore, if by \( \tilde{A} \) we represent the set of non-negative random variables, then \( \tilde{A} \subset A^* \). This is because all cash flows \( Z \in \tilde{A} \) are in fact arbitrages and hence should always be acceptable. Indeed, if \( Z \geq 0 \) then \( V(Z) \geq 0 \) and \( X - e^{rT}w \geq 0 \). Then, one may at \( t = 0 \) borrow \( w \) under the the risk-free interest and buy \( X \) for its value \( V(X) \), since \( V(X) \geq w \). Then, at time \( t = T \), they receive \( X \) and pay back \( e^{rT}w \), hence a risk-free profit.

We also observe that \( \tilde{A} \) and \( A^* \) are both convex sets and cones, with the latter, in particular, being a half-space.

Now, let us consider an illiquid market. Then, \( A^* \) turns out to be too large to be the set of acceptable cash flows \( A \). An intuitive interpretation of this, is that the market is not actually trading in risk-neutral prices: the value of the traded asset is dependent on the “direction” (buy or sell) of the trade, i.e. whether the contract is sold or bought. More specifically, the prices that are quoted by the market are the bid price, which is the highest price someone is willing to pay in order to acquire a contract and the ask price, which is the lowest price someone is willing to receive in order to sell a contract. Hence, since the “fair price” quoted is not unique, but depends on the market direction, the price that is calculated under the one-price framework is not actually the arbitrage-free price, since it is well established that for the bid and ask prices quoted in the market the following relationship holds:

\[ \text{bid}(X) \leq \text{risk-neutral}(X) \leq \text{ask}(X). \] (1.1)

The difference between the bid and ask prices is called “bid-ask spread” and the less liquid the market is, the larger this spread is. Thus, Conic Finance offers a “two-price” framework, in which we consider two values of importance, i.e. the bid and ask price. In an illiquid market, the set of acceptable cash flows \( A \) still contains the set of non-negative risks, since arbitrages are always acceptable. However, it turns out to be more restricted than \( A^* \), hence \( \tilde{A} \subset A \subset A^* \). To be able to characterize \( A \) in an illiquid and incomplete market, we require the notion of coherent risk measures, which we describe in the following section.
1.2 Coherent Risk Measures

A risk measure \( \rho \) is a functional that assigns a real number to a risk \( X \). One may interpret the risk measure in the following manner: consider a situation in which there is a risk \( X \) with a potentially large payoff. Then, if one were to promise to pay \( X \) at \( t = T \), they would demand an amount of cash \( \rho(X) \) at \( t = 0 \) in order for the transaction to be acceptable. Then, the “riskier” \( X \) is (i.e. has larger potential payoff), the larger \( \rho(X) \) they would demand. Hence, the discounted value of \( \rho(X) \) may be considered as the value that must be paid at time \( t = 0 \), in order for someone to promise the risk (payoff) \( X \) at time \( t = T \), i.e. its price. In the following, we provide the definition of a coherent risk measure.

**Definition 1.2.** A functional \( \rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} \) is called a coherent risk measure if it satisfies the following properties:

1. **Normalization:**
   \[ \rho(0) = 0; \] (1.2)

2. **Translativity:** For all \( X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) and \( c \in \mathbb{R} \), we have
   \[ \rho(X + c) = \rho(X) + c; \] (1.3)

3. **Sub-additivity:** For all \( X, Y \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \), we have
   \[ \rho(X + Y) \leq \rho(X) + \rho(Y); \] (1.4)

4. **Positive homogeneity:** For any \( X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) and constant \( c > 0 \), we have
   \[ \rho(cX) = c\rho(X); \] (1.5)

5. **Monotonicity:** For all \( X, Y \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \), we have
   \[ \mathbb{P}(X \leq Y) = 1 \Rightarrow \rho(X) \leq \rho(Y). \] (1.6)

**Remark 1.3.** The aforementioned properties may be interpreted as such:

1. Holding no assets entails zero risk.

2. The cash \( \rho(X + c) \) that needs to be added to make \( X + c \) acceptable, is the amount of cash to make \( X \) acceptable (i.e. \( \rho(X) \)) plus the amount of cash \( c \), which implies that holding cash is risk-free.

3. The amount of cash that needs to be added to make a portfolio consisting of \( X + Y \) acceptable, is not greater than the amount of cash that needs to be added in order to make two separate portfolios consisting of \( X \) and \( Y \) acceptable. This property thus rewards portfolio diversification.
4. Taking the same position multiple times does not contribute to diversification. Hence, buying \(c\) units of the same risk \(X\), increases the risk by that factor and as such requires that much more cash to be acceptable.

5. If \(X\) is considered to be almost surely of no greater risk than \(Y\), then the cash that should be held to make \(X\) acceptable should be no greater than the cash to make \(Y\) acceptable.

In Remark 1.8, we argue about the choice of axioms for our definition of coherent risk measures. We now present a result about the correspondence between coherent risk measures and acceptability sets and an important representation result for coherent risk measures, both by Artzner et al. [3].

**Proposition 1.4.** Let \(\rho\) be a coherent risk measure. Then the set \(A_\rho := \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : \rho(-X) \leq 0\}\) induced by \(\rho\) is a closed acceptability set, as per the axioms of Definition 1.1.

**Proof.** By subadditivity, positive homogeneity and normalization, we have that \(\rho\) is a convex function on \(L^\infty(\Omega, \mathcal{F}, \mathbb{P})\) and hence a continuous function. As such, \(A_\rho\) is a closed, convex and positively homogeneous cone. Continuing, let \(X \geq 0 \Rightarrow -X \leq 0\). By monotonicity we have \(\rho(-X) \leq 0\), hence \(L^+ \subset A_\rho\). Finally, to prove that \(L^- \cap A_\rho = \emptyset\), assume that there exists \(X \in L^-\) such that \(\rho(-X) < 0\). Then, by monotonicity we have that \(0 = \rho(0) < 0\), a contradiction. Hence, if \(X \in L^-\), then \(\rho(-X) > 0 \Leftrightarrow X \notin A_\rho\).

**Proposition 1.5.** Let a functional \(\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}\) where \(\Omega\) is finite. Then, \(\rho\) is a coherent risk measure if and only if there exists a non-empty set of probability measures \(\mathcal{M}\) on \((\Omega, \mathcal{F}, \mathbb{P})\), such that:

\[
\rho(X) = \sup_{Q \in \mathcal{M}} E_Q[X]. \tag{1.7}
\]

**Proof.** “\(\Leftarrow\):”

Let \(\mathcal{M} \neq \emptyset\) a set of probability measures and a functional \(\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow [0, \infty]\) such that (1.7) holds. Then, to prove that \(\rho\) is a coherent risk measure, it suffices to show that \(\rho\) satisfies the properties of Definition 1.2. In the following, let \(X, Y \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})\).

1. \[\rho(0) = \sup_{Q \in \mathcal{M}} E_Q[0] = 0;\]

2. For \(c \in \mathbb{R}\):

\[
\rho(X + c) = \sup_{Q \in \mathcal{M}} E_Q[X + c] = \sup_{Q \in \mathcal{M}} (E_Q[X] + E_Q[c])
\]

\[
= \sup_{Q \in \mathcal{M}} (E_Q[X] + c) = \sup_{Q \in \mathcal{M}} E_Q[X] + c
\]

\[= \rho(X) + c;\]
3. 
\[ \rho(X + Y) = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X + Y] = \sup_{Q \in \mathcal{M}} (\mathbb{E}_Q[X] + \mathbb{E}_Q[Y]) \]
\[ \leq \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X] + \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[Y] = \rho(X) + \rho(Y); \]

4. For \( c > 0: \)
\[ \rho(cX) = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[cX] = c \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X] = c\rho(X); \]

5. Let \( \hat{Q} \in \mathcal{M} \) such that \( \forall Q \in \mathcal{M}: \)
\[ \mathbb{E}_Q[X] \leq \mathbb{E}_{\hat{Q}}[X] \Rightarrow \mathbb{E}_{\hat{Q}}[X] = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X]. \]

Then, we have
\[ X \leq Y \text{ a.s.} \Rightarrow \mathbb{E}_Q[X] \leq \mathbb{E}_Q[Y], \quad \forall Q \in \mathcal{M} \]
\[ \Rightarrow \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X] \leq \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[Y] \]
\[ \Rightarrow \rho(X) \leq \rho(Y). \]

“\( \Rightarrow \)”: Here we follow the proof of Huber [22], who also consider a finite \( \Omega \). Let a functional \( \rho : L^\infty \to \mathbb{R} \) satisfying the properties of Definition 1.2, where \( L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \). We want to show that there exists a non-empty set of probabilities \( \mathcal{M} \) such that \( \rho(X) = \sup_{Q \in \mathcal{M}} \int X dQ \). We see that it suffices to show that for every \( X_0 \in L^\infty \), there exists a probability measure \( Q \) such that for every \( X \in L^\infty \) it holds \( \int X dQ \leq \rho(X) \) and \( \int X_0 dQ = \rho(X_0) \). We note that (1.3) and (1.5) imply positive affine homogeneity for \( \rho \), i.e.
\[ \rho(aX + b) = a\rho(X) + b, \quad \forall a \geq 0, b \in \mathbb{R}. \] (1.8)

We start by taking an arbitrary \( X_0 \in L^\infty \). Because of (1.8), we may assume without loss of generality that \( \rho(X_0) = 1 \) and define
\[ U \coloneqq \{ X \in L^\infty : \rho(X) < 1 \}. \]

From (1.6) and (1.8) we have that \( U \) is open. Indeed, for every \( X \in U \), we set \( \epsilon = 1 - \rho(X) \) and see that for every \( Y \) such that \( Y < X + \epsilon \) it holds \( \rho(Y) < \rho(X + \epsilon) = 1 \Rightarrow Y \in U \). We furthermore see that because of (1.4), \( U \) is convex. Hence, because \( U \) is open and convex and \( X_0 \notin U \), there exists a linear functional \( \lambda \) separating \( U \) from \( X_0 \), i.e.
\[ \lambda(X) < \lambda(X_0), \quad \forall X \in U. \]

Since \( \lambda \) is linear and \( X \equiv 0 \in U \), we have that \( \lambda(X_0) > \lambda(0) = 0 \), hence we may normalize
\( \lambda(X_0) \) by setting \( \lambda(X_0) = 1 = \rho(X_0) \). We thus have

\[
\rho(X) < 1 \Rightarrow \lambda(X) < 1. \tag{1.9}
\]

Because of (1.6) and (1.8), we have

\[
X \leq 0 \Rightarrow \rho(X) \leq \rho(0) = 0 \Rightarrow X \in U.
\]

Hence, for \( X \geq 0 \) and \( c > 0 \) we have from (1.9)

\[
c\lambda(X) = -\lambda(-cX) > -1.
\]

Hence \( \lambda(X) \geq -1/c, \forall c > 0 \), so \( \lambda \) is a positive functional. We now claim that \( \lambda(1) = 1 \). To show this, we start by taking \( c < 1 \); then because \( c \in U \), we have from (1.9) that \( \lambda(c) < 1, \forall c < 1 \), which implies that \( \lambda(1) \leq 1 \). Now, let \( c > 1 \). We have \( \rho(2X_0 - c) = 2 - c < 1 \Rightarrow 2X_0 - c \in U \), so, from (1.9):

\[
\lambda(2X_0 - c) = 2 - c\lambda(1) < 1 \Rightarrow \lambda(1) > 1/c, \quad \forall c > 1.
\]

We thus have shown that \( \lambda(1) = 1 \). We now have from (1.8) and (1.9) that for every \( c \in \mathbb{R} \) it holds:

\[
\rho(X) < c \Rightarrow \lambda(X) < c.
\]

which implies \( \lambda(X) \leq \rho(X) \) for every \( X \in L^\infty \), which is what we wanted to show, with the probability measure \( Q(A) := \lambda(1_A) \in \mathcal{M} \). \( \square \)

**Remark 1.6.** We see that the probability measure \( \mathcal{M} \) of Proposition 1.5 is an arbitrary set. However, since we assumed that \( \Omega \) is finite, i.e. \( \Omega = \{ \omega_1, \ldots, \omega_n \} \) for some \( n \in \mathbb{N} \) we may assume, without loss of generality, that \( \mathcal{M} \) is closed and convex, since it can be identified with a subset of the simplex \( \{ (p_1, \ldots, p_n) : \sum p_i = 1, p_i \geq 0, \forall i \} \) [22].

If \( \Omega \) is an infinite set, we have the following theorem.

**Theorem 1.7.** Let \( \rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R} \) be a coherent risk measure and \( \Omega \) not finite. Then, the following are equivalent:

1. There is an \( L^1(\Omega, \mathcal{F}, P) \)-closed, convex set of probability measures \( \mathcal{M} \), all of them being absolutely continuous with respect to \( P \) and such that for every \( X \in L^\infty(\Omega, \mathcal{F}, P) \):

\[
\rho(X) = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[X]
\]

2. \( \rho \) satisfies the Fatou property, i.e.

\[
\rho(X) \leq \liminf \rho(X_n)
\]

for any sequence of random variables \( (X_n) \) uniformly bounded by 1 and \( X_n \overset{P}{\rightarrow} X \).
Proof. For a proof and an extended version of this theorem, see Theorem 3.2 in Delbaen [12]. □

We are now ready to define the set of acceptable cash flows $A$ for illiquid markets.

### 1.3 Illiquid Markets

From now on, we consider an illiquid and potentially incomplete market. Since the market is incomplete, it does not have a unique risk-neutral measure with which we can unequivocally define the set of acceptable cash flows $A$ as we did in Section 1.1. To deal with this situation, we consider a coherent risk measure $\rho$ satisfying the Fatou property. Then, by Theorem 1.7, we have that there exists an $L^1(\Omega, \mathcal{F}, \mathbb{P})$-closed, convex set of probability measures $\mathcal{M}$ with each probability measure in it absolutely continuous with respect to $\mathbb{P}$, such that $\rho(X) = \sup_{Q \in \mathcal{M}} E_Q[X]$, for every $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. We furthermore assume, without loss of generality, that $\mathcal{M}$ contains at least one risk-neutral measure $Q^*$, since $Q^* \sim \mathbb{P}$. The probability measures of $\mathcal{M}$ are interpreted as different “scenarios” on the state space $\Omega$ and are called “generalized scenarios” in Artzner et al. [3] and “test measures” by Carr et al. [8]. From Proposition 1.4, we have that $\rho$ induces the set $A_\rho = \{X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : \rho(-X) \leq 0\}$, which is an acceptability set. We thus define $A = A_\rho$ and hence, for a cash flow $Z \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ to be acceptable we have that:

$$Z \in A \iff \rho(-Z) \leq 0 \iff \sup_{Q \in \mathcal{M}} E_Q[-Z] \leq 0 \iff \inf_{Q \in \mathcal{M}} E_Q[Z] \geq 0$$

or, equivalently,

$$A = \left\{ Z \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : \inf_{Q \in \mathcal{M}} E_Q[Z] \geq 0 \right\}.$$

Since $A$ is an acceptability set and we have considered that a risk-neutral measure $Q^*$ is included in $\mathcal{M}$, we have that $\hat{A} \subset A \subset A^*$. Continuing, we have that if the market accepts to buy $X$ for the price $b$, it holds (from the point of view of the market) that:

$$X - e^{rT}b \in A \iff e^{-rT}E_Q[X - e^{rT}b] \geq 0, \quad \forall Q \in \mathcal{M}$$

or

$$e^{-rT}E_Q[X] \geq b, \quad \forall Q \in \mathcal{M},$$

while if the market accepts to sell $X$ for the price $a$, it holds that:

$$e^{rT}a - X \in A \iff e^{-rT}E_Q[e^{rT}a - X] \geq 0, \quad \forall Q \in \mathcal{M}$$

or

$$e^{-rT}E_Q[X] \leq a, \quad \forall Q \in \mathcal{M}.$$
which \( a \equiv b = w \). Now, given the definition of the bid and ask price, we are able to reach an explicit definition for the bid and ask prices, which are central in Conic Finance:

\[
\text{bid}(X) = e^{-rT} \inf_{Q \in \mathcal{M}} E_Q[X] \tag{1.10}
\]

and

\[
\text{ask}(X) = e^{-rT} \sup_{Q \in \mathcal{M}} E_Q[X]. \tag{1.11}
\]

Hence, given the representation result of Theorem 1.7, we have that for (1.11) it holds:

\[
\text{ask}(X) = e^{-rT} \sup_{Q \in \mathcal{M}} E_Q[X] = e^{-rT} \rho(X), \tag{1.12}
\]

i.e. we can express the ask price of \( X \) as a coherent risk measure, up to discounting. In addition, since buying a random cash flow \( X \) is equal to selling its negative \(-X\), we have that

\[
\text{ask}(X) = -\text{bid}(-X). \tag{1.13}
\]

Thus, we can also express the bid price with respect to a coherent risk measure.

**Remark 1.8.** It is common in the literature, including Artzner et al. [3], to define the coherent risk measure as such: a functional \( \rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R} \) that satisfies the normalization, subadditivity and positive homogeneity conditions as given in Definition 1.2, but with the translativity and monotonicity given as \( \rho(X + c) = \rho(X) - c \) and \( \mathbb{P}(X \leq Y) = 1 \Rightarrow \rho(X) \geq \rho(Y) \) respectively, for any \( X, Y \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \) and \( c \in \mathbb{R} \). Then, the risk measure is defined as \( \rho(X) = \sup_{Q \in \mathcal{M}} E_Q[-X] \) and comparing with (1.7), we see that \( \rho(X) = \rho'(-X) \). Since we follow Madan & Schoutens [29] for this chapter, we use their version of coherent risk measure, described in Definition 1.2. A reasoning for the different sign in this definition is that Artzner et al. [3] use the coherent risk measure in order to quantify risk in the traditional sense (e.g. market risk, currency risk etc), while we use it as a pricing functional. Hence, under the definition of [3], for a risk to be acceptable, there needs to hold \( \rho'(X) \leq 0 \), or \( \rho'(-X) \leq 0 \), which we interpret as the (undiscounted) premium that the market demands in order for it to sell \( X \) (thus promising a payoff \( X \)), which is non positive because of the direction of the cash flow. Also, since buying \( X \) is equivalent to selling \(-X\), we see that \( \rho(X) = \sup_{Q \in \mathcal{M}} E_Q[X] \geq 0 \) is equivalent to \( \rho'(X) = \sup_{Q \in \mathcal{M}} E_Q[-X] \leq 0 \). It should furthermore be noted that the results about coherent risk measures cited in this chapter were developed for the original definition of Artzner et al. [3] and we present them by taking into account the transformation \( X \mapsto -X \).

### 1.4 No Arbitrage in different markets

Given that we model the market using a convex cone \( \mathcal{A} \) derived from a convex set of probability measures \( \mathcal{M} \), the question arises whether it would be possible to have arbitrage by exploiting the potentially different convex cones of different markets. More specifically, if there are two markets with acceptable zero cost cash flow convex cones \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) derived from \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \),
would it be possible to have arbitrage by buying from one market and selling to the other? From (1.10) and (1.11) we get that, for $i \in \{1, 2\}$:

$$
\begin{align*}
\text{ask}_i(X) &= e^{-rT} \sup_{Q \in \mathcal{M}_i} E_Q[X] \\
\text{bid}_i(X) &= e^{-rT} \inf_{Q \in \mathcal{M}_i} E_Q[X]
\end{align*}
$$

If $\mathcal{M}_1 \cap \mathcal{M}_2 \neq \emptyset$ then there exists $Q^* \in \mathcal{M}_1 \cap \mathcal{M}_2$ and as such we have:

$$
\begin{align*}
\text{ask}_1(X) &\geq e^{-rT} E_{Q^*}[X] \geq \text{bid}_2(X) \\
\text{ask}_2(X) &\geq e^{-rT} E_{Q^*}[X] \geq \text{bid}_1(X)
\end{align*}
$$

It is thus not possible to create an arbitrage opportunity between the two markets, since the value of selling $X$ to any market cannot be higher than the value of buying $X$ from any market. This result can be clearly extended to any number of markets, as long as $\mathcal{M}_i \cap \mathcal{M}_j \neq \emptyset, \forall i \neq j$.

### 1.5 Distorted Expectations

Our goal now is to reach an analytic, easy to compute formula for the bid and ask prices. In what follows, we consider $(\Omega, \mathcal{F}, P)$ to be atomless. We start by giving the following definitions.

**Definition 1.9.** A functional $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is called law invariant if, for $X, Y \in L^\infty(\Omega, \mathcal{F}, P)$ with $F_X \equiv F_Y$, it follows that $\rho(X) = \rho(Y)$, where by $F_X (F_Y)$ we denote the cumulative distribution function of $X (Y)$.

**Definition 1.10.** A pair of random variables $X, Y : \Omega \to \mathbb{R}$ are called comonotone if for every couple $\omega_1, \omega_2 \in \Omega$, it holds $(X(\omega_2) - X(\omega_1)) \cdot (Y(\omega_2) - Y(\omega_1)) \geq 0$ a.s..

Comonotonicity thus implies that two risks are driven by one single factor. Another way to describe comonotonicity, is that there exists a random variable $U$ on the unit interval, such that $X = F_X^{-1}(U)$ and $Y = F_Y^{-1}(U)$. This, as well as Definition 1.10 are in fact (part of) proposition 4.5 from Dennberg [13], to which we refer for a rigorous and extensive treatment of comonotonicity.

**Definition 1.11.** A functional $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is called comonotonic if $\rho(X + Y) = \rho(X) + \rho(Y)$ for any comonotone pair $X, Y \in L^\infty(\Omega, \mathcal{F}, P)$.

Hence, if a risk measure is comonotonic it is additive, which leads to

$$
\begin{align*}
\text{bid}(X + Y) &= \text{bid}(X) + \text{bid}(Y) \\
\text{ask}(X + Y) &= \text{ask}(X) + \text{ask}(Y)
\end{align*}
$$
Comonotonicity implying additivity may be interpreted as the loss of diversification, due to the common dependency of the risks. Lastly, we give the definition of a distorted (probability) measure.

**Definition 1.12.** A non-decreasing concave (distribution) function $\Psi : [0, 1] \rightarrow [0, 1]$ with $\Psi(0) = 0$ and $\Psi(1) = 1$ is called a concave distortion function.

We may now present the following result by Kusuoka [27].

**Theorem 1.13.** Let $\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. Then, the following are equivalent.

1. $\rho$ is a law invariant and comonotonic coherent risk measure with the Fatou property.
2. There exists a non-decreasing concave function $\Psi$ with $\Psi(0) = 0$ and $\Psi(1) = 1$, such that for every $0 \leq X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$
   $$\rho(X) = \int_0^\infty \mu(X > x)dx,$$
   (1.14)
   where $\mu : \mathcal{F} \rightarrow [0, 1]$ with $\mu(A) = \Psi(\mathbb{P}(A)), A \in \mathcal{F}$.

**Proof.** For a proof and a slightly more general version of the theorem, see [27].

Our goal now, is to extend this result to all $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. To do this, we require the concept of Choquet integrals\textsuperscript{2}, introduced by Choquet [10]. We provide their definition from [17].

**Definition 1.14.** Let $c : \mathcal{F} \rightarrow [0, 1]$ be any set function that is normalized, i.e. $c(\emptyset) = 0$, $c(\Omega) = 1$ and monotone, i.e. $c(A) \leq c(B)$ if $A \subset B$. Then, for every $0 \leq X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, we define the Choquet integral of $X$ with respect to $c$ as:

$$(C) \int Xdc := \int_0^\infty c(X > x)dx.$$ 

Furthermore, for $X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ we have:

$$(C) \int Xdc := \int_{-\infty}^0 (c(X > x) - 1)dx + \int_0^{+\infty} c(X > x)dx,$$

with the integrals in the right hand side being usual Riemann integrals.

These integrals are non-additive, which is a desirable property, as our goal is to assign non-uniform probability weights to all outcomes. If $c$ is a probability measure, then the Choquet integral coincides with the usual Lebesgue integral. Moreover, we have the following result.

**Theorem 1.15.** Let $c$ be a normalized monotone set function. Then the following are equivalent.

1. $\rho(X) = (C) \int Xdc, X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is a a coherent risk measure
2. The set function $c$ is submodular, i.e. $c(A \cap B) + c(A \cup B) \leq c(A) + c(B), \forall A, B \in \mathcal{F}.$

\textsuperscript{2}For a more rigorous treatment of non-additive (Choquet) integrals, we refer to Denneberg [13].
Proof. For a proof and an extended version of this theorem, see [17].

The normalized and monotone set function \( \mu = \Psi \circ P : \mathcal{F} \to [0, 1] \) as defined in Theorem 1.13, is also submodular (see Eberlein et al. [15]), and we thus have:

\[
\rho(X) = (C) \int X d\mu = \int_{-\infty}^{0} (\mu(X > x) - 1)dx + \int_{0}^{+\infty} \mu(X > x)dx, \quad X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}),
\]

which yields:

\[
-\rho(-X) = - (C) \int (-X) d\mu
= - \int_{-\infty}^{0} (\mu(-X > x) - 1)dx - \int_{0}^{\infty} \mu(-X > x)dx
= \int_{-\infty}^{0} (1 - \mu(X < -x))dx - \int_{0}^{\infty} \mu(X < -x)dx
= \int_{0}^{\infty} (1 - \mu(X < x))dx - \int_{-\infty}^{0} \mu(X < x)dx
= - \int_{-\infty}^{0} \Psi(F_X(x))dx + \int_{0}^{\infty} (1 - \Psi(F_X(x)))dx
= \int_{-\infty}^{+\infty} x d\Psi(F_X(x)).
\]

With this result available, we now consider the cash flow \( Z = X - e^{rT}b \), in which the market buys a risk \( X \) for a price \( b \). Then, from previous results, for the cash flow to be acceptable there has to exist a non-empty convex closed set of probability measures \( \mathcal{M} \) containing a risk-neutral measure \( Q^* \) such that:

\[
Z \in \mathcal{A} \iff \inf_{Q \in \mathcal{M}} E_Q[Z] \geq 0 \iff - \inf_{Q \in \mathcal{M}} E_Q[Z] \leq 0
\]

\[
\iff \sup_{Q \in \mathcal{M}} E_Q[-Z] \leq 0 \iff \rho(-Z) \leq 0
\]

\[
\iff \int_{-\infty}^{+\infty} zd\Psi(F_Z(z)) \geq 0.
\]

Now, following [31], since \( F_Z(z) = F_X(e^{rT}b + z) \), we have:

\[
\int_{-\infty}^{+\infty} zd\Psi(F_X(e^{rT}b + z)) \geq 0
\]

and by assigning \( z = x - e^{rT}b \):

\[
\int_{-\infty}^{+\infty} (x - e^{rT}b)d\Psi(F_X(x)) \geq 0.
\]
Hence, by the definition of the bid price, we get:

\[
\text{bid}(X) = e^{-rT} \int_{-\infty}^{+\infty} x d\Psi(F_X(x)). \tag{1.15}
\]

Additionally, because of (1.13), the ask price of \( X \) is given by:

\[
\text{ask}(X) = -e^{-rT} \int_{-\infty}^{+\infty} x d\Psi(F_{-X}(x)). \tag{1.16}
\]

We thus have an analytic formula for both the bid and the ask price of any risk \( X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \).

**Remark 1.16.** We observe that the identity function \( I : [0, 1] \to [0, 1], I(u) = u \) satisfies the definition of a distortion function and that by taking \( \Psi = I \) we get

\[
\text{bid}(X) = e^{-rT} \int_{-\infty}^{+\infty} x dF_X(x) = \text{ask}(X),
\]

which is the risk-neutral price. We thus again see that the Conic Finance framework is not only consistent with the classic one-price framework, but that in fact it incorporates it as a special case.

In the bid equation (1.15), because of the concavity of \( \Psi \), a larger probability weight is assigned to the lower quantiles and a smaller probability weight is assigned to the higher quantiles. This re-weights the bid price downwards and assures that it will be lower than the mid price. Correspondingly, we see that for the ask price the lower quantiles are assigned a smaller probability weight and the higher quantiles are assigned a larger probability weight, which leads the ask price being larger than the mid price. This leads this approach to be more conservative, since we assign more weight to the quantiles associated with losses and less weight to quantiles associated with gains. Moreover, it guarantees that condition (1.1) is satisfied. Additionally, if \( \Psi(u) \) is differentiable and we denote the density function of \( X \) as \( f_X(x) = \frac{dF_X(x)}{dx} \), we have

\[
\text{bid}(X) = e^{-rT} \int_{-\infty}^{+\infty} x \Psi(F_X(x)) f_X(x) dx. \tag{1.17}
\]

Finally, in correspondence with the formula

\[
\mathbb{E}[X] = \int_{-\infty}^{+\infty} x dF_X(x) = -\int_{-\infty}^{0} F_X(x) dx + \int_{0}^{+\infty} (1 - F_X(x)) dx,
\]

and given a distortion function \( \Psi \), we may define a Choquet expectation operator \( \mathcal{D}^\Psi \) as:

\[
\mathcal{D}^\Psi[X] := \int_{-\infty}^{+\infty} x d\Psi(F_X(x)) = -\int_{-\infty}^{0} \Psi(F_X(x)) dx + \int_{0}^{+\infty} (1 - \Psi(F_X(x))) dx, \tag{1.18}
\]

hence \( \text{bid}(X) = e^{-rT} \mathcal{D}^\Psi[X] \). Moreover, by introducing the complementary distortion

\[
\hat{\Psi}(u) = 1 - \Psi(1 - u)
\]
which is convex and bounded above by the identity function, we can rewrite (1.15) and (1.16) in a more symmetric way as:

\[
\text{bid}(X) = e^{-rT} \left( -\int_0^\infty \Psi(1 - F_X(x)) \, dx + \int_0^\infty \hat{\Psi}(1 - F_X(x)) \, dx \right)
\]

and

\[
\text{ask}(X) = e^{-rT} \left( -\int_0^\infty \Psi(1 - F_X(x)) \, dx + \int_0^\infty \hat{\Psi}(1 - F_X(x)) \, dx \right),
\]

which shows that the bid and ask prices are a discounted expectation under a non-additive probability.

### 1.6 Examples of distortion functions

Here we cite some potential distortion functions, defined on the unit interval, following Madan & Schoutens [29], to which we refer for a more thorough presentation.

1. The MINVAR Distortion Function:

\[
\Psi_{\lambda}^{\text{MINVAR}}(u) = 1 - (1 - u)^{1+\lambda}, \quad \lambda \geq 0
\]

2. The MAXVAR Distortion Function:

\[
\Psi_{\lambda}^{\text{MAXVAR}}(u) = u^{1+\lambda}, \quad \lambda \geq 0
\]

3. The MAXMINVAR Distortion Function:

\[
\Psi_{\lambda}^{\text{MAXMINVAR}}(u) = \left( 1 - (1 - u)^{1+\lambda} \right)^{\frac{1}{1+\lambda}} = \Psi_{\lambda}^{\text{MAXVAR}} \left( \Psi_{\lambda}^{\text{MINVAR}}(u) \right), \quad \lambda \geq 0
\]

4. The MINMAXVAR Distortion Function:

\[
\Psi_{\lambda}^{\text{MINMAXVAR}}(u) = 1 - \left( 1 - u \right)^{1+\lambda}, \quad \lambda \geq 0
\]

5. The Wang Transform:

\[
\Psi_{\lambda}^{\text{WANG}}(u) = N(N^{-1}(u) + \lambda), \quad \lambda \geq 0, \ u \in (0, 1)
\]

where \( N \) is the standard normal cumulative distribution function and \( N^{-1} \) its reverse function.

We now present some plots of the above distortion functions for different values of \( \lambda \).
We thus see the different behaviour of each distortion function. Also of interest are the extreme cases when $\lambda = 0$ and $\lambda \to \infty$.

Figure 1.2: Distortion functions for $\lambda = 0$ and $\lambda = \infty$.

1.7 Conic Finance Pricing

In this section we present two widely used pricing methods, namely the Binomial Tree and Monte Carlo methods, applied under the Conic Finance framework.

1.7.1 Conic Binomial Trees

Here we assume the simple and well-know setting of a one-step binomial tree. Namely, we consider two discreet time points, $t_0 = 0$ in which our non-divident-paying asset has a value of $S_0$ and $t = T$, in which our asset has two possible states, an ‘up’ state $uS_0$ and a ‘down’ state
\( dS_0, \text{ for } 0 \leq d < e^{\rho T} < u, \) where \( r \) is the risk-free rate. If \( p \) is the risk-neutral probability that the asset moves to the up state, then the value of a derivative under the risk-neutral framework at \( t_0 \) is

\[
f = e^{-rT}(pf_u + (1-p)f_d)
\]

where \( f_u \) denotes its value at the up state, \( f_d \) its value at the down state and we assume \( f_u, f_d \geq 0 \). We thus get that the risk-neutral probability of the asset to move to the up state is

\[
p = \frac{e^{\rho T} - d}{u - d}.
\]

Let us now consider the Conic Finance approach. Under the same setting, we further assume a distortion function \( \Psi \) and that \( f_d < f_u \), which would be the case of e.g. a European call option. Since we must have bid\((X) \leq \) risk-neutral\((X) \leq \) ask\((X) \), we need to assign probability weights appropriately. Hence, for the bid price case, we want the down state to have a higher probability than that of the risk-neutral approach. As such, we take \( \Psi(1-p) \) to be the probability that the asset moves to the ‘down’ state, which is precisely the distorted risk-neutral ‘down’ state probability. Hence, in this case, the bid price is

\[
f_{\text{bid}}^{\text{call}} = e^{-rT}[(1 - \Psi(1-p))f_u + \Psi(1-p)f_d]
\]

For the ask price, we want to re-weight upwards the probability to move unto the ‘up’ state, so this is the risk-neutral probability that we distort, which gives us

\[
f_{\text{ask}}^{\text{call}} = e^{-rT}[\Psi(p)f_u + (1 - \Psi(p))f_d]
\]

Now, let us consider a derivative for which \( f_u < f_d \), e.g. a European put option. In this case, when calculating the bid price, we want to re-weight upwards the ‘up’ state. We thus have

\[
f_{\text{bid}}^{\text{put}} = e^{-rT}[\Psi(p)f_u + (1 - \Psi(p))f_d]
\]

Similarly, for the ask we distort the ‘down’ state probability, which yields

\[
f_{\text{ask}}^{\text{put}} = e^{-rT}[(1 - \Psi(1-p))f_u + \Psi(1-p)f_d]
\]

We thus see that, in the Conic Finance framework, the ‘up’ and ‘down’ state probabilities (and as such the derivative prices) are market direction dependent.

The following graph is a geometric representation of two-state conic binomial trees and risk-neutral probability \( p \). We consider the asset to be in the up-state and the derivative to have a positive up-state payoff and zero down-state payoff (e.g. a European call option).
Here, we consider axis $y$ to consist of the up-state payoffs and axis $x$ to consist of the down-state payoffs of a derivative $X$. The red line is the risk-neutral line, defined by the risk-neutral probability $p$, containing all derivatives with zero expectation under the risk-neutral measure. As such, the half-space “above” it is the set $\mathcal{A}^*$ of risk-neutral acceptable cash flows, as defined in section 1.1. The blue (green) line represents the payoffs with zero bid (ask) values under the distorted probability measures defined before. The intersection of the half-spaces defined by the three lines is a cone and it is the space of accepted cash flows $\mathcal{A}$. To find the bid price, one needs to project the point of the payoff along the direction of the bid line unto the line $y = x$ and then its projection unto the $x$ (or $y$) axis is the bid price, up to discounting. The same procedure holds for the mid and ask prices.

1.7.2 Conic Monte Carlo

Using Monte Carlo simulations is a well-established method for pricing both vanilla and exotic derivatives. Its main idea is to simulate $N$ paths of the underlying, which give $N$ different payoff values, under the risk-neutral measure $Q$. Then, for $N \to \infty$ we get that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \text{payoff}_i = \mathbb{E}_Q[\text{payoff}] = e^{rT} \cdot \text{premium}$$

i.e., for a sufficiently large $N$,

$$\text{premium} \simeq e^{-rT} \frac{1}{N} \sum_{i=1}^{N} \text{payoff}_i$$

Thus, we can consider that each simulation is assigned a weight of $1/N$. Under the Conic Finance framework, the approach is the following: after running $N$ simulations, we sort the simulated
payoffs in an ascending order, i.e.,

$$\text{payoff}(i) \leq \text{payoff}(i+1), \forall i \in \{1, \ldots, N-1\}.$$ 

It obviously holds that

$$e^{-rT} \frac{1}{N} \sum_{i=1}^{N} \text{payoff}_i = e^{-rT} \frac{1}{N} \sum_{i=1}^{N} \text{payoff}(i).$$

Working under the Conic Finance framework, we assign higher weights to lower payoffs and lower weights to higher payoffs to calculate the bid price and vice versa to calculate the ask price. As such, instead of assigning a uniform weight to all simulated payoffs, we use a distortion function \(\Psi\) to assign weights in the aforementioned manner to the sorted payoffs. Indeed, we use weights \(p^*_i\) for the bid price and \(\tilde{p}_i\) for the ask price, given by:

\[
\begin{align*}
p^*_i &= \Psi\left(\frac{i}{N}\right) - \Psi\left(\frac{i-1}{N}\right), \quad i \in \{1, \ldots, N\} \\
\tilde{p}_i &= \Psi\left(\frac{N-i+1}{N}\right) - \Psi\left(\frac{N-i}{N}\right), \quad i \in \{1, \ldots, N\},
\end{align*}
\]

which we can compare to the “uniform” Monte Carlo weights as such: let a random variable \(X \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})\), a distortion function \(\Psi\) and an i.i.d.\(^3\) empirical sample \(x_1, \ldots, x_N\) of \(X\). Then we can estimate its expected value using the empirical distribution function, i.e.

\[
E[X] \approx \int_{-\infty}^{+\infty} xd\hat{F}_X(x) = \sum_{i=1}^{N} x_i \Delta \left(\frac{i}{N}\right) = \frac{1}{N} \sum_{i=1}^{N} x_i.
\]

For its distorted expectation we thus have

\[
D^\Psi[X] \approx \int_{-\infty}^{+\infty} xd\Psi(\hat{F}_X(x)) = \sum_{i=1}^{N} x_{(i)} \Delta \Psi\left(\frac{i}{N}\right) = \sum_{i=1}^{N} x_{(i)} \left(\Psi\left(\frac{i}{N}\right) - \Psi\left(\frac{i-1}{N}\right)\right),
\]

where we used \(x_{(i)}\) instead of \(x_i\) because of the different weights assigned through \(\Psi\). Moreover, we see that (1.19) are indeed probability weights, as \(p^*_i, \tilde{p}_i \geq 0 \forall i \in \{1, \ldots, N\}\) and

\[
\sum_{i=1}^{N} p^*_i = \sum_{i=1}^{N} \tilde{p}_i = 1.
\]

We also see that if \(\Psi\) is the identity function, then \(p^*_i = \tilde{p}_i \equiv 1/N\), which is the probability weight assigned in the one-price framework. Thus, for \(N\) large enough, we have the following

\(^3\)Independent and Identically Distributed.
equation for the bid price:

$$\text{bid} \simeq e^{-rT} \sum_{i=1}^{N} p_i \text{payoff}_{(i)}$$

$$= e^{-rT} \sum_{i=1}^{N} \left( \Psi \left( \frac{i}{N} \right) - \Psi \left( \frac{i-1}{N} \right) \right) \text{payoff}_{(i)} \quad (1.20)$$

while the corresponding equation for the ask price is given by:

$$\text{ask} \simeq e^{-rT} \sum_{i=1}^{N} \tilde{p}_i \text{payoff}_{(i)}$$

$$= e^{-rT} \sum_{i=1}^{N} \left( \Psi \left( \frac{N-i+1}{N} \right) - \Psi \left( \frac{N-i}{N} \right) \right) \text{payoff}_{(i)} \quad (1.21)$$

We also observe that the weights $p_i^*$ and $\tilde{p}_i$ only depend on the (pre-selected) distortion function $\Psi$ and on the number of simulations $N$, but not on the simulated paths themselves. As such, the probability weights can be calculated either before or after the simulation procedure takes place.
2 Swaptions and the SABR model

In this chapter we describe the model that has been utilized in this thesis, i.e. the (displaced) SABR model. We also give a description of the swaption, which is the financial product that we price. Since swaptions are options having swaps as the underlying, we begin by giving a brief introduction of swaps.

2.1 Swaps

A (vanilla) interest rate swap is an agreement between two parties: one party pays interest at a fixed rate on a notional for a predetermined time period, while receiving interest at a floating rate on the same notional, for the same time period. The time period between the start of the swap and the last payment is called maturity. The floating rate is usually based on the London Interbank Offered Rate\(^1\) (LIBOR), which is the interest rate that London based banks with credit rating AA charge in order to lend to each other (Hull [24]). An interest rate swap can be further characterized depending on which party pays the fixed interest and which the floating interest. Indeed, from the point of view of the party that pays the fixed interest rate, the swap is called a payer swap, while for the party paying the floating interest rate, it is called a receiver swap. In the following, we describe how the cash flows associated with swaps were traditionally calculated, using a Single Curve framework and how this has changed into a Multi-Curve framework after the 2008 financial crisis.

2.1.1 Single Curve Framework

In the following, we follow Filipovic [16]. Consider that there are \(n\) payments taking place at dates \(T_0 < T_1 < \ldots < T_n\), with \(T_i - T_{i-1} = \delta\). We denote \(N\) to be the notional and \(K\) the fixed interest rate. Then the value of the fixed interest payments (“fixed leg”) can be calculated as the sum of the discounted cash flows, i.e.:

\[
p_{\text{fix}}(t) = NK\delta \sum_{i=1}^{n} P(t, T_i) + NP(t, T_n), \quad t \leq T_1
\]

where by \(P(t, T)\) we denote the zero coupon bond at time \(t\), with maturity \(T\), which represents the discounting. The \(NP(t, T_n)\) part of the right hand side represents the return of the (discounted) notional during the last payment at \(T_n\). The value of the floating interest payments (“floating leg”) can be calculated as the sum of the discounted cash flows:

\[
p_{\text{float}}(t) = N\delta \sum_{i=1}^{n} P(t, T_i)F(T_{i-1}, T_i) + NP(t, T_n), \quad t \leq T_0,
\]

\(^1\)Depending on the currency, there are other IBORs used, such as EURIBOR for the euro.
where the fixed rate $K$ has been replaced with $F(T_{i-1}, T_i)$, which is an interest rate for the time interval $[T_{i-1}, T_i]$, fixed at $T_{i-1}$. Note that here we have assumed for simplicity that the number of cash flows for the fixed and the floating leg is the same, hence $\delta_{\text{fix}} = \delta_{\text{float}} = \delta$, but this is not necessarily the case. Thus, the value of a payer swap at time $t \leq T_0$ is $\Pi_p(t) = p_{\text{float}}(t) - p_{\text{fix}}(t)$, which can be calculated (Filipovic [16]) to be:

$$\Pi_p(t) = N \left( P(t, T_0) - P(t, T_n) - K\delta \sum_{i=0}^{n} P(t, T_i) \right).$$

For the receiver swap, we need only change the sign of the cash flows, thus getting

$$\Pi_r(t) = -\Pi_p(t).$$

Hence, we are now able to calculate the “fair” value of the fixed rate, i.e. the value of $K$ such that $\Pi_p(t) = \Pi_r(t) = 0$. For $t = 0$, we call this the swap rate and denote it as $R_{\text{swap}}(0)$, which is calculated to be

$$R_{\text{swap}}(0) = \frac{P(0, T_0) - P(0, T_n)}{\delta \sum_{i=1}^{n} P(0, T_i)},$$

hence at time $t \leq T_0$, we have

$$R_{\text{swap}}(t) = \frac{P(t, T_0) - P(t, T_n)}{\delta \sum_{i=1}^{n} P(t, T_i)}.$$

As can be seen, the swap rate is calculated using only one interest rate curve. Additionally, we may compute the forward swap rate for any time before the first payment $0 < t \leq T_0$:

$$R_{\text{swap}}^{\text{ fwd}} \bigg|_{t=0} = \mathbb{E} \left[ R_{\text{swap}}(t) \bigg| \mathcal{F}_0 \right] = \mathbb{E} \left[ \frac{P(t, T_0) - P(t, T_n)}{\delta \sum_{i=1}^{n} P(t, T_i)} \bigg| \mathcal{F}_0 \right].$$

### 2.1.2 Multi-Curve Framework

Traditionally, LIBOR was considered to entail insignificant risk and as such was used as a proxy for the risk-free interest rate (Hull and White [23]). Hence, LIBOR was used for both defining the floating interest rate and for discounting, leading to the Single Curve framework discussed before. However, the events of the financial crisis of 2008, showed that this was not the case and led to a different approach. Namely, the Overnight Index Swap $^{2}$ (OIS) rate has become increasingly popular as a risk-free rate proxy. The reason for this is that, because OIS allows the interbank loans to be rolled into the next day, it entails less risk than LIBOR. As such, it is becoming more and more common to use OIS as the discount rate (Henrard [21]). For more details on OIS and the events leading it gradually replacing LIBOR for the purpose of discounting, we refer to [21] and [23].

Under this framework, we denote by $D(t, T)$ the discounting derived from the OIS curve and by $F(T_{i-1}, T_i)$ the interest rate of the time interval from $T_{i-1}$ to $T_i$, derived from the forward curve $^{2}$“An OIS is a swap where a fixed rate for a period is exchanged for the geometric average of the overnight rates during the period” (Hull [24]).
(LIBOR or otherwise). We now examine the more general case where, for the fixed leg we have $m$ payments and consider $T^\text{fix}_j - T^\text{fix}_{j-1} \equiv \delta^\text{fix}$ for every $j \in \{1, \ldots, m\}$. Correspondingly, for the floating leg we consider $n$ payments and $T^\text{float}_i - T^\text{float}_{i-1} \equiv \delta^\text{float}$ for every $i \in \{1, \ldots, n\}$. Then, we calculate the swap rate at $t = 0$ to be

$$R_{\text{swap}}(0) = \frac{\delta^\text{float} \sum_{i=1}^{n} F(T^\text{float}_{i-1}, T^\text{float}_i)D(0, T^\text{float}_i) \delta^\text{fix} \sum_{j=1}^{m} D(0, T^\text{fix}_j)}{\sum_{j=1}^{m} D(0, T^\text{fix}_j)} \quad (2.1)$$

and at $t \leq T_1$ we have:

$$R_{\text{swap}}(0) = \frac{\delta^\text{float} \sum_{i=1}^{n} F(T^\text{float}_{i-1}, T^\text{float}_i)D(0, T^\text{float}_i) \delta^\text{fix} \sum_{j=1}^{m} D(0, T^\text{fix}_j)}{\sum_{j=1}^{m} D(0, T^\text{fix}_j)}. \quad (2.2)$$

Hence, the forward swap rate is calculated to be

$$R_{\text{swap}}^{\text{fwd}}(t) \bigg|_{t=0} = \mathbb{E} \left[ R_{\text{swap}}(t) \bigg| \mathcal{F}_0 \right] = \frac{\delta^\text{float}}{\delta^\text{fix}} \mathbb{E} \left[ \frac{\sum_{i=1}^{n} F(T^\text{float}_{i-1}, T^\text{float}_i)D(t, T^\text{float}_i)}{\sum_{j=1}^{m} D(t, T^\text{fix}_j)} \bigg| \mathcal{F}_0 \right]. \quad (2.3)$$

### 2.2 Swaptions

Swaptions are derivatives whose underlying is an interest rate swap. They give the holder the right (but not the obligation) to enter a swap at a predetermined swap rate, at a predetermined future date (Hull [24]). They are defined by three dates: the date in which the swaption is issued, called the start date, the date in which the holder of the swaption decides whether to exercise the option or not, called the expiry date and the date describing how long the swap payments last after the option is exercised, called the maturity date. There are two main types of swaptions: the “swap-settled” swaptions, in which exercising the swaption leads to entering the underlying swap and the “cash-settled” swaptions, in which exercising the swaption leads to a cash settlement upon the expiry date (Pietersz and Sengers [35]). Swap-settled swaptions are popular when USD is the currency, while cash-settled swaptions are popular when EUR or GBP is the currency [35]. Following Mercurio [33] and applying the Multi Curve framework discussed in the previous section with $t_0 < T < T_1 < \ldots < T_n = \tau$, the payoff of the swap-settled swaptions at $t = T$ is:

$$\text{payoff}_{\text{swap}} = N \cdot A_T^{\text{swap}} \cdot (\omega \cdot (S_T - K))^+$$

where

$$A_T^{\text{swap}} = \sum_{i=1}^{n} \delta_i D(T, T_i) \quad (2.4)$$

and we define

$$A_t^{\text{swap}} := D(t, T) A_T^{\text{swap}}.$$
The payoff of the cash-settled swaptions at $t = T$ is:

$$\text{payoff}_{\text{cash}} = N \cdot A^\text{cash}_T \cdot (\omega \cdot (S_T - K))^+$$  \hspace{1cm} (2.5)

where

$$A^\text{cash}_T = \sum_{i=1}^{n} \frac{\frac{1}{m} R(T)\mid_{t_i}}{(1 + \frac{1}{m} R(T)\mid_{t_i})^i}.$$  \hspace{1cm} (2.6)

Again, we define

$$A^\text{cash}_t := D(t,T)A^\text{cash}_T$$

In the previous formulas, we have used the following notation:

- $t_0$: start date
- $T$: expiry date of the swaption
- $\tau$: the maturity date of the swaption
- $D(t_1,t_2)$: discounting factor from $t_1$ to $t_2$, based on the discounting curve
- $n$: number of floating payments of the underlying swap
- $m$: number of fixed leg payments per annum
- $\delta_i$: year fraction corresponding to the $i$-th fixed leg payment
- $R(T)\mid_{t_0}$: the forward swap rate for expiry $T$ and tenor $\tau$, at $t_0$, as described in (2.3)
- $S_T$: the value of the underlying swap at time $t = T$
- $K$: the strike of the option
- $N$: the notional of the underlying swap
- $\omega$: $\omega = 1$ if it is a payer swap, $\omega = -1$ if it is a receiver swap

We call (2.4) the swap annuity, or present value of a basis point (PVBP). The equivalent martingale measure $Q^\text{swap}$ using (2.4) as the numeraire, is called the (swap-settled) annuity measure, under which payoff$_{\text{swap}}$ is a martingale (Henrard [21]). Hence, using the same notation as before, the premium of a swap-settled swaption is given by:

$$V_{\text{swap}}(t) = NA^\text{swap}_t E_{Q^\text{swap}} [(\omega(S_T - K))^+ \mid \mathcal{F}_t]$$  \hspace{1cm} (2.7)

or, through (2.4):

$$V_{\text{swap}}(t) = ND(t,T) \sum_{i=1}^{n} \delta_i D(T, T_i) E_{Q^\text{swap}} [(\omega(S_T - K))^+ \mid \mathcal{F}_t].$$
Correspondingly, we call (2.6) the cash annuity and the premium of a cash-settled swaption is given by:

$$V_{\text{cash}}(t) = D(t, T) \mathbb{E}_{Q^T} \left[ \text{payoff}_{\text{cash}} | \mathcal{F}_t \right]$$

where $Q^T$ is the $T$-forward measure. In the swap-settled swaption case, the swap annuity is essentially a portfolio of zero-coupon bonds and as such may conveniently be used as a numeraire. However, the cash-settled annuity is merely an auxiliary quantity and thus has no representation as a financial product, therefore it cannot be used as a numeraire. Hence, to calculate the premium, we make the assumption that the cash-settled swaption is a martingale under the annuity measure, hence:

$$V_{\text{cash}}(t) = \frac{V_{\text{swap}}(t)}{A_t^{\text{swap}}} \mathbb{E}_{Q^{\text{swap}}} \left[ \frac{A_t^{\text{cash}}}{A_t^{\text{swap}}} \left( \omega(S_T - K) \right)^+ | \mathcal{F}_t \right]$$

or

$$V_{\text{cash}}(t) = NA_t^{\text{swap}} \mathbb{E}_{Q^{\text{swap}}} \left[ \frac{A_t^{\text{cash}}}{A_T} \left( \omega(S_T - K) \right)^+ | \mathcal{F}_t \right]. \quad (2.8)$$

While the Single-Curve framework was the standing market practice, $A_t^{\text{cash}}$ was considered as an approximation of $A_t^{\text{swap}}$, hence $A_t^{\text{swap}} \approx A_t^{\text{cash}}$, making (2.8):

$$V_{\text{cash}}(t) = NA_t^{\text{cash}} \mathbb{E}_{Q^{\text{swap}}} \left[ \left( \omega(S_T - K) \right)^+ | \mathcal{F}_t \right] \quad (2.9)$$

or

$$V_{\text{cash}}(t) = ND(t, T) \sum_{i=1}^{n} \frac{1}{1 + \frac{1}{m} R(T)|_{t}^i} \mathbb{E}_{Q^{\text{swap}}} \left[ \left( \omega(S_T - K) \right)^+ | \mathcal{F}_t \right]. \quad (2.10)$$

This is a mathematically wrong result, since the cash-settled swaption premium is not a martingale under the swap-settled measure. However, under the Single Curve framework, the errors of this approach were considered small enough to justify its use. Under the Multi-Curve framework, these errors have become non-insignificant and have led to quoting different implied volatilities for swap/cash settled swaptions, quoting different premiums for ATM payer and receiver cash settled swaptions, as well as quoting non-negligible prices for cash-settled zero-width collars\(^3\) (Pietersz and Sengers \[35\]). However, (2.9) is still the “the market-standard formula for cash-settled swaptions” \[35\] and as such we consider (2.10) as our starting point for pricing cash-settled swaptions.

---

\(^3\)A zero-width collar is buying a payer swaption and selling a receiver swaption with the same strikes. Its value for cash-settled swaptions not being zero, implies that put-call parity does not hold.
2.3 SABR

In this section, we describe the SABR model, which was introduced by Hagan et al. [19] and unless otherwise stated, we follow their original work.

2.3.1 Original SABR Model

The SABR model is a stochastic volatility model given by the equations:

$$
\begin{align*}
\frac{dF_t}{F_t} &= \sigma_t F_t^{\beta} dW_t^{(1)}, \quad F_0 = f \\
\frac{d\sigma_t}{\sigma_t} &= \nu \sigma_t dW_t^{(2)}, \quad \sigma_0 = \alpha
\end{align*}
$$

(2.11)

where $F_t$ represents the forward value of the underlying, $\sigma_t$ the volatility, $\nu$ is the volatility of the volatility (or “volvol”) and for the parameters it holds that $0 \leq \beta \leq 1$, $\nu > 0$. Furthermore, we assume that the two Brownian motions are correlated:

$$dW_t^{(1)} dW_t^{(2)} = \rho dt, \quad -1 \leq \rho \leq 1. \quad (2.12)$$

We also impose the boundary condition that $F = 0$ is absorbing. In our swaption pricing case, we consider $f = R_{\text{fwd}}^{\text{swap}}(T)|_{t=0}$ as defined in (2.3), with $T$ being the expiry date of the swaption. The name of the model itself is derived from its parameters: SABR is an acronym for Stochastic Alpha Beta Rho. It is worth mentioning that the second equation of (2.11) can be solved analytically. Indeed, by applying Itô’s lemma for $f(x) = \log x$, we get

$$\sigma_t = \sigma_0 \exp \left( -\frac{1}{2} \nu^2 t + \nu W_t^{(2)} \right). \quad (2.13)$$

Moreover, for the special cases in which $\beta = 0$ or $\beta = 1$, then both equations of (2.11) have an explicit solution, with the underlying having a normal distribution in the $\beta = 0$ case and a log-normal distribution in the $\beta = 1$ case.

SABR was developed trying to solve the weaknesses of its contemporary standard market models. Namely, on one hand, the celebrated Black’s [5] model which, even though dominant in the market, made the rather strong assumption that the volatility remains constant. Since we have ample historical data of markets being (relatively) stable in some periods and outward chaotic in other periods, considering the volatility to be constant is unrealistic.

To deal with this problem, “local volatility” models were developed, in which the volatility was considered to be dynamic. In the work of e.g. Dupire [14], the proposed model was:

$$dF_t = \sigma_{\text{loc}}(t, F_t) F_t dW_t, \quad F_0 = f$$

under the forward measure, with $\sigma_{\text{loc}}(t, F_t) F_t$ being Markovian. Then, the local volatility was obtained by calibrating the model to market values of liquid derivatives. The problem with this approach is that it captures the wrong dynamics of the implied volatility curve [19], i.e. for an increase of the underlying price it predicts that the implied volatility curve shifts to lower prices and vice versa, which is contradictory to the real market behavior.
Thus, building on the idea that volatility should not be considered constant, (2.11) and (2.12) were used to define the SABR model, which captures the correct dynamics. Another advantage of SABR is that by using singular perturbation techniques, an analytic formula for the implied volatility can be calculated. Indeed, for not ATM strikes, i.e. when $K \neq f$ we have the formula:

$$
\sigma^{IMP}(K, f) = \frac{\alpha}{(fK)^{(1-\beta)/2}} \left[ 1 + \frac{(1-\beta)^2}{24} \log^2 \left( \frac{f}{K} \right) + \frac{(1-\beta)^4}{1920} \log^4 \left( \frac{f}{K} \right) + \ldots \right] \cdot \left( \frac{z}{\chi(z)} \right).
$$

(2.14)

with

$$
z = \frac{\nu}{\sigma} \left( fK \right)^{(1-\beta)/2} \log \left( \frac{f}{K} \right)
$$

and

$$
\chi(z) = \log \left( \frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho} \right).
$$

When $K = f$, (2.14) reduces to:

$$
\sigma^{IMP}_{ATM} = \sigma^{IMP}(f, f) = \frac{\alpha}{f^{(1-\beta)}} \left[ 1 + \frac{(1-\beta)^2}{24} \frac{\alpha^2}{f^{2-2\beta}} + \frac{1}{4} \frac{\alpha \beta \rho \nu}{f^{1-\beta}} + \frac{2 - 3 \rho^2}{24} \nu^2 \right] T + \ldots
$$

(2.15)

The extra terms in (2.14) and (2.15) are omitted because they are much smaller than the first terms. For more details on the derivation of (2.14) and (2.15), we refer to the original paper by Hagan et al. [19]. The implied volatility calculated by these formulas can then be used as input to Black’s formula, to obtain the premium’s value.

The parameters $\beta$ and $\rho$ affect the $\sigma^{IMP}(K)$ curve in a similar way, which is why it is standard practice to fix a constant value for $\beta$ a priori and then, for fixed expiry and maturity dates, calibrate the rest of the parameters to fit a curve $\sigma^{IMP}_{MKT}(K)$ of market observed implied volatilities. A way to do this is by least-square minimizing, that is, as seen e.g. in Rouah, F. D. [36]:

$$
(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \arg \min_{\alpha, \rho, \nu} \sum_{i: K_i} \left( \sigma^{IMP}_{MKT}(K_i) - \sigma^{IMP}(K_i, f; \alpha, \rho, \nu) \right)^2,
$$

(2.16)

with $i$ ranging through the available strikes $K_i$ of the fixed expiry and maturity. Then we plug the derived $\hat{\alpha}, \hat{\rho}, \hat{\nu}$ values into (2.14) and (2.15) to calculate the implied volatilities for all the strike values of the fixed expiry and maturity date. Another common practice is to first calculate $\alpha$ through the ATM equation (2.15), using the value for the ATM implied volatility and calibrating with respect to $\rho$ and $\nu$. Then, given the known $\alpha$ and $\beta$, $\rho$ and $\nu$ are calibrated...
as before. Again following [36], we may write this procedure as:

$$(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \text{arg min}_{\alpha, \rho, \nu} \sum_{i\in K_i} \left( \sigma_{MKT}^\text{IMP}(K_i) - \sigma^\text{IMP}(K_i, f; \alpha(\rho, \nu), \rho, \nu) \right)^2$$

(2.17)

In the $\sigma^\text{IMP}(K)$ curve, the role of the parameters is as such: $\alpha$ controls the overall height of the curve, $\nu$ controls how much “smile” the curve has and $\rho$ controls the curve’s skew.

### 2.3.2 Negative Interest Rates SABR adaptations

Because of the $F^\beta_t$ term in (2.11) and the fact that $0 \leq \beta \leq 1$, $F_t$ is not allowed to have negative values, except when $\beta = 0$ or $\beta = 1$. In the swaption pricing case, this property had the intuitive explanation that interest rates could never be negative, hence this property of SABR was considered an advantage. However, in the current post-crisis, low interest rate environment, negative interest rates are a reality and so SABR needed to be extended in a way that allows negative interest rates as well, that is to allow $F_t < 0$. Two such extensions are the “Free Boundary” SABR and the “Displaced” (or “Shifted”) SABR.

**Free Boundary SABR**

In the Free Boundary SABR, the absorbing boundary condition is dropped and (2.11) is modified to:

$$dF_t = \sigma_t |F_t|^\beta dW_t^{(1)}$$

$$d\sigma_t = \nu \sigma_t dW_t^{(2)}$$

with (2.12) remaining the same. Hence, since $|F_t| \neq 0$, $\forall t \geq 0$, $F_t$ is allowed to have negative values. For more information on the Free Boundary SABR, see Antonov et al. [2].

**Displaced SABR**

On the other hand, the Displaced (or Shifted) SABR, whose use we employ in this thesis, deals with the possible negative values of $F_t$ by introducing a deterministic (potentially non-constant) displacement factor $d > 0$ such that $F'_t := F_t + d$. The displacement factor needs to be large enough to “absorb” all foreseeable negative interest rates, i.e. $F'_t \geq 0$, for every $t$ in the time horizon. Thus, we replace $F_t$ with $F'_t$ in (2.11), getting:

$$dF'_t = \sigma_t |F'_t|^\beta dW_t^{(1)}$$

$$d\sigma_t = \nu \sigma_t dW_t^{(2)}$$

(2.18)

with (2.12) remaining again unaltered. An advantage of this approach is that, given the original SABR setting, all that is needed to extend to the displaced SABR is a minor modification. Indeed, if we define the displaced strike as $K' := K + d$ and take the displaced underlying $F'_t$ defined as before, we see that:

$$\mathbb{E}[(F_t - K)^+] = \mathbb{E}[(F_t + d) - (K + d)^+] = \mathbb{E}[(F'_t - K')^+]$$
under any probability measure. Hence, pricing a derivative under the original or the Displaced SABR is the same, if we displace both the underlying and the strike. Also, (2.14) becomes

\[
\sigma^{IMP}(K', f') = \left( f' K' \right)^{(1-\beta)/2} \frac{\alpha}{24} \log^2 \left( \frac{f'}{K'} \right) + \frac{(1-\beta)^4}{1920} \log^4 \left( \frac{f'}{K'} \right) + \ldots \cdot \frac{z'}{\chi(z')}
\]

\[
\cdot \left[ 1 + \frac{(1-\beta)^2}{24} \alpha^2 + \frac{1}{4} \frac{\alpha \beta \rho \nu}{f' (1-\beta)} + \frac{2 - 3 \rho^2}{24} \nu^2 T + \ldots \right]
\]

(2.19)

with

\[
z' = \frac{\nu}{\sigma} \left( f' K' \right)^{(1-\beta)/2} \log \left( \frac{f'}{K'} \right)
\]

and

\[
\chi(z) = \log \left( \frac{\sqrt{1 - 2 \rho z + z^2} + z - \rho}{1 - \rho} \right)
\]

while (2.15) becomes

\[
\sigma^{IMP}_{ATM} = \sigma^{IMP}(f', f') = \frac{\alpha}{f'^{(1-\beta)}} \left[ (1-\beta) + \frac{(1-\beta)^2}{24} \alpha^2 + \frac{1}{4} \frac{\alpha \beta \rho \nu}{f'^{(1-\beta)}} + \frac{2 - 3 \rho^2}{24} \nu^2 \right] T + \ldots
\]

(2.20)

A disadvantage of this approach is that the displacement factor cannot be directly observed from the market. Its value can be determined indirectly from the market, by observing the implied volatilities curve extracted by the market premiums. It can also be determined by considering the displacement factor as an extra SABR parameter for calibration/curve fitting. For more details on the displaced SABR, see Hagan & Kumar [20].
3 Methodology

We now present the data we have used and the methodology that we have followed.

3.1 Data

The data used in testing consisted of two sets, one for swap-settled and one for cash-settled swaptions. Each data set contained the (undiscounted) market quoted premiums of swaptions with several expiry/tenor dates, for multiple strike values. Each set also contained a 3 month interest rate curve, a 6 month interest rate curve and an OIS interest rate curve for multiple future dates, in their respective currency (USD for swap-settled, EUR for cash-settled). Both were retrieved from ICAP and had 31/10/2017 as their start date.

3.1.1 Cash Settled (EUR)

For the cash-settled case, the quoted market premiums expiries and maturities belong in the set \{1, 2, 5, 10, 15, 20, 30\} (in years), with each swaption being a combination of two of these numbers denoted as “XyZy”. Hence, e.g., in the following, the “2y5y” swaption corresponds to the swaption with expiry date 2 years after the starting date and swap maturity 5 years after the swaption expiry date. The interest rate curves (3m/6m/OIS) each consist of 45 values ranging from 1 day after the starting date to 60 years after the starting date, thus containing values for the entire time horizon, up to the furthest date of the 30y30y swaption. Since we used the market practice of considering the “spot date\(^1\)” for the expiry/maturity dates, we applied linear interpolation to calculate interest rates of dates not available in our interest rate curve data. The payment frequencies are: once per year for the fixed leg, while for the floating leg every 3 months if the swap maturity is 1 year and every 6 months otherwise.

Then, for a swaption with fixed expiry and maturity dates, there is a premium corresponding to each strike value. The strike values \(K\) are expressed as the distance from the ATM swap rate (2.3), measured in basis points\(^2\) (“bp”) i.e.

\[
K = R_{\text{swap}}^{\text{fwd}}(T)\big|_{t=0} + \text{bp}
\]

The available strikes are for 0, ±6.25, ±12.5, ±25, ±50, ±75, ±100, ±200 and ±300 basis points, with the bp = 0 case clearly being the ATM strike. In our analysis, we ignored all ±300bp swaptions, as well as swaptions with a zero quoted premium. The premiums corresponding to positive basis points belong to payer swaptions, while those of negative basis points belong to receiver swaptions. Hence, from (2.5) we see that all non-ATM swaptions examined here are Out-of-The-Money (“OTM”). The ATM premium is quoted as the sum of a payers ATM and a receivers ATM swaption. However, as discussed in Section 2.2, the payer and receiver cash-settled (EUR) ATM premiums have a non-negligible spread. Hence, we used zero-width collar

\(^1\)Two business days after the actual date.
\(^2\)1 basis point = 10\(^{-4}\).
quotes from ICAP in order to derive the unequal payer and receiver ATM premiums. Lastly, a displacement factor is given for swaptions with maturities of 1, 5, 10, 20 and 30 years; for maturities of 2 and 15 years, we calculate the displacement factor by using linear interpolation.

3.1.2 Swap Settled (USD)

The data given for the swap-settled swaptions were equivalent to the cash-settled swaptions, with the following exceptions. First, the expiry dates belong in the set \{1, 2, 5, 10, 15, 20\} (no 30 years expiries available). Then, the spread between payer and receiver ATM is considered to be negligible, hence the quoted ATM premium is calculated as the sum of the ATM payers and ATM receivers premium. Also, the strikes quoted are for the basis points \(0, \pm 6.25, \pm 12.5, \pm 25, \pm 37.5, \pm 50, \pm 75, \pm 100, \pm 150, \pm 200\) and \(\pm 300\) (\(\pm 37.5\) is included, unlike with the cash-settled swaptions). However, the quoted premium for all \(-300\)bp and almost all \(+300\)bp swaptions was 0, so we omitted all \(+300\)bp swaptions from our analysis. The payment frequencies are: every 3 months for the floating leg and every 6 months for the fixed leg. Lastly, the displacement factor is identical for all maturity dates.

3.2 Calibration

For the calibration of the SABR parameters \(\alpha, \beta, \nu, \rho\), we proceeded with two different approaches, one deterministic and one stochastic. In both approaches, we take \(\beta = 0.5\) and calibrate the parameters for fixed expiry/maturity dates.

Deterministic Approach

In the first approach, we solved (2.20) with respect to \(\alpha\), by picking two arbitrary values for \(\rho\) and \(\nu\) (e.g. \(\rho = 0.5\) and \(\nu = 0.3\)) and taking

\[
\sigma_{ATM}^{\text{IMP}} = \frac{2}{\sqrt{T}} \frac{1}{N^{-1}} \left[ \frac{1}{2} \left( \frac{V_4(0)}{D(0,T) N A_T^*} + 1 \right) \right]
\]

i.e. the solution of Black’s formula for swaptions with respect to the implied volatility, for the ATM case. Here we use the notation of Chapter 2, with \(N^{-1}\) being the inverse of the standard normal cumulative distribution and \(*\in\{\text{swap, cash}\}\). Then, we get \(\rho\) and \(\nu\) through (2.17). A problem with this approach is that it does not guarantee that the minimum it finds is in fact the global minimum. Hence, a poor choice of initial conditions might lead to a local, instead of the global, minimum. A way to improve the local minimum in our case, is to apply the algorithm once using arbitrary values for \(\rho, \nu\) and then use the algorithm’s resulting \(\rho', \nu'\) values as the initial condition of a new algorithm application. This obviously makes this method more time consuming, but leads to solving the problem of a potentially bad choice of initial conditions.

Stochastic Approach

Our second approach was using the Differential Evolution (DE) algorithm, which is a stochastic, global optimization algorithm developed by Storn & Price [37]. In it, the objective function \(f : X \subseteq \mathbb{R}^D \to \mathbb{R}\) with \(X \neq \emptyset\), where \(D\) is the number of parameters to be optimized, does not have to be differentiable, continuous or linear. Differential Evolution allowed us to calibrate all SABR parameters \((\alpha, \rho, \nu)\) simultaneously, without the need of providing initial conditions. The
version of DE that we have applied is DE/best/1/bin, of which we give a brief description.

First, we provide a population number $NP \geq 4$, i.e. the number of $\mathbb{R}^D$ vectors that will be used for the optimization procedure, upper and lower boundaries $x_U, x_L \in \mathbb{R}^D$ to define the space $X$, a “crossover constant” $CR \in [0,1]$ and a termination criterion (either a tolerance level or a maximum number of “generations”). Unless specified otherwise, whenever we make a random decision, a uniform distribution is assumed. The initial vectors are chosen randomly from the entire $X$. Hence, we have

$$x_L \leq x_{i,0} \leq x_U$$

where $x_{i,0} \in \mathbb{R}^D$, for every $i \in \{1, \ldots, NP\}$. We denote the vectors of generation $G \in \mathbb{N}$ as $x_{i,G}$. We call these vectors “agents”. Given that we are in generation $G$, the steps of the algorithm are as such:

1. **Mutation:** For each agent $x_{i,G}$, $i \in \{1, \ldots, NP\}$, randomly choose two indices $r_{1,G}, r_{2,G} \in \{1, \ldots, NP\}$, such that $x_{r_{1,G}} \neq x_{r_{2,G}}, x_{r_{1,G}} \neq x_{i,G}$ and $x_{r_{2,G}} \neq x_{i,G}$. Then, assign

$$v_{i,G+1} := g + F \cdot (x_{r_{1,G}} - x_{r_{2,G}})$$

where $F \in [0,2]$ is a constant factor and $g = (g_1, \ldots, g_D)$ is the agent corresponding to the best position in $X$ until this point, i.e. $f(g) \leq f(x_{i,k}) \forall i \in \{1, \ldots, NP\}$ and $\forall k \in \{1, \ldots, G\}$. We call $v_{i,G+1} = (v_{1i,G+1}, \ldots, v_{Di,G+1}) \in \mathbb{R}^D$ “mutant vectors”.

2. **Crossover:** Now, for every $i \in \{1, \ldots, NP\}$ we introduce a trial vector $u_{i,G+1} = (u_{ji,G+1})_j$, $j \in \{1, \ldots, D\}$ as such:

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & \text{if } rndb(j) \leq CR \text{ or } j = R_i \\ x_{ji,G} & \text{if } rndb(j) > CR \text{ and } j \neq R_i \end{cases}$$

where $rndb(j) \sim \text{Unif}[0,1]$, drawn separately for each $j \in \{1, \ldots, D\}$ and $R_i$ is an index chosen randomly from $\{1, \ldots, D\}$. This ensures that every trial vector contains at least one value from its respective mutant agent.

3. **Selection:** Finally, to decide which agents will be used in generation $G + 1$, we have for every $i \in \{1, \ldots, NP\}$:

$$x_{i,G+1} = \begin{cases} u_{i,G+1} & \text{if } f(u_{i,G+1}) \leq f(x_{i,G}) \\ x_{i,G} & \text{otherwise} \end{cases}$$

Steps 1 to 3 are then repeated until a termination criterion is reached. For more on DE and its other versions, we refer to Fulcher [18] and Pedersen [34].

### 3.3 Simulation

We now describe the methods we have used in order to simulate the value of the underlying at the expiry of the swaption, i.e. $S_T$. These simulations were run multiple times, in order to
approximate $\mathbb{E}[(\omega(S_T - K)^+)]$ through Monte Carlo.

### 3.3.1 Euler Scheme

We start by considering a time horizon $[0, T]$ and dividing it into $N$ time steps $t_i$, where $t_0 = 0$, $t_N = T$ and $t_{i+1} - t_i \equiv \Delta t$, for every $i = 0, \ldots, N - 1$, i.e. $\Delta t = 1/N$. Hence, the discretized version of (2.18)

$$
\begin{align*}
    dF'_t &= \sigma_t F'_t \beta dW^{(1)}_t \\
    d\sigma_t &= \nu \sigma_t dW^{(2)}_t
\end{align*}
$$

may be written as:

$$
\begin{align*}
    F'_{t_{i+1}} &= \sigma_{t_i} F'_t \beta (W^{(1)}_{t_{i+1}} - W^{(1)}_{t_i}) + F'_t \\
    \sigma_{t_{i+1}} &= \nu \sigma_{t_i} (W^{(2)}_{t_{i+1}} - W^{(2)}_{t_i}) + \sigma_{t_i}
\end{align*}
$$

(3.2)

and since for a Brownian Motion $W$ it holds that $W(t) - W(s) \sim N(0, t - s)$, we can rewrite (3.2) as

$$
\begin{align*}
    F'_{t_{i+1}} &= \sigma_{t_i} F'_t \beta (\epsilon^{(1)}_1 \rho + \epsilon^{(2)}_2 \sqrt{1 - \rho^2}) \sqrt{\Delta t} + F'_t \\
    \sigma_{t_{i+1}} &= \nu \sigma_{t_i} \epsilon^{(2)}_1 \sqrt{\Delta t} + \sigma_{t_i}
\end{align*}
$$

(3.3)

with $\epsilon^{(1)} \sim N(0, 1)$ and $\epsilon^{(2)} \sim N(0, 1)$, correlated under (2.12), where $N(0, 1)$ is the standard normal distribution.

Continuing, we apply the Cholesky decomposition to (2.12), giving us:

$$
\begin{align*}
    dW^{(1)}_t &= \rho dW_t + \sqrt{1 - \rho^2} dZ_t \\
    dW^{(2)}_t &= dW_t
\end{align*}
$$

where $W$ and $Z$ are independent standard Brownian Motions. Hence, (3.3) becomes

$$
\begin{align*}
    F'_{t_{i+1}} &= \sigma_{t_i} F'_t \beta \left( \epsilon_1 \rho + \epsilon_2 \sqrt{1 - \rho^2} \right) \sqrt{\Delta t} + F'_t \\
    \sigma_{t_{i+1}} &= \nu \sigma_{t_i} \epsilon_1 \sqrt{\Delta t} + \sigma_{t_i}
\end{align*}
$$

(3.4)

where $\epsilon_1$ and $\epsilon_2$ are independent random variables with $\epsilon_1 \sim N(0, 1)$ and $\epsilon_2 \sim N(0, 1)$. Furthermore, we take initial conditions $F'_0 = \mathbb{E} [R_{\text{swap}}(T) | F_0] + d$ (see (2.3)) with $T$ being the swaption’s expiry date and $d$ the appropriate displacement factor and $\sigma_0 = \alpha$, derived from the curve fitting procedure discussed before.

### Remark 3.1

For the Euler scheme to have a guaranteed rate of strong or weak convergence (for a definition of strong and weak convergence see e.g. [25]), we need that the coefficients of (2.11) are globally Lipschitz continuous [28]. Hence, for the SABR, the classical Euler scheme has no theoretical result for a guaranteed convergence. However, we use it because of its wide
application in mathematical finance.

### 3.3.2 Semi-Exact

Our second simulation method is based on the 2017 paper by Cai, Song and Chen [7]. Their approach focuses on directly simulating the conditional expectation of $F_T$ given $\sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds$ and $F_0$, instead of simulating the entire path in order to reach its terminal value at $T$. By doing so, the inherent bias of time discretization techniques, as in the case of the Euler or Milstein scheme is avoided. Cai et al. [7] present an exact simulation for two special cases and provide a “semi-exact” simulation for the general case. We present the final results of the conditional distribution of $F_T$ and then, for the general case, we describe their approach to simulate the “intermediate” distributions of $\sigma_T$ and $\int_0^T \sigma_s^2 ds$, needed for the simulation of the conditional distribution of $F_T$. Before doing so, we first present some special functions that are used afterwards.

#### Special Functions

A non-central chi-squared random variable $\chi^2(\mu; \lambda)$ with $\mu$ degrees of freedom and noncentrality parameter $\lambda$ has probability density function

$$q_{\chi^2}(x; \mu, \lambda) = \frac{1}{2} \exp \left( -\frac{x + \lambda}{2} \right) \left( \frac{x}{\lambda} \right)^{\frac{a-2}{2}} I_{\mu-2}(\sqrt{\lambda x})$$

where $I_a(\cdot)$ is the modified Bessel function of the first kind with index $a$ given by

$$I_a(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{a+2k}}{k! \Gamma(a+k+1)}$$

where $\Gamma(\cdot)$ is the gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, with $x > 0$ and $a > -1$. If $\lambda = 0$, then we have the special case of the central chi-squared random variable $\chi^2(\mu)$ with $\mu$ degrees of freedom and probability density function

$$q_{\chi^2}(x; \mu) = \frac{e^{-x/2} x^{\mu/2-1}}{2^{\mu/2} \Gamma(\mu/2)}$$

We denote their cumulative distribution functions by $Q_{\chi^2}(x; \mu, \lambda)$ and $Q_{\chi^2}(x; \mu)$ respectively.

Let us now present the results for the conditional distribution of $F_T$. The following proposition can be found in Islah [26].

**Proposition 3.2.** Fix $T > 0$ and suppose that $\sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds$ and $F_0$ are given.

1. Let $\beta = 1$. Then, $F_t$ is log-normally distributed and through (2.13), we have that

$$\log F_t \sim N \left( \log F_0 - \frac{1}{2} \int_0^T \sigma_s^2 ds + \frac{\rho}{\nu} (\sigma_T - \sigma_0), (1 - \rho^2) \int_0^T \sigma_s^2 ds \right)$$
(2) Let \(0 \leq \beta < 1, \rho = 0\) and assume that \(F_t\) has an absorbing boundary at 0. Then:

\[
P[F_T = 0 \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A_0; \frac{1}{1-\beta} \right)
\]

and

\[
P[F_T \leq u \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A_0; \frac{1}{1-\beta}, C_0(u) \right), \quad u > 0
\]

where

\[
A_0 = \frac{1}{\int_0^T \sigma_s^2 ds} \left( F_0^{1-\beta} \right)^2 \quad \text{and} \quad C_0(u) = \frac{1}{\int_0^T \sigma_s^2 ds} \cdot \frac{u^{2(1-\beta)}}{(1-\beta)^2}
\]

(3) Let \(\rho \neq 0, 0 \leq \beta < 1\) and assume that \(F_t\) has an absorbing boundary at 0. Then:

\[
P[F_T = 0 \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A; 1 + \frac{\beta}{(1-\beta)(1-\rho^2)} \right)
\]

and

\[
P[F_T \leq u \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A; 1 + \frac{\beta}{(1-\beta)(1-\rho^2)}, C(u) \right), \quad u > 0
\]

where

\[
A = \frac{1}{(1-\rho^2) \int_0^T \sigma_s^2 ds} \left( F_0^{1-\beta} \frac{\rho}{\nu} (\sigma_T - \sigma_0) \right)^2
\]

and

\[
C(u) = \frac{1}{(1-\rho^2) \int_0^T \sigma_s^2 ds} \cdot \frac{u^{2(1-\beta)}}{(1-\beta)^2}. \quad (3.8)
\]

In our setting, we have calculated values for \(\sigma_0\) and \(F_0\), hence we only need to simulate values for \(\sigma_T\) (straightforward through (2.13)) and \(\int_0^T \sigma_s^2 ds\). Thus we focus on the simulation of \(\int_0^T \sigma_s^2 ds\), given \(\sigma_0\) and \(\sigma_T\). Cai et al. [7] used results from Yor [38] to show the following.

**Proposition 3.3.** The density function of \(\int_0^T \sigma_s^2 ds\), conditioning on \(\sigma_0\) and \(\sigma_T\) is given by

\[
P \left[ \int_0^T \sigma_s^2 ds \in dw \mid \sigma_0, \sigma_T \right] = \frac{\nu \sqrt{2\pi T}}{w} \exp \left\{ \frac{1}{2\nu^2} \left( \frac{1}{T} \left[ \ln \left( \frac{\sigma_T}{\sigma_0} \right) \right]^2 - \frac{\sigma_T^2 + \sigma_0^2}{w} \right) \right\} \cdot I_0(\hat{r}) f_\nu(\nu^2 T) dw
\]
where $I_a(\cdot)$ is the modified Bessel function of the first kind with index $a$, $\hat{\tau} \equiv \hat{\tau}(w) = \sigma_0 \sigma_T / (\nu^2 w)$ and $f_r(\cdot)$ with $r > 0$ is the Hartman-Watson density function:

$$f_r(t) = \frac{1}{I_0(r)} \frac{r}{\sqrt{2\pi t}} \exp\left(\frac{\pi^2}{2t}\right) \cdot \int_0^\infty \exp\left(-\frac{y^2}{2t}\right) \exp\left(-r \cosh(y)\right) \sinh(y) \sin\left(\frac{\pi y}{t}\right) dy, \ t > 0.$$ 

Even though (3.3) provides an explicit expression for $\int_0^T \sigma_s^2 ds$, the Hartman-Watson density function is “extremely difficult to compute for small $t$” [7], because the term outside the integral $(r/\sqrt{2\pi^3 t}) \cdot \exp(\pi^2/2t)$ exponentially, while the term $\sin(\pi y/t)$ inside the integral changes its sign more and more frequently as $t \to 0$. To achieve acceptable results, high-precision computations are required, something that is not always computationally feasible. To tackle this problem, Cai et al. [7] employed the following approach: instead of numerically calculating the conditional distribution directly, they used an appropriate bijection $h : \mathbb{R}^+ \to \mathbb{R}^+$, such that the Laplace transform of the cumulative distribution function of $h(\int_0^T \sigma_s^2 ds)$, given $\sigma_0$ and $\sigma_T$ has a closed form. We denote the cumulative distribution function of $h(\int_0^T \sigma_s^2 ds)$ given $\sigma_0$ and $\sigma_T$ as $L_h(\cdot)$ and its Laplace transform as $\hat{L}_h(\cdot)$ with

$$\hat{L}_h(\theta) := \int_0^\infty e^{-\theta u} L_h(u) du, \ \theta > 0.$$

Then, we have the following proposition.

**Proposition 3.4.** For $h(x) = 1/x$, the Laplace transform of $L_h(\cdot)$ is:

$$\hat{L}_h(\theta) = \frac{1}{\theta} \exp\left\{ -\frac{\left[\phi_{ln(\sigma_T/\sigma_0)}(\theta \nu^2 / \sigma_0^2)\right]^2 - \left[\ln(\sigma_T/\sigma_0)\right]^2}{2\nu^2 T} \right\}$$

with

$$\phi_x(\lambda) = \text{arcosh}(\lambda e^{-x} + \cosh(x)),$$

where $\cosh(y) = (e^y + e^{-y})/2$ and $\text{arcosh}(z) = (\ln(z + \sqrt{z^2 + 1})$.

We thus see that choosing $h(x) = 1/x$ leads to a Laplace transform that consists only of relatively simple functions. Continuing, applying the Euler inversion algorithm by Abate and Whitt [1], we may calculate $L_h(\cdot)$ from $\hat{L}_h(\cdot)$ by

$$L_h(u) = \frac{e^{M/2}}{2u} \text{Re}\left(\hat{L}_h\left(\frac{M}{2u}\right)\right) + \frac{e^{M/2}}{2u} \sum_{k=1}^\infty (-1)^k \text{Re}\left(\hat{L}_h\left(\frac{M - 2k\pi i}{2u}\right)\right) - e_d, \tag{3.9}$$

where $\text{Re}(x)$ is the real part of $x$, $i$ is the imaginary unit $i = \sqrt{-1}$, $M$ is a positive constant and


\[ e_d = e_d(L_h, u, M) = \sum_{k=1}^{\infty} e^{-kM} L_h((2k + 1)u). \]

Since \( L_h(\cdot) \) is a cumulative distribution function, it is non-decreasing and \( 0 \leq L_h(\cdot) \leq 1 \). Hence, if \( L_h(\cdot) \equiv 0 \) we have \( e_d = 0 \), while if \( L_h(\cdot) \equiv 1 \), then

\[ e_d = \sum_{k=1}^{\infty} e^{-kM} = \frac{1}{e^M - 1} = \frac{e^{-M}}{1 - e^{-M}}. \]

Hence, we can bound \( e_d \)

\[ 0 \leq e_d \leq \frac{e^{-M}}{1 - e^{-M}} \simeq e^{-M} \tag{3.10} \]

for sufficiently large \( M \). Additionally, we approximate the alternating power series of (3.9) by using the Euler transformation

\[ \sum_{k=0}^{\infty} (-1)^k a_k \simeq E(m, n) := \sum_{k=0}^{m} \frac{m!}{k!(m-k)!} 2^{-m} S_{n+k} \tag{3.11} \]

with

\[ S_j := \sum_{k=0}^{j} (-1)^k a_k. \]

In our implementation, we follow the paper’s suggestion and take \( M = 20 \), \( m = 20 \) and \( n = 35 \). Hence, (3.9) using the approximations (3.10) and (3.11) becomes:

\[
L_h(u) = \frac{e^{M/2}}{2u} \text{Re} \left( \hat{L}_h \left( \frac{M}{2u} \right) \right) \\
+ \frac{e^{M/2}}{u} \sum_{k=0}^{m} \left\{ \frac{m!}{k!(m-k)!} 2^{-m} \sum_{j=0}^{n+k} (-1)^j \text{Re} \left( \hat{L}_h \left( \frac{M-2j\pi i}{2u} \right) \right) \right\} \\
- \frac{e^{M/2}}{u} \text{Re} \left( \hat{L}_h \left( \frac{M}{2u} \right) \right) - e^{-M}. \tag{3.12}
\]

Thus, through (3.12) we have a computationally efficient formula for the cumulative distribution function of \( h(\int_0^T \sigma^2_s ds) = 1/ \int_0^T \sigma^2_s ds \) given \( \sigma_0, \sigma_T \). Hence, to sample \( \int_0^T \sigma^2_s ds \) given \( \sigma_0, \sigma_T \), we work as such: we simulate \( U \sim \text{Unif}[0, 1] \) and solve \( L_h(V) = U \) numerically using some root
finder algorithm. Then, $V$ has the same distribution as $1/ \int_0^T \sigma_s^2 ds$ given $\sigma_0, \sigma_T$, so

$$1/V \sim \int_0^T \sigma_s^2 ds \text{ given } \sigma_0, \sigma_T.$$

We are now ready to simulate the distribution of $F_T$ given $F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds$, using (3.5) and (3.6). To do so, we work as before and simulate $U \sim \text{Unif}[0,1]$. Then, if

$$U \leq \mathbb{P}[F_T = 0 \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A; 1 + \frac{\beta}{(1 - \beta)(1 - \rho^2)} \right)$$

we set $F_T = 0$. Otherwise, we numerically solve

$$U = \mathbb{P}[F_T \leq \hat{V} \mid F_0, \sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds] = 1 - Q_{\chi^2} \left( A; 1 + \frac{\beta}{(1 - \beta)(1 - \rho^2)}, C(\hat{V}) \right)$$

and assign $F_T = \hat{V}$ using the inversion technique. We thus have a procedure to directly simulate $F_T$ given $\sigma_0, \sigma_T, \int_0^T \sigma_s^2 ds$ and $F_0$.

### 3.4 Bid and Ask

#### 3.4.1 Market Values

Our original data set contained only the mid price (undiscounted) premiums. However, our ultimate goal was to work under the Conic Finance framework and as such estimate the bid and ask prices directly. In order to attain the bid and ask market prices for our start date, we used the Bloomberg Terminal. There, we were able to find separate bid and ask prices only for ATM swaptions. We extracted these values, along with the mid price values quoted in the Bloomberg Terminal and created Mid/Bid and Mid/Ask ratios, with which we divided our original data in order to construct the market bid/ask premiums. However, as can be seen in Tables 3.1 and 3.2, the ratios are very close to 1, which can be attributed to their high liquidity. Indeed, for the EUR swaptions the largest deviation of a bid or ask price from the mid price is almost 3.8%, for the 1y1y swaption, while for most swaptions the ratios are equal to 1 up to four decimal points. For USD swaptions, we have bid = mid = ask for most cases. We thus have that, in general, bid $\simeq$ mid $\simeq$ ask, and hence selecting the identity function as the distortion function $\Psi$ of (1.15) and (1.16) would not cause significant errors; however, this would be identical to working under the one-price framework. To deal with this problem, we added an extra artificial spread by directly adding (subtracting) a quantity to the ATM Mid/Bid (Mid/Ask) ratio. Then, we needed to reflect the fact that the furthest OTM a swaption is, the less liquid it is. To do so, we first applied the (augmented) ATM ratio on all the swaptions of the expiry/maturity date in question. Then, for the furthest OTM premium, we added to it an extra spread, calculated as 5% of its premium after the ATM ratio was applied. Then, for the rest of the premiums, we added a spread in a similar way, with the percentage linearly reducing with respect to the bp’s from ATM, such that the ATM premium received 0% extra spread. In the following, we present the ATM Mid/Bid and Mid/Ask ratios of EUR and USD swaptions with expiry dates up to 10 years.
Table 3.1: ATM Mid/Bid and Mid/Ask ratios, for EUR swaptions

<table>
<thead>
<tr>
<th>Swaption</th>
<th>Mid/Bid</th>
<th>Mid/Ask</th>
</tr>
</thead>
<tbody>
<tr>
<td>1y1y</td>
<td>1.037865</td>
<td>0.962135</td>
</tr>
<tr>
<td>1y2y</td>
<td>1.033688</td>
<td>0.966312</td>
</tr>
<tr>
<td>1y5y</td>
<td>1.021159</td>
<td>0.978841</td>
</tr>
<tr>
<td>1y10y</td>
<td>1.002777</td>
<td>0.999723</td>
</tr>
<tr>
<td>1y15y</td>
<td>1.000094</td>
<td>0.999887</td>
</tr>
<tr>
<td>1y20y</td>
<td>1.000044</td>
<td>0.999956</td>
</tr>
<tr>
<td>1y30y</td>
<td>1.000063</td>
<td>0.999937</td>
</tr>
<tr>
<td>2y1y</td>
<td>1.002937</td>
<td>0.996815</td>
</tr>
<tr>
<td>2y2y</td>
<td>1.003024</td>
<td>0.996994</td>
</tr>
<tr>
<td>2y5y</td>
<td>1.000037</td>
<td>0.999963</td>
</tr>
<tr>
<td>2y10y</td>
<td>1.000872</td>
<td>0.999112</td>
</tr>
<tr>
<td>2y15y</td>
<td>1.000038</td>
<td>0.999952</td>
</tr>
<tr>
<td>2y20y</td>
<td>1.000063</td>
<td>0.999937</td>
</tr>
<tr>
<td>2y30y</td>
<td>1.000063</td>
<td>0.999937</td>
</tr>
</tbody>
</table>

Table 3.2: ATM Mid/Bid and Mid/Ask ratios, for USD swaptions

<table>
<thead>
<tr>
<th>Swaption</th>
<th>Mid/Bid</th>
<th>Mid/Ask</th>
</tr>
</thead>
<tbody>
<tr>
<td>1y1y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y2y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y5y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y10y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y15y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y20y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1y30y</td>
<td>1.003279</td>
<td>0.996742</td>
</tr>
<tr>
<td>2y1y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y2y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y5y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y10y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y15y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y20y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2y30y</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Swaption</th>
<th>Mid/Bid</th>
<th>Mid/Ask</th>
</tr>
</thead>
<tbody>
<tr>
<td>5y1y</td>
<td>1.000094</td>
<td>0.999811</td>
</tr>
<tr>
<td>5y2y</td>
<td>1.007048</td>
<td>0.99305</td>
</tr>
<tr>
<td>5y5y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5y10y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5y15y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5y20y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5y30y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y1y</td>
<td>1.000077</td>
<td>0.999919</td>
</tr>
<tr>
<td>10y2y</td>
<td>1.004777</td>
<td>0.995299</td>
</tr>
<tr>
<td>10y5y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y10y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y15y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y20y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y30y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10y10y</td>
<td>1.000021</td>
<td>0.999979</td>
</tr>
<tr>
<td>10y15y</td>
<td>1.000031</td>
<td>0.999969</td>
</tr>
<tr>
<td>10y20y</td>
<td>1.000347</td>
<td>0.999658</td>
</tr>
<tr>
<td>10y30y</td>
<td>1.000212</td>
<td>0.999791</td>
</tr>
</tbody>
</table>

Table 3.2: ATM Mid/Bid and Mid/Ask ratios, for USD swaptions

**3.4.2 Pricing**

In order to reach a direct numerical pricing formula for the bid and ask values, we work as such: for a fixed expiry and maturity date, we simulate $M$ values $\tilde{S}_T$ of the underlying at expiry, using either approach of Section 3.3. Then, from (1.20), we have that the bid price at time $t = 0$ can be approximated by:

$$\text{bid}_*(0) = NA_t \sum_{i=1}^M \left\{ \left( \Psi \left( \frac{i}{M} \right) - \Psi \left( \frac{i-1}{M} \right) \right) \left[ \omega(\tilde{S}_{T(i)} - K) + \right] \right\},$$

(3.13)
where \( \ast \in \{ \text{swap}, \text{cash} \} \), \( \Psi \) is a concave distortion function as described in Section 1.5, \( \tilde{S}_{T(i)} \) denotes the ordered simulated values \( \tilde{S}_{T(i)} \), i.e. \( \tilde{S}_{T(i)} \leq \tilde{S}_{T(j)} \) for every \( i, j \in \{1, \ldots, M\} \) with \( i \leq j \), while the rest follows the notation of Section 2.2. Also, for the ask price we have through (1.21) that:

\[
\text{ask}_\ast(0) = N A_1^\ast \sum_{i=1}^{M} \left\{ \left( \Psi \left( \frac{M - i + 1}{M} \right) - \Psi \left( \frac{M - i}{M} \right) \right) \left[ \omega (\tilde{S}_{T(i)} - K^+) \right] \right\}.
\]

with the same notation. We have considered the distortion functions \( \text{MINVAR}, \text{MAXVAR} \) and \( \text{MINMAXVAR} \) of Section 1.6 for \( \Psi \), which are parametric functions depending on \( \lambda \). We did not consider \( \text{MAXMINVAR} \) because its results were too similar to \( \text{MINMAXVAR} \) and \( \text{WANG} \) because it was too computationally expensive. For fixed expiry and tenor dates, each strike \( K \) corresponds to a different swaption, with a different bid and ask value. Thus, we calibrate \( \lambda \) for each strike \( K \), through least squares minimization of the corresponding bid and ask price, i.e.

\[
\lambda(K) = \arg \min_{\lambda \geq 0} \left( (\text{ask}_\ast(\lambda, K) - \text{ask}_{MKT}^\ast(K))^2 + (\text{bid}_\ast(\lambda, K) - \text{bid}_{MKT}^\ast(K))^2 \right).
\]

This approach also allows us to price products with strikes that are not quoted, through interpolation or extrapolation.
4 Results

4.1 Implied Volatility Curve Fitting

Stochastic vs Deterministic Optimization
We now present results from the deterministic and stochastic calibration approach, which we compare in order to decide which of them better suits our needs.

In Fig. 4.1 we present the implied volatilities curve fitting for the 2y5y EUR and USD swaption, along with their respective absolute residuals in table 4.1. These results were derived using $\rho = 0.5$ and $\nu = 0.1$ as initial conditions for the deterministic curve fitting.

![Figure 4.1: Implied Volatilities Curve Fitting, 2y5y EUR/USD Swaption.](image)

We see that in both cases, the differential evolution algorithm is more accurate, with the exception of some points close to ATM in the USD case. This is because we take the SABR parameter $\alpha$ to be the solution of (2.15), using the analytic value of $\sigma^{IMP}_{ATM}$ from (3.1), hence “forcing” the curve to pass through its ATM market value and thus be precise. We see that in the USD case, the differential evolution results are closer to the market implied volatilities (with the exception of those close to ATM), while the deterministic approach yields slightly better results in the EUR case.
Table 4.1: 2y5y USD absolute residuals.

However, for cash-settled (EUR) swaptions, the ATM payers and ATM receivers premiums are quoted differently. To deal with this, we work separately for payers (i.e. \( bp \geq 0 \)) and receivers (\( bp \leq 0 \)) swaptions, which leads to the following implied volatility curve fitting:

![Figure 4.2: Distinct EUR payers/receivers calibration](image)

In Fig. 4.2 below, we see the absolute residuals, where by \( \sigma_{MKT} \) we denote the market quoted implied volatility, by \( \sigma_{det} \) the implied volatility calibrated through the deterministic curve fitting and \( \sigma_{stoch} \) the implied volatility calibrated through the stochastic curve fitting.
We therefore see that differential evolution is again more accurate. In Figures 4.1 and 4.2, we used $\rho = 0.5$ and $\nu = 0.1$ as initial conditions for the deterministic calibration. We now present the same curve fitting plots as Fig. 4.2, but instead of running a single deterministic calibration, the values of $\rho$ and $\nu$ are now derived as such: first, we run the algorithm using arbitrary $\rho, \nu$ values (in this case $\rho = 0.5, \nu = 0.1$). Then, we take the resulting values of $\rho$ and $\nu$ and use them as the initial values input of a new algorithm run. In Fig. 4.3, we applied this scheme for 10 iterations and compare it to the stochastic calibration.

As we can see, the two approaches yield almost identical results, especially for the payers swaptions. Hence we may consider the stochastic approach to be the “asymptotic” result of multiple deterministic calibration feedback loops. In the following tables we compare the results of the stochastic calibration versus the results of a) 50 “nested” deterministic calibration and b) a single deterministic calibration for the 2y5y EUR swaption. By $\sigma$ we denote the implied volatilities, while $-0$ and $+0$ corresponds to receivers and payers ATM.

| bp  | $|\sigma_{\text{det}} - \sigma_{\text{MKT}}|$ | $|\sigma_{\text{stoch}} - \sigma_{\text{MKT}}|$ |
|-----|---------------------------------|---------------------------------|
| -200.00 | 1.537669e-02                  | 5.294844e-03                  |
| -150.00 | 3.853840e-04                  | 9.475504e-03                  |
| -100.00 | 7.478148e-03                  | 2.543871e-03                  |
| -75.00 | 1.256775e-02                  | 2.821493e-03                  |
| -50.00 | 1.344750e-02                  | 4.889329e-03                  |
| -25.00 | 8.674814e-03                  | 2.502501e-03                  |
| -12.50 | 4.676801e-03                  | 6.343639e-05                  |
| -6.25 | 2.400014e-03                  | 1.623611e-03                  |
| 0.00  | 3.608225e-16                  | 3.323990e-03                  |

(a) Receivers

| bp  | $|\sigma_{\text{det}} - \sigma_{\text{MKT}}|$ | $|\sigma_{\text{stoch}} - \sigma_{\text{MKT}}|$ |
|-----|---------------------------------|---------------------------------|
| 0.00 | 2.775558e-17                   | 8.791231e-05                   |
| 6.25 | 3.708313e-04                   | 5.019230e-05                   |
| 12.50 | 5.200562e-04                   | 4.793734e-05                   |
| 25.00 | 5.916608e-04                   | 2.406848e-05                   |
| 50.00 | 1.426937e-04                   | 7.725024e-06                   |
| 75.00 | 7.756792e-04                   | 2.130045e-05                   |
| 100.00 | 1.916782e-03                  | 2.264022e-05                   |
| 150.00 | 4.368556e-03                  | 4.405846e-06                   |
| 200.00 | 6.759454e-03                  | 1.277863e-05                   |

(b) Payers

Table 4.2: 2y5y EUR absolute residuals, Payers/Receivers calibrated separately.
Table 4.3: Calibration comparison after 50 “nested” deterministic runs.

| bp    | $\sigma_{det}$ | $\sigma_{stoch}$ | $\sigma_{MKT}$ | $|\sigma_{det} - \sigma_{MKT}|$ | $|\sigma_{stoch} - \sigma_{MKT}|$ | $|\sigma_{det} - \sigma_{stoch}|$ |
|-------|-----------------|------------------|----------------|-------------------------------|-------------------------------|-------------------------------|
| -200.00 | 0.474417 | 0.474022 | 0.479705 | 5.287723e-03 | 5.303226e-03 | 1.550283e-05 |
| -150.00 | 0.318725 | 0.318016 | 0.308539 | 1.018595e-02 | 9.476374e-03 | 7.095761e-04 |
| -100.00 | 0.242977 | 0.242314 | 0.239764 | 3.212411e-03 | 2.549768e-03 | 6.626430e-04 |
| -75.00  | 0.220443 | 0.220220 | 0.223035 | 2.592139e-03 | 2.814700e-03 | 2.225613e-04 |
| -50.00  | 0.207547 | 0.208230 | 0.213113 | 5.566681e-03 | 4.883886e-03 | 6.827947e-04 |
| -25.00  | 0.204177 | 0.206165 | 0.208666 | 4.485802e-03 | 2.501179e-03 | 1.987323e-03 |
| -12.50  | 0.205489 | 0.208163 | 0.208101 | 2.612042e-03 | 6.20721e-05 | 2.550014e-03 |
| -6.25   | 0.206716 | 0.209720 | 0.208099 | 1.383270e-03 | 1.607822e-03 | 2.375123e-04 |
| 0.00    | 0.208250 | 0.211569 | 0.208250 | 2.775558e-17 | 3.19742e-03 | 3.19742e-03 |
| +0.00   | 0.208516 | 0.208428 | 0.208516 | 2.775558e-17 | 8.788541e-05 | 8.788541e-05 |
| 6.25    | 0.208920 | 0.208830 | 0.208780 | 1.403833e-04 | 5.020379e-05 | 9.017950e-05 |
| 12.50   | 0.209472 | 0.209381 | 0.209333 | 1.388192e-04 | 4.793939e-05 | 9.088327e-05 |
| 25.00   | 0.210941 | 0.210853 | 0.210829 | 1.123232e-04 | 2.404800e-05 | 8.828427e-05 |
| 50.00   | 0.214909 | 0.214837 | 0.214845 | 6.404105e-05 | 7.762815e-06 | 5.627823e-05 |
| 75.00   | 0.219641 | 0.219594 | 0.219615 | 2.577713e-05 | 2.133728e-05 | 4.439856e-06 |
| 100.00  | 0.224669 | 0.224649 | 0.224672 | 3.344939e-06 | 2.266588e-05 | 1.932094e-05 |
| 150.00  | 0.234686 | 0.234722 | 0.234718 | 3.196650e-05 | 4.157696e-06 | 2.755073e-05 |
| 200.00  | 0.244032 | 0.244119 | 0.244106 | 7.425140e-05 | 1.282885e-05 | 6.142255e-05 |

Table 4.4: Calibration comparison after one deterministic run with $\rho = 0.5$ and $\nu = 0.1$. 

51
Also of interest are the values of the calibrated parameters:

\[ (\alpha, \rho, \nu)_{\text{det}1} = (0.03120965, 0.34491170, 0.37975913) \]
\[ (\alpha, \rho, \nu)_{\text{det}50} = (0.03091384, 0.31036060, 0.40415198) \]
\[ (\alpha, \rho, \nu)_{\text{stoch}} = (0.03089498, 0.30851779, 0.40560207) \]

with

\[ |(\alpha, \rho, \nu)_{\text{det}1} - (\alpha, \rho, \nu)_{\text{stoch}}| = (0.03146712, 3.63939149, 2.58429411) \times 10^{-2} \]
\[ |(\alpha, \rho, \nu)_{\text{det}50} - (\alpha, \rho, \nu)_{\text{stoch}}| = (0.00188574, 1.84281471, 1.45009262) \times 10^{-3}. \]

Which further shows the parameters calibrated through the deterministic approach converge to the ones calibrated through the stochastic approach. Finally, to demonstrate the sensitivity of the curve fitting with respect to the choice of the initial conditions, we present some plots with different initial \( \rho, \nu \) conditions.

![Figure 4.4: Deterministic Curve Fitting for different initial conditions.](image)

**SABR calibration**

We now present the curve fitting results for the swaptions that we have considered, including the average sum of the squares residual. We start with EUR swaptions. All results are with the Differential Evolution algorithm.
We now present the plot corresponding to the 20y30y EUR swaption, in order to more clearly demonstrate the ATM jump that led us to calibrate payers and receivers separately, for the EUR swaptions.

Finally, we present some the implied volatility curve fitting plots for the USD swaptions we have considered.

4.2 Distortion functions Comparison and $\lambda$ interpolation

In this section, we start by comparing the different distortion functions and then we examine the possibility of interpolating the distortion parameter $\lambda$ for swaptions with unknown bid and
ask prices. In what follows, we consider the 1y2y USD swaption, for which we simulated 100,000 values through the Semi-Exact approach, in order to apply (Conic) Monte Carlo. In what follows, our bid and ask values are constructed by applying an extra 10% spread to prices corresponding to all strikes, plus an additional diminishing 5% spread to the OTM values. The following table contains the bid values acquired through each distortion function:

<table>
<thead>
<tr>
<th>bp</th>
<th>bid</th>
<th>MINMAXVAR</th>
<th>MINVAR</th>
<th>MAXVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>-200.00</td>
<td>0.008636</td>
<td>0.005252</td>
<td>0.008177</td>
<td>0.006792</td>
</tr>
<tr>
<td>-150.00</td>
<td>0.166250</td>
<td>0.156518</td>
<td>0.160176</td>
<td>0.176369</td>
</tr>
<tr>
<td>-100.00</td>
<td>1.542273</td>
<td>1.523006</td>
<td>1.490643</td>
<td>1.672679</td>
</tr>
<tr>
<td>-75.00</td>
<td>4.139091</td>
<td>4.092620</td>
<td>3.992828</td>
<td>4.434330</td>
</tr>
<tr>
<td>-50.00</td>
<td>9.839091</td>
<td>9.903433</td>
<td>9.566129</td>
<td>10.575179</td>
</tr>
<tr>
<td>-37.50</td>
<td>14.589205</td>
<td>14.673125</td>
<td>14.19980</td>
<td>15.547120</td>
</tr>
<tr>
<td>-25.00</td>
<td>20.995227</td>
<td>21.072735</td>
<td>20.454542</td>
<td>22.149683</td>
</tr>
<tr>
<td>-12.50</td>
<td>29.335312</td>
<td>29.280243</td>
<td>28.551976</td>
<td>30.526782</td>
</tr>
<tr>
<td>-6.25</td>
<td>34.264560</td>
<td>34.119178</td>
<td>33.339929</td>
<td>35.427707</td>
</tr>
<tr>
<td>0.00</td>
<td>79.354545</td>
<td>78.834747</td>
<td>77.713083</td>
<td>80.362501</td>
</tr>
<tr>
<td>6.25</td>
<td>34.446094</td>
<td>34.298346</td>
<td>33.480807</td>
<td>35.736678</td>
</tr>
<tr>
<td>12.50</td>
<td>29.743125</td>
<td>29.715325</td>
<td>28.939567</td>
<td>31.082289</td>
</tr>
<tr>
<td>25.00</td>
<td>21.745057</td>
<td>21.970848</td>
<td>21.256052</td>
<td>23.167915</td>
</tr>
<tr>
<td>37.50</td>
<td>15.579830</td>
<td>15.910880</td>
<td>15.286312</td>
<td>16.907993</td>
</tr>
<tr>
<td>50.00</td>
<td>10.979205</td>
<td>11.309147</td>
<td>10.796021</td>
<td>12.103912</td>
</tr>
<tr>
<td>75.00</td>
<td>5.271989</td>
<td>5.397746</td>
<td>5.154557</td>
<td>5.849643</td>
</tr>
<tr>
<td>100.00</td>
<td>2.384318</td>
<td>2.475576</td>
<td>2.39566</td>
<td>2.712386</td>
</tr>
<tr>
<td>150.00</td>
<td>0.481250</td>
<td>0.498987</td>
<td>0.471522</td>
<td>0.556851</td>
</tr>
<tr>
<td>200.00</td>
<td>0.103636</td>
<td>0.112352</td>
<td>0.102355</td>
<td>0.126712</td>
</tr>
</tbody>
</table>

Table 4.5: Simulated bid values for the 1y2y USD swaption.

We may see the residuals in the following plots:

![1y2y USD bid residuals](image1)

![1y2y USD bid relative residuals](image2)

Figure 4.5: 1y2y USD bid residuals.

We see that, with the exception of the extreme values of $-200bp$ and $+200bp$, the relative residuals of the MINVAR and MINMAXVAR distortions are always under 5%. This is because these values are very small and as such any deviation from the market value causes a large relative residual. Indeed, if we restrict ourselves to products with bid and ask values larger than
1.00, we have the following residuals:

![1y2y USD bid residuals](image)

Figure 4.6: 1y2y USD bid residuals for $-100 \leq \text{bp} \leq +100$.

We therefore see that the residuals of MINVAR and MINMAXVAR are small everywhere. In order to be able to select the distortion that gives the best results, we also need to see the simulated values for the ask price, which we cite below:

<table>
<thead>
<tr>
<th>bp</th>
<th>ask</th>
<th>MINMAXVAR</th>
<th>MINVAR</th>
<th>MAXVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>-200.00</td>
<td>0.011667</td>
<td>0.009726</td>
<td>0.007074</td>
<td>0.011536</td>
</tr>
<tr>
<td>-150.00</td>
<td>0.219028</td>
<td>0.211548</td>
<td>0.186023</td>
<td>0.220410</td>
</tr>
<tr>
<td>-100.00</td>
<td>1.981667</td>
<td>1.966102</td>
<td>1.794436</td>
<td>2.008044</td>
</tr>
<tr>
<td>-75.00</td>
<td>5.252222</td>
<td>5.214001</td>
<td>4.827488</td>
<td>5.327143</td>
</tr>
<tr>
<td>-50.00</td>
<td>12.330000</td>
<td>12.384190</td>
<td>11.706091</td>
<td>12.569802</td>
</tr>
<tr>
<td>-37.50</td>
<td>18.168750</td>
<td>18.240131</td>
<td>17.374208</td>
<td>18.520175</td>
</tr>
<tr>
<td>-25.00</td>
<td>25.983611</td>
<td>26.050213</td>
<td>24.994136</td>
<td>26.459550</td>
</tr>
<tr>
<td>-12.50</td>
<td>36.079062</td>
<td>36.032820</td>
<td>34.784747</td>
<td>36.628016</td>
</tr>
<tr>
<td>-6.25</td>
<td>42.009983</td>
<td>41.883091</td>
<td>40.554395</td>
<td>42.575648</td>
</tr>
<tr>
<td>0.00</td>
<td>96.988889</td>
<td>96.533511</td>
<td>95.059436</td>
<td>97.632351</td>
</tr>
<tr>
<td>6.25</td>
<td>42.232552</td>
<td>42.104586</td>
<td>40.632085</td>
<td>42.821727</td>
</tr>
<tr>
<td>12.50</td>
<td>36.580625</td>
<td>36.556656</td>
<td>35.183966</td>
<td>37.161594</td>
</tr>
<tr>
<td>25.00</td>
<td>26.911597</td>
<td>27.104787</td>
<td>25.976146</td>
<td>27.469667</td>
</tr>
<tr>
<td>37.50</td>
<td>19.402431</td>
<td>19.683706</td>
<td>18.783102</td>
<td>19.872914</td>
</tr>
<tr>
<td>50.00</td>
<td>13.758750</td>
<td>14.037167</td>
<td>13.332176</td>
<td>14.118494</td>
</tr>
<tr>
<td>75.00</td>
<td>6.689792</td>
<td>6.793978</td>
<td>6.353886</td>
<td>6.838802</td>
</tr>
<tr>
<td>100.00</td>
<td>3.063611</td>
<td>3.138348</td>
<td>2.909939</td>
<td>3.133785</td>
</tr>
<tr>
<td>150.00</td>
<td>0.634028</td>
<td>0.648143</td>
<td>0.588075</td>
<td>0.645609</td>
</tr>
<tr>
<td>200.00</td>
<td>0.140000</td>
<td>0.146866</td>
<td>0.132641</td>
<td>0.142865</td>
</tr>
</tbody>
</table>

Table 4.6: Simulated ask values for the 1y2y USD swaption.

In the same manner as with the bid prices, we present the plots of the ask (relative) residuals below.
We observe that, as with the bid case, we have large residuals for the outermost OTM swaptions. We also see that in this case MAXVAR outperforms MINVAR, while MINMAXVAR performs consistently well. We now present the residuals for basis points which correspond to ask values larger than 1.00:

![Figure 4.7: 1y2y USD ask residuals.](image)

We thus conclude that MINMAXVAR performs better than both MINVAR and MAXVAR, as it produces good results for both bid and ask values. Indeed, we have the following for the absolute residuals of the bid values:

\[
\sum_{i: K_i} |\text{bid}_i - \text{bid}_i^{MIN}| = 7.6644720
\]

\[
\sum_{i: K_i} |\text{bid}_i - \text{bid}_i^{MAX}| = 14.158785
\]

\[
\sum_{i: K_i} |\text{bid}_i - \text{bid}_i^{MINMAX}| = 2.3288720
\]
and for the absolute residuals of the asks:

\[ \sum_{i:K_i} |\text{ask}_i - \text{ask}^{\text{MIN}}_i| = 13.258215 \]
\[ \sum_{i:K_i} |\text{ask}_i - \text{ask}^{\text{MAX}}_i| = 5.7201290 \]
\[ \sum_{i:K_i} |\text{ask}_i - \text{ask}^{\text{MINMAX}}_i| = 1.9886130. \]

Hence, from now on, we will only focus on the \textit{MINMAX} distortion function. Also of interest are the values of the calibrated \( \lambda = \lambda(K) \), a plot of which we present here, for strikes corresponding to bid/ask values more than 1:

![Figure 4.9: MINMAX distortion parameter \( \lambda(K) \).](image)

The spike at \( K = 0 \) is expected, because of the swaption quoting convention that the ATM quoted premium is equal to the sum of the ATM payer and receiver swaption. We now examine the case in which we want to calculate the bid and ask value of a swaption which has no bid and ask quoted values, hence we can not directly calibrate \( \lambda \) for it. In this case, we (linearly) interpolate \( \lambda \) using its two closest adjacent points. Indeed, let us consider the case in which we do not have bid and ask values corresponding to the strike \( K = +37.5bp \) and so we linearly interpolate \( \lambda(37.5) \) using \( \lambda(25) \) and \( \lambda(50) \). Then we get \( \lambda_{\text{int}}(37.5) = 0.03344238 \), instead of the “real” value \( \lambda(37.5) = 0.03321885 \) which yields the following results (in parentheses we cite the values for \( \lambda(37.5) \)):

\[
\begin{align*}
\text{bid}_{\text{int}}(K = 37.5) &= 15.89934852 \quad (15.91087951) \\
\text{ask}_{\text{int}}(K = 37.5) &= 19.69728469 \quad (19.68370587)
\end{align*}
\]

which have an absolute residual of:

\[
\begin{align*}
|\text{bid}_{\text{int}}(37.5) - \text{bid}(37.5)| &= 0.31951898 \quad (0.33104997) \\
|\text{ask}_{\text{int}}(37.5) - \text{ask}(37.5)| &= 0.29485413 \quad (0.28127532)
\end{align*}
\]

Moving further OTM, we interpolate \( \lambda(75) \) using \( \lambda(50) \) and \( \lambda(100) \). Then we get \( \lambda_{\text{int}}(75) = \)
0.02709932 instead of \( \lambda(75) = 0.02706849 \) leading to:

\[
\begin{align*}
\text{bid}_\text{int}(K = 75) &= 5.39703149 \quad (5.39774597) \\
\text{ask}_\text{int}(K = 75) &= 6.79484049 \quad (6.79397799)
\end{align*}
\]

and absolute residuals

\[
\begin{align*}
|\text{bid}_\text{int}(75) - \text{bid}(75)| &= 0.12504285 \quad (0.12575734) \\
|\text{ask}_\text{int}(75) - \text{ask}(75)| &= 0.10504882 \quad (0.10418632).
\end{align*}
\]

We therefore see that the results from the interpolated values of \( \lambda \) yield results that are quite close to the results obtained through the calibrated value of \( \lambda \).

### 4.3 Bid and Ask pricing

After presenting and analyzing the bid and ask simulated values for the 1y2y USD swaption under the Semi-Exact approach for all strikes in Figures 4.5 and 4.7 and for strikes corresponding to \(-100 \leq \text{bp} \leq +100\) in Figures 4.6 and 4.8, we now present some results of bid and ask pricing, separately for the Euler scheme and the Semi-Exact approach. In both cases, we have applied a 10% extra spread to prices corresponding to all strikes, plus an additional diminishing 5% spread to the OTM values.

#### 4.3.1 Semi-Exact approach

**Swap settled swaptions**

We demonstrate the results for the 2y2y and 5y5y USD swaptions, in which we have simulated 100,000 Monte Carlo values.

![Figure 4.10: 2y2y USD bid and ask residuals, Semi-Exact.](image)

where the large relative error for \( K = 200 \text{bp} \), is again due to the very small value of the swaption (market quoted premium=3.16, compared to 148.07 for the ATM). We also interpolated \( \lambda(K = -75) \) through \( \lambda(K = -50) \) and \( \lambda(K = -100) \):

\[
\begin{align*}
\lambda(-75) &= 0.03629910 \\
\lambda_{\text{int}}(-75) &= 0.03636964
\end{align*}
\]
yielding the following absolute residuals (in parentheses we have the respective values for the
original simulated value of $\lambda$):

$$|\text{bid}_{\text{int}}(-75) - \text{bid}(-75)| = 0.32662380 \quad (0.33125675)$$
$$|\text{ask}_{\text{int}}(-75) - \text{ask}(-75)| = 0.28411401 \quad (0.27860461).$$

For the 5y5y swaption we have the following residuals:

![Residuals for 5y5y USD bid and ask](image)

Figure 4.11: 5y5y USD bid and ask residuals, Semi-Exact.

Here we have interpolated $\lambda(12.5)$ through $\lambda(6.25)$ and $\lambda(25)$, getting:

$$\lambda(12.5) = 0.04604132$$
$$\lambda_{\text{int}}(12.5) = 0.04602239$$

and residuals:

$$|\text{bid}_{\text{int}}(12.5) - \text{bid}(12.5)| = 3.65846364 \quad (3.67029706)$$
$$|\text{ask}_{\text{int}}(12.5) - \text{ask}(12.5)| = 3.21843770 \quad (3.20488623).$$

We thus see that the interpolated value of the parameter $\lambda$ yields bid and ask values that are
a close approximation to the values we would get for a known quoted value. We now proceed
with the EUR swaptions.

**Cash settled swaptions**
We now present results for the 1y2y, 2y2y and 5y5y cash settled (EUR) swaptions, under the
same setting (number of simulations, augmented bid/ask spread) as before.
In this case we have omitted the $-150$ and $-200$ basis points, because the corresponding market quoted premium is 0. In the same manner as with the USD case, we see that the furthest OTM swaptions have large relative residuals but small (absolute) residuals, which is because they have a very small premium; indeed, the premium for $-100$, $-75$, $-50$, $+150$ and $+200$ bps is less than 1.00. We now interpolate $\lambda(25)$ through $\lambda(12.5)$ and $\lambda(50)$, with:

\[
\lambda(25) = 0.03610967
\]
\[
\lambda_{\text{int}}(25) = 0.03731506
\]

and residuals:

\[
|\text{bid}_{\text{int}}(25) - \text{bid}(25)| = 0.35559177\quad (0.32055127)
\]
\[
|\text{ask}_{\text{int}}(25) - \text{ask}(25)| = 0.20271702\quad (0.24816871).
\]

For the 2y2y EUR swaption we have:

\[
\lambda(-50) = 0.02921708
\]
\[
\lambda_{\text{int}}(-50) = 0.03043271
\]
and

\[ |\text{bid}_{\text{int}}(-50) - \text{bid}(-50)| = 0.19088570 \quad (0.22038018) \]
\[ |\text{ask}_{\text{int}}(-50) - \text{ask}(-50)| = 0.21996955 \quad (0.18463692). \]

Finally, for the 5y5y EUR swaption:

\[ |\text{bid}_{\text{int}}(150) - \text{bid}(150)| = 0.47101590 \quad (0.44810682) \]
\[ |\text{ask}_{\text{int}}(150) - \text{ask}(150)| = 0.32756097 \quad (0.35637495). \]

\[ \lambda(150) = 0.03458495 \]
\[ \lambda_{\text{int}}(150) = 0.03467072 \]

\[ |\text{bid}_{\text{int}}(150) - \text{bid}(150)| = 0.47101590 \quad (0.44810682) \]
\[ |\text{ask}_{\text{int}}(150) - \text{ask}(150)| = 0.32756097 \quad (0.35637495). \]

4.3.2 Euler Scheme

We now present the equivalent results, obtained by simulating values using the Euler scheme. We simulated 100,000 paths, discretizing each year of the expiry into 10,000 time steps for each path. In order to avoid repetition, instead of presenting the results of \( \lambda \) interpolation for each swaption as before, we will collectively present them together with those of the Semi-Exact approach, in the next sub-section.

Swap settled swaptions

For the 1y2y USD swaption, we have the following results.
Comparing with the respective results from the previous section, we see that the Semi-Exact approach yields generally more accurate results. Continuing, we present the residuals for the 2y2y and 5y5y swaptions.

Cash settled swaptions
We conclude this section by demonstrating results for the 1y2y, 2y2y and 5y5y EUR swaptions, simulated through the Euler scheme.
4.3.3 Comparison

USD swaptions

Here we have gathered the sum of the absolute residuals from both the Semi-Exact method and the Euler scheme. We have also added the “normalized” sum of absolute residuals, which we did by dividing them with the respective quoted ATM premium, in order to demonstrate the relevant magnitude of the residuals.
We see that for small expiries, the results yielded by the Semi-Exact approach are better that those of the Euler scheme. Of interest are also the computational times of these simulations. Indeed, the Euler scheme required around 22, 44 and 105 minutes for the 1y2y, 2y2y and 5y5y swaptions respectively, while for the Semi-Exact scheme the required time was around 25 minutes for each swaption. We thus conclude that the Semi-Exact approach, being both faster and more accurate, is strictly better than the Euler scheme for small expiries, while still being preferable for larger expiries, because of its speed and similar results.

We now present a table containing the residuals of the bid and ask prices for which we interpolated $\lambda$, for both methods.

### Table 4.7: Method residuals comparison USD

<table>
<thead>
<tr>
<th>Method</th>
<th>1y2y</th>
<th>2y2y</th>
<th>5y5y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-Exact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sum</td>
<td>bid_{res}</td>
<td>$</td>
<td>2.32887203</td>
</tr>
<tr>
<td>$\sum</td>
<td>ask_{res}</td>
<td>$</td>
<td>1.98861305</td>
</tr>
<tr>
<td>norm. bid res.</td>
<td>0.02667971</td>
<td>0.07091421</td>
<td>0.15194299</td>
</tr>
<tr>
<td>norm. ask res.</td>
<td>0.02278168</td>
<td>0.06093139</td>
<td>0.13227729</td>
</tr>
<tr>
<td>( K )</td>
<td>( \lambda_{cal} )</td>
<td>( \lambda_{int} )</td>
<td>(</td>
</tr>
<tr>
<td>1y2y</td>
<td>37.5</td>
<td>0.0332188</td>
<td>0.03344238</td>
</tr>
<tr>
<td>2y2y</td>
<td>37.5</td>
<td>-75</td>
<td>0.03629910</td>
</tr>
<tr>
<td>5y5y</td>
<td>12.5</td>
<td>0.04604132</td>
<td>0.04602239</td>
</tr>
<tr>
<td>Euler</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sum</td>
<td>bid_{res}</td>
<td>$</td>
<td>3.58735899</td>
</tr>
<tr>
<td>$\sum</td>
<td>ask_{res}</td>
<td>$</td>
<td>3.08995857</td>
</tr>
<tr>
<td>norm. bid res.</td>
<td>0.04109702</td>
<td>0.08302220</td>
<td>0.10852172</td>
</tr>
<tr>
<td>norm. ask res.</td>
<td>0.03539877</td>
<td>0.07151001</td>
<td>0.09577320</td>
</tr>
</tbody>
</table>

We thus see that the bid/ask results given by the interpolated $\lambda$ values yield acceptable results in both methods.

### Table 4.8: Residuals for swaptions with interpolated $\lambda$, USD

We now proceed with presenting the equivalent results for the simulated EUR swaptions.
We see that the results are in accordance with the previous, as the Semi-Exact approach is more accurate for the 1y2y and 2y2y swaptions, while slightly less accurate for the 5y5y swaption. The computational time required by the Euler scheme was 44, 117 and 220 minutes for the 1y2y, 2y2y and 5y5y swaptions respectively, while for the Semi-Exact approach it was around 45 minutes for each. As before, we now present a table containing the results of the interpolated \( \lambda \) swaptions.

Table 4.9: Method residuals comparison EUR

|     | \( \sum |\text{bid}_{res}| \) | \( \sum |\text{ask}_{res}| \) | norm. bid res. | norm. ask res. |
|-----|------------------------------|------------------------------|----------------|----------------|
| 1y2y| 2.41542162                   | 1.84898861                   | 0.05905676     | 0.04520755     |
| 2y2y| 3.10198635                   | 2.59436321                   | 0.03581970     | 0.02995800     |
| 5y5y| 30.04753876                  | 25.62646334                  | 0.05669347     | 0.04835182     |

Euler

|     | \( \sum |\text{bid}_{res}| \) | \( \sum |\text{ask}_{res}| \) | norm. bid res. | norm. ask res. |
|-----|------------------------------|------------------------------|----------------|----------------|
| 1y2y| 3.88019520                   | 2.90193309                   | 0.09487030     | 0.07095191     |
| 2y2y| 4.29575004                   | 3.53692335                   | 0.04960450     | 0.04084207     |
| 5y5y| 26.14355771                  | 22.24112175                  | 0.04932747     | 0.04196438     |

Table 4.10: Residuals for swaptions with interpolated \( \lambda \), EUR

Demonstrating again the accuracy of the bid/ask values calculated through interpolated \( \lambda \) values.
Discussion

The purpose of this thesis was to provide a pricing procedure for the bid and ask values of EUR and USD swaptions, through the Conic Finance theory. To the best of our knowledge, this was the first such attempt. To do so, we had to calibrate the SABR parameters by curve fitting the market implied volatilities and then, using them we simulated multiple values of the underlying which were used for Conic Monte Carlo. To this end, we first applied two standard techniques used in finance, i.e. using deterministic square minimization for the implied volatility curve fitting and simulating the underlying through the Euler discretization scheme. Then, we applied two less frequently used techniques: the Differential Evolution algorithm for the parameter calibration and the Semi-Exact approach for the underlying simulation. We concluded that both Differential Evolution and the Semi-Exact approach were better suited for our situation, the first being independent of potentially ill-chosen initial values and the second being less biased and of far better computational speed.

Given our bid and ask results, we are led to the conclusion that the Conic Finance theory offers a solid pricing framework, which is valid in both liquid and illiquid markets.

Concerning future research, we have the following suggestions. Firstly, the products that we considered were liquid enough to have an almost zero bid-ask spread. We circumvented this by adding an artificial extra spread, but it would have been more natural to apply our study to less liquid products, something that we suggest for further studies. Continuing, even though we demonstrated that the $MINMAXVAR$ distortion function yielded the best results in our case, there is still room for improvement, namely considering the $MINMAXVAR2$ distortion function, as seen in [32], which is dependent on two parameters $\lambda, \gamma$ instead of just one. This function would clearly require more computational time for calibration purposes, but offers greater versatility and as such should yield better results. Continuing, an open, to the best of our knowledge, problem in finance is the issue of the cash-settled (EUR) annuity. As discussed in Section 2.2, the current standard market formula is mathematically wrong, because of the use of the Multi Curve Framework. One suggestion would be to consider the cash-settled annuity as a stochastic process of its own and try to model it separately. Further, the poorer results of the Semi-Exact approach for larger expiries are expected and discussed in Cai et al. [7], who offer the solution of splitting the expiry into smaller time intervals and applying the Semi-Exact piece-wise. Finally, instead of the classical Euler discretization scheme, we would suggest the use of a modified Euler scheme that guarantees convergence for models with non-Lipschitz parameters, as seen e.g. in [4] and [6].
Popular summary

If the market is liquid and complete, then the classical financial framework offers an unambiguous pricing method for all financial products, through the use of the risk-neutral measure. We thus have the rule of the “one price”. However, in illiquid markets, the price of a product depends on the market direction, i.e. whether it is being bought or sold. This leads to the difference between the buying and the selling price, called the “bid-ask spread”, to become non-negligible. Conic Finance offers a framework in which easy to compute formulas are derived for the direct calculation of both the bid and the ask prices.

The SABR (“Stochastic Alpha Beta Rho”) model is a stochastic volatility model used to calculate forward values of a derivative’s underlying, for example the price of a stock or an interest rate. In the current post-crisis environment, interest rates are very low and sometimes even negative. To deal with this situation, the Displaced SABR extension of SABR was developed.

In this thesis, we follow the Conic Finance approach and use the Displaced SABR model in order to directly price the bid and ask prices of swaptions, i.e. derivatives in which the underlying is a swap. To do so, we first needed to calibrate the SABR parameters and then simulate multiple values of the underlying. In both cases, we used a well-known (deterministic least squares and Euler discretization) and a less known method (Differential Evolution Algorithm, “Semi-Exact” approach). We concluded that for the parameter calibration the Differential Evolution algorithm yielded better results and for the underlying simulation the Semi-Exact approach was more accurate in most cases and computationally less expensive in general.
Bibliography


