MSc Thesis Quantitative Risk Management

Simulation and Arbitrage-Free Estimation of the FX Implied Volatility Surface

Karim Moussa
k.moussa@student.vu.nl
2110482
Vrije Universiteit Amsterdam

Supervisors:
Drs. Arnout Dieterman, *ABN Amro Bank*
Dr. Svetlana Borovkova, *Vrije Universiteit Amsterdam*

September 30, 2016

Abstract

The contribution of this thesis is twofold. First, we propose a method to simulate realistic FX option price data in the form of volatility surfaces, consistent with conventions of the FX market. Second, using the simulated prices, we compare several arbitrage-free methods to construct the volatility surface, namely the SABR model, the lognormal mixture model, constrained moving least-squares and arbitrage-free SVI, to determine which method is optimal for FX option price data. In addition, all methods are applied to market prices of EUR/USD and USD/JPY options. We find that the arbitrage-free SVI model is most suitable to estimate the FX volatility surface.

Keywords: FX options simulation; Implied volatility; Volatility surface; Arbitrage;

Acknowledgements

Several people have helped me during the writing of this thesis, and I would like to express my gratitude towards them.
First of all, I am grateful to my supervisors Svetlana Borovkova and Arnout Dieterman, as I have greatly benefited from their advice and practical knowledge.
I would like to thank the members of the department of quantitative analysis at ABN Amro, Geert Ceuppens, Raoul Pietersz, Arnout Dieterman (once more), Chris Klaver, Panos Nikolopoulos, Vincent Leijdekker, Jelle Kok, André Roukema and Nancy Appels, for giving me the opportunity to learn from them, and most importantly, for making my time there enjoyable.
I am grateful for the advice of Sergiy Ladokhin from ABN Amro Clearing Bank, which has aided me in writing this thesis.
I thank Evgeny Garmaev, as I have learned a lot from several discussions on volatility surfaces we had.
I am thankful to Bas Henke for several helpful comments on improving this thesis.
Lastly, I am most thankful to my family and all the people close to me for their continuous support.

Amsterdam, September 30, 2016

Contents

1 Introduction

This thesis is concerned with two closely related objectives. The first objective is to propose a method to simulate realistic foreign exchange (FX) option price data. The second objective is to determine which method is *optimal* for constructing the arbitrage-free volatility surface for FX options by means of a comparison study. The simulation of FX option price data appears not to have been addressed in the literature, although realistic simulation of option prices is important, as it allows for efficient testing and comparison of option pricing methods. The proposed simulation method first computes arbitrage-free option prices by using a model for the risk-neutral dynamics of the underlying, which are then transformed into options traded in the FX market: ATM implied volatilities, risk reversals and strangles. To simulate prices, the generated FX options are perturbed with random noise to mimic market noise caused by supply and demand. We will use the proposed simulation method to compare several methods that construct the volatility surface in a simulation study, after which all methods are applied to market prices of EUR/USD and USD/JPY options.
As the Black-Scholes formula provides an invertible relationship between implied volatilities and European option prices, construction of the volatility surface is equivalent to the problem of determining European option prices for arbitrary strike prices and time to expiries, given that the market provides only a limited number of prices. Apart from determining the values of European options, the volatility surface is used for the calculation of Greeks (thus as a hedging instrument), for the generation of stress scenarios and for computation of the local volatility, a method often used to price exotic options. As such, the volatility surface is an element of crucial importance in every pricing library.
The word *optimal* refers to the quality of the volatility surface, which can be measured by the following three objectives:

1. The surface must be arbitrage-free;
2. The surface needs to fit the market prices well; and
3. The method constructing the surface must allow for fast calibration to market prices.

The first objective stems directly from the axiom of no-arbitrage, which is the most important, and perhaps the only rule in pricing options. Therefore, the focus of this thesis is exclusively on methods that construct the volatility surface while trying to ensure the exemption of arbitrage. The other two objectives arise from practical requirements, although it can be reasoned that they too are implied by the axiom of no-arbitrage. Objectives two and three can be quantified easily, which allows us to compare the different methods.
The rest of this thesis will proceed as follows. In the remainder of this Section, the interested reader can find a short review of a crucial part of option pricing history related to the volatility surface and some examples of methods to construct the volatility surface. In addition, we provide an overview of FX market conventions. In Section 2, we discuss several necessary and sufficient no-arbitrage conditions for the volatility surface in terms of shape constraints on the call price surface; we will refer to these constraints in Section

3 where several methods aimed at constructing the arbitrage-free volatility surface are discussed. In Section 4 we propose a method to simulate FX option price data in the form of volatility surfaces, which will be used in a simulation study in Section 5 for method comparison. In Section 6 all methods are applied to real EUR/USD and USD/JPY European option price data. Section 7 concludes.

1.1 Option pricing and the axiom of no-arbitrage

The field of option pricing has its roots in Paris of 1900, where Louis Bachelier was the first person to apply advanced mathematics to the pricing of options in his PhD thesis "*Théorie de la spéculation*".[1] In his thesis, Bachelier (1900) derived an option pricing formula in which the underlying follows an arithmetic Brownian motion:

$$dS_t = \mu dt + \sigma dW_t, \tag{1}$$

where S_t denotes the current value of the underlying, μ denotes the expected return relative to the length of the process, σ denotes the corresponding volatility and W_t denotes a standard one-dimensional Brownian motion. To derive his option pricing formula, Bachelier made the assumption that the expected profit of an investor is zero, and emphasized its importance by calling it the *fundamental principle*:

"*L'ésperance mathématique du spéculateur est nulle.*"
"*The mathematical expectation of a speculator is zero.*"

Bachelier justified this assumption by means of the following equilibrium argument:

"*It seems that the market, the aggregate of speculators can believe in neither a market rise nor a market fall, since for each quoted price, there are as many buyers as sellers.*"

The statement that as the number of sellers for any product equals the number of buyers, their expectations must offset each other, can be seen as an early version of the efficient market hypothesis. The work of Bachelier on option pricing appears to have been forgotten after it was published, until 1955 when it brought under the attention of Paul Samuelson. The work of Bachelier inspired Samuelson, who would later win the Nobel Prize for Economics, as well as many others to further develop option pricing theory.
In 1908, shortly after Bachelier's work on option pricing, the German mathematician Vinzenz Bronzin published a book named "*Theorie der Prämiengeschäfte*" (Theory of Premium Contracts), in which he derived several important results in option pricing theory, among which are an option pricing formula similar to Bachelier's formula, a formulation of put-call parity and a formula for the density implied by the prices of options. Contrary to Bachelier, Bronzin had derived his option pricing formula based on arbitrage arguments instead of an equilibrium argument. Bronzin's work, like Bachelier's, appears to have been forgotten after publication, only to be rediscovered recently (see Zimmermann and Hafner (2007)); despite the important results it contained, Bronzin's work has

[1] Bachelier is also credited to be the first person to model the Brownian motion in the same thesis, hereby predating the famous Einstein (1905) by five years, who published on the topic in his PhD thesis "*The theory of the Brownian movement*". Notably, Bachelier's supervisor was none other than Henri Poincaré, one of the greatest mathematicians of his time.

had little influence on the option pricing developments that followed.
After Bachelier and Bronzin, several other option pricing formulas (see Sprenkle (1961), Boness (1964) and Chen (1970) among others) had been derived based on an equilibrium argument. The formulas were similar to the later derived Black-Scholes formula, although with the drawback of having at least one parameter for which calibration is unclear.
The papers of Black and Scholes (1973) and Merton (1973) caused a paradigm shift in option pricing, going from equilibrium based pricing to arbitrage based pricing. In accordance with no-arbitrage pricing, Black and Scholes (1973) derived a partial differential equation for the pricing of options when the returns of the underlying follow a normal process:

$$\frac{dS_t}{S_t} = d\ln S_t = \mu dt + \sigma dW_t. \tag{2}$$

In the case of European options, a closed-form option pricing formula was derived corresponding to process (2):

$$\begin{aligned} C^{BS}(K,\tau;\sigma) &= S_t N(d_1) - e^{-r\tau} K N(d_2), \\ d_1 &= \frac{\ln(S_t/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \\ d_2 &= d_1 - \sigma\sqrt{\tau}, \end{aligned} \tag{3}$$

where $\tau \equiv T - t$ denotes the time to expiry in years, r denotes the constant risk-free short rate and K denotes the strike of the option.
Black and Scholes had originally derived their partial differential equation by ensuring consistency with the Capital Asset Pricing Model, after which Merton provided them the following derivation that is based only on the exemption of arbitrage:

1. Consider the price of a derivative $V_t = V_t(S_t, t)$ that depends on the value of the underlying S_t driven by (2). Using Itô's formula, we find

$$dV_t = \left[\frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}\right] dt + \sigma S_t \frac{\partial V}{\partial S_t} dW_t.$$

2. We can create a deterministic portfolio Π_t by going long one unit of the derivative V_t and short $\partial V/\partial S_t$ units of the underlying (the weight of the underlying must be adjusted continuously for the portfolio to become entirely riskless):

$$\begin{aligned} \Pi_t &= V_t - \frac{\partial V}{\partial S_t} S_t, \\ d\Pi_t &= \left[\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2}\right] dt. \end{aligned}$$

3. As the portfolio is riskless, its return should be equal to the risk-free rate r:

$$\frac{d\Pi_t}{\Pi_t} = r dt, \text{ or equivalently: } d\Pi_t = r\Pi_t dt,$$

from which we find the Black-Scholes partial differential equation by rearranging:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2} + r S_t \frac{\partial V}{\partial S_t} - r V_t = 0.$$

"*Scholes' and Black's insight, which was a critical insight, was that hedging an option removes its systematic risk. At first I said, Thats impossible. But I looked into it. I went back to them and said, You guys are absolutely right, but for the wrong reason. The hedge removes all risk. There are two derivations of the formula: theirs and mine. They were nice enough to include mine in their paper.*".

- Merton (2014)

The insight that although derivatives and the underlying assets they depend on are risky it is possible combine them into a *riskless* portfolio by hedging, provided a method to price these derivatives and noted the start of the *risk-neutral valuation* of derivatives. In addition to the hedging argument, Merton (1973) derived several theoretical bounds for the prices of call and put options by arbitrage arguments in his landmark paper "*The theory of rational option pricing*". In effect, Black, Scholes and Merton had shown that rational determination of the price of any derivative should be done in accordance with the axiom of no-arbitrage:

Axiom 1. (No-arbitrage pricing)
The rationally determined price of a derivative is such that it admits no arbitrage opportunities.

The year 1973 also noted the establishment of the first options exchange when the Chicago Board Options Exchange was founded. Option pricing reportedly took an enormous rise in popularity, and in retrospect, the succes of the Black-Scholes model can be attributed to two factors:

1. The hedging argument, leading to a paradigm shift in option pricing from equilibrium-based pricing to no-arbitrage pricing; and

2. The fact that only the volatility parameter σ cannot be observed directly in the market, but can be estimated easily from historical data, made it easy to price derivatives (which was not the case for the predecessors of the Black-Scholes formula, apart from Bachelier's formula which did not ensure positivity of the underlying).

The theory of no-arbitrage pricing was developed further in Cox and Ross (1976) and Harrison and Kreps (1979) among others, and it is a result from the latter paper that we now bring under attention. Therein, Harrison and Kreps had proven that the exemption of arbitrage coincides with the existence of a probability measure, called an *equivalent martingale measure* or a *risk-neutral measure*, equivalent to the real-world probability measure, but under which the underlying has drift equal to the risk-free rate. A remarkable result, because it provides a link between the concepts of probability and arbitrage, which at first sight appear to be unrelated. The result is perceived to be so important that it was later named the *First Fundamental Theorem of Asset Pricing.*

Theorem 1. (First Fundamental Theorem of Asset Pricing)
The market defined by an underlying $(S_t)_{t\geq 0}$ is arbitrage-free if and only if there exists a probability measure $\mathbb{Q}$ equivalent to the real-world probability measure $\mathbb{P}$, under which the discounted price process $(\tilde{S}_t)_{t\geq 0} \equiv (e^{-\int_0^t r(s)ds} S_t)_{t\geq 0}$ is a martingale. The value of any derivative is then given by its discounted expected payoff under the measure $\mathbb{Q}$.

1.2 Implied volatility

In the years following the introduction of the Black-Scholes model, the volatility parameter σ was estimated using historical data of the underlying. Latané and Rendleman (1976) introduced the concept of implied volatility, the volatility σ_{IV} such that the price computed using the Black-Scholes formula (3) matches the market price C^{MKT}, i.e.

$$C^{BS}(K, \tau; \sigma_{IV}) = C^{MKT}(K, \tau).$$

In their article, Latané and Rendleman showed that a weighted average of implied volatilities is a better predictor of future volatility than the volatility estimated based on historical data. The initial use of implied volatility was thus to improve estimation of the volatility parameter, and not to question the validity of the Black-Scholes model.
After the Black-Scholes model was introduced, the lognormal dynamics specified by (2) were generally accepted as a reasonable approximation to the true data generating process of the underlying in most options. This consensus appears to have changed after the stock market crash in 1987, in which asset prices made sudden severe downward movements inconsistent with model (2). Skewed patterns of implied volatility started to show in the equity options market (see e.g. Rubinstein (1994)), and soon patterns of implied volatility started to occur in other markets too.
The concept of implied volatility allows for visualization of the model misspecification of (2); by plotting the implied volatilities against strike prices for a specific maturity we can often observe a 'smile' or 'smirk'-like pattern, which implies a violation of the assumptions of the Black-Scholes model, in which the volatility is assumed to be a constant. Figure 1 shows such a pattern, along with a possible Black-Scholes volatility. We can interpret these patterns and make relative statements about the implied risk-neutral density of the underlying's returns, by noting that a constant implied volatility corresponds to a normal distribution. For example, the pattern in Figure 1 shows that the prices of out of the money put options are priced higher than would be expected using the Black-Scholes model, whereas out of the money call options are priced slightly lower; the risk-neutral distribution of the underlying is left-skewed compared to a normal distribution.
A pattern can often also be seen with respect to the time expiry dimension, although usually less pronounced. Figure 2 shows a plot of the implied volatilities against a range of strike prices and time to expiries. The function $\sigma(K, \tau) : \mathbb{R}_+^2 \to \mathbb{R}_+$ that maps *any* strike and time to expiry to an implied volatility is called the *volatility surface*.
Despite its inconsistencies with observed market prices, the Black-Scholes option pricing formula is still widely used, as most exchanges quote the prices of options in terms of implied volatilities. Lee (2005) notes that to understand why the inversion of option prices in an *incorrect* formula should deserve the large amount of attention it is given in current research, it is helpful to regard the Black-Scholes implied volatility as a *language* in which to express an option price.

"*The language of implied volatility is, moreover, a useful alternative to raw prices. It gives a metric by which option prices can be compared across different strikes, maturities, and underlyings, and by which market prices can be compared to assessments of fair value. It is a standard in industry, to the extent that traders quote option prices in "vol" points, and exchanges update implied volatility indices in real time.*"

- Lee (2005)

1.3 Beyond Black-Scholes

The occurrence of volatility smiles indicates clear misspecification of the Black-Scholes model and has called for an option pricing model that does a better job at fitting the observed market prices. Many alternatives to the Black-Scholes model have been proposed, and most of them can be categorized into two main strategies:

1. Specify a process with more realistic risk-neutral dynamics for the underlying than (2); and
2. Directly model the volatility surface.

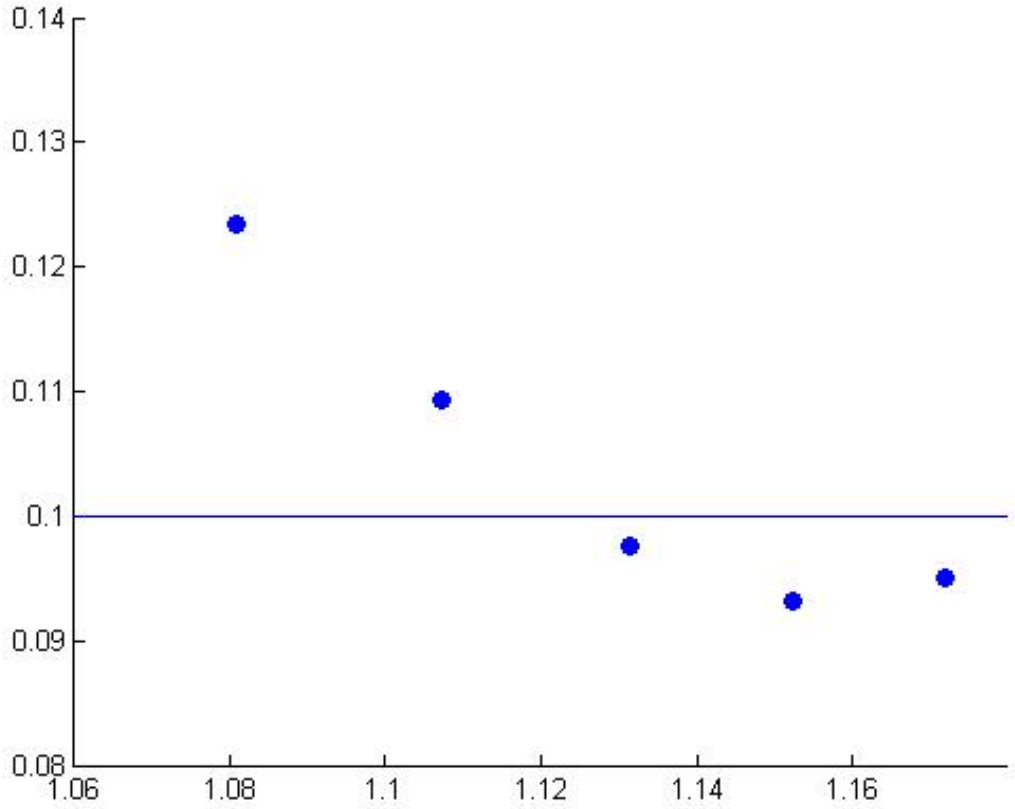

Figure 1: Implied volatilities for European options on the EUR/USD exchange rate (S_t = 1.1309) with a time to expiry of one year as of 15-06-2016. The blue line represents a possible Black-Scholes implied volatility estimate over all options.

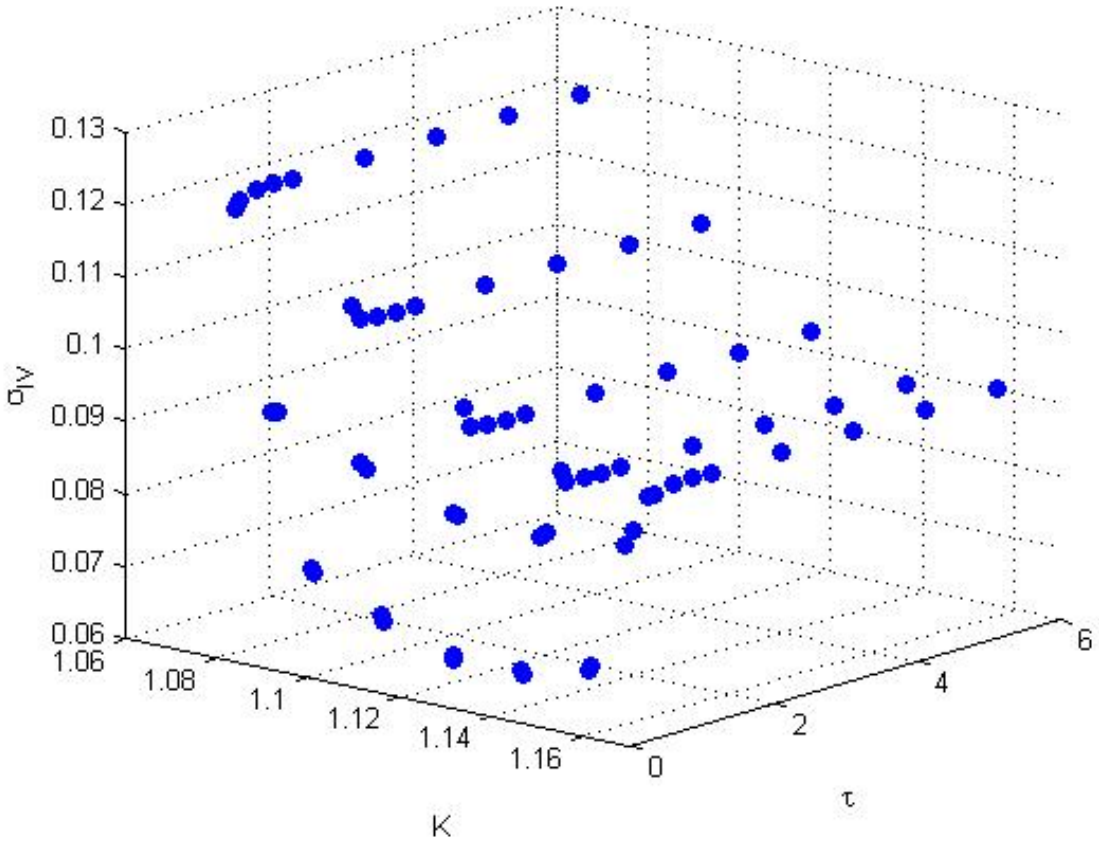

Figure 2: Implied volatilities for European options on the EUR/USD exchange rate (S_t = 1.1309) with time to expiries ranging from one day until five years as of 15-06-2016.

1.3.1 Dynamic specification of the underlying

As the occurrence of volatility smiles implies that the 'true' risk-neutral distribution for the returns of the underlying is not normal, one way of matching prices more closely to the market's is to assume dynamics for the underlying different from (2). It is important here to note, that in such a strategy, the specification for the underlying must be in risk-neutral form, i.e. the discounted underlying must be a martingale. In this way, we can ensure the exemption of arbitrage by the First Fundamental Theorem of Asset Pricing.

Example 1. (Stochastic volatility with jumps; SVJ)
The Bates (1996) model extends the Black-Scholes model by including both stochastic volatility as well as jumps into the specification for the dynamics of the underlying, and was devised to improve the pricing of currency options. The model is a generalization of the stochastic volatility model of Heston (1993) and the jump diffusion model of Merton (1976). The dynamics of the underlying S and the variance v under the risk-neutral measure are characterized by the following equations:

$$
\begin{aligned}
dS_t &= (r^d - r^f - \lambda\mu_J)S_t dt + \sqrt{v_t}S_t dW_t + J_t S_t dN_t, \\
dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}d\tilde{W}_t, \\
\mathbb{C}\text{orr}[W, \tilde{W}] &= \rho, \\
\ln(1 + J_t) &= N\left(\ln(1 + \mu_J) - \frac{\sigma_J^2}{2}, \sigma_J^2\right),
\end{aligned} \tag{4}
$$

where in the SDE for the underlying, r^d denotes the domestic risk-free short rate and r^f denotes its foreign counterpart, $\sqrt{v_t}$ is the stochastic volatility, J_t is a log-normally distributed jump-size with mean μ_J and standard deviation σ_J and N_t is a Poisson process with intensity λ, yielding probability of a jump in continuous time of λdt. In the SDE of the volatility, θ represents the long-run variance, κ denotes the speed of mean reversion, σ denotes the volatility of the volatility and $\tilde{W}_t$ is another standard one-dimensional Brownian motion.
Setting μ_J and σ_J equal to zero yields Heston's stochastic volatility model, and setting κ and σ equal to zero yields Merton's jump diffusion model. The model allows for the generation of implied volatility skews by a nonzero average jump size or by a nonzero correlation between the underlying and the volatility process. The jump component is complementary to the stochastic volatility component, in the sense that jumps allow for a pronounced smile effect for short time to expiries, whereas stochastic volatility allows for modeling smile effects for longer time to expiries. The explanation is that jumps could occur at any time, hence its inclusion is also noticeable for short to expiries, where the effect of jumps with (close to) zero means average out in the long-run. On the other hand, stochastic volatility implies a positive relation between the option's time to expiry and the likelihood of a large movement of the underlying between the start of the option and expiration.

In addition to ensuring exemption of arbitrage, assuming more realistic dynamics for the underlying than (2) has the advantage of parameter estimates that allow for an easy interpretation. The major disadvantage of these models is its calibration speed. Kilin (2006) finds in a comparison study that the fastest method of computing option prices for models of the underlying with known characteristic function is using direct integration of a complicated function (see Section 5.1), which becomes computationally onerous when calibrating such models to market prices.
To make matters worse, calibration of models specifying the dynamics of the underlying do not always result in convex optimization problems (see Figure 3), which complicates model fitting even further. Although local optimization algorithms such as Newton's method are fast, non-convexity of the criterion function causes these methods to be sensitive to the starting values. A frequently used class of global optimization methods designed to tackle non-convex optimization problems is the class of evolutionary based methods (e.g. the Differential Evolution method of Storn and Price (1997)). Here, one starts with a set of points in the parameter space and draws new points using generated random numbers for several iterations; if the newly drawn points improve upon original points, the original points are replaced. The set of points will converge to an optimum, of which it is hoped to be the global optimum. Such methods are able to deal with local optima, and are often combined with local optimization methods to converge faster to a solution. However, global optimization methods require many evaluations of the criterion

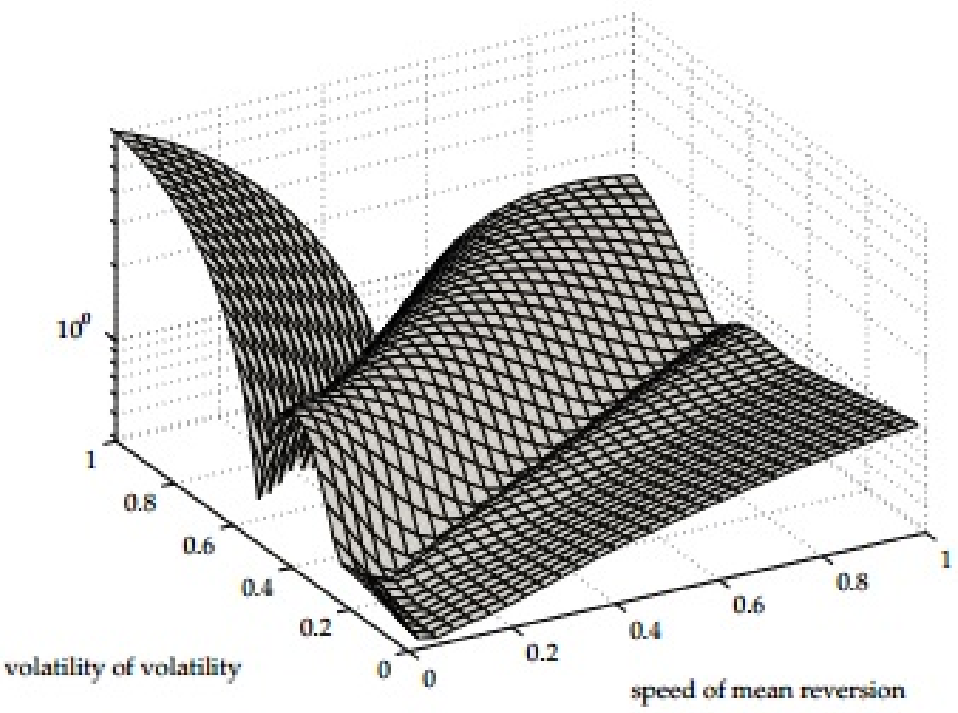

Figure 3: A search space for the Heston model when using a mean absolute percentage error criterion function, from Gilli and Schumann (2011). The fitting of the Heston model is seen to be a non-convex optimization problem, which complicates calibration.

function, which in turn requires the computation of multiple model prices and hence numerical evaluation of integrals. One is thus left with the trade-off being either sensitive to local optima and thus likely not getting the most out of the model, or suffering from even longer calibration time. In the hope to gain speed to a level acceptable in practice, we will consider two models for the dynamic specification of the underlying that do not require the evaluation of integrals to compute option prices: the SABR model introduced by Hagan, Kumar, Lesniewski, and Woodward (2002) and the lognormal-mixture model of Brigo and Mercurio (2002b).

1.3.2 Modeling the volatility surface directly

Due to the large computational demands of most models that assume more realistic dynamics for the underlying, modeling the volatility surface directly is often seen as an attractive alternative. An example of a method that directly models the volatility surface is the *stochastic volatility inspired* (SVI) model introduced by Gatheral (2004).

Example 2. (Stochastic Volatility Inspired; SVI).
The original SVI representation defines a parametric shape for the (total) variance for a fixed time to expiry given by

$$w(x; a, b, c, \rho, m) = a + b(\rho(x - m) + \sqrt{(x - m)^2 + c^2}), \tag{5}$$

where $w \equiv \sigma^2\tau$ denotes the total implied variance, $x \equiv \ln(K/F_{t,T})$ with $F_{t,T}$ the value of the forward contract that expires at time T, and a, b, c, m and ρ are parameters that determine the shape of the smile. The main advantages of the SVI model are:

1. Its flexibility, which makes it possible to fit many possible implied volatility shapes;
2. The tails of the formula are consistent with the moment condition of Lee (2004), hence the implied volatility does not grow faster than $\sqrt{|x|}$ as $|x| \to \infty$; and
3. The large-maturity limit of the Heston implied volatility smile agrees algebraically with SVI, which implies SVI is not just an arbitrary specification.

Disadvantages of the SVI model are that the corresponding optimization problem is not ensured to be convex, and that it requires a separate time to expiry inter- and extrapolation method to construct the entire volatility surface.

A general disadvantage of modeling the surface directly is that ensuring exemption of arbitrage is a non-trivial matter. For this reason, recent methods that model the surface directly aim to exclude *static* arbitrage, a simplified version of arbitrage (see Section 2).

1.4 Foreign exchange market conventions

The foreign exchange (FX) market has several important conventions of which we provide an overview here. We will follow Clark (2011) closely.
First, the time t spot rate S_t for a currency pair quoted as CCY1/CCY2, is the number of units of CCY2 (the domestic currency) required to buy one unit of CCY1 (the foreign currency). For example, 'EUR/USD' refers to the number of dollars per euro, contrary to what the slash between the currencies might suggest. Furthermore, as was shown in Example 1, the FX market has to deal with two 'riskless' interest rates instead of one, and the forward is defined as[2]

$$F_{t,T} = S_t e^{(r^d - r^f)\tau}.$$

The existence of multiple interest rates also results in an adjusted version of the Black-Scholes formula, as derived by Garman and Kohlhagen (1983):

$$C(K,\tau) = e^{-r^f\tau} S_t N(d_1) - e^{-r^d} K N(d_2), \tag{6}$$

where $N(\cdot)$ denotes the cumulative standard normal distribution and d_1 and d_2 are given by

$$\begin{aligned} d_1 &= \frac{\ln(S_t/K) + (r^d - r^f + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \\ d_2 &= d_1 - \sigma\sqrt{\tau}. \end{aligned}$$

Formula (6) is equivalent to the Black-Scholes formula for European option prices with continuous time dividends correction when we interpret the foreign risk-free short rate as the continuous dividend yield. Different from the original Black-Scholes model is, that the FX market knows multiple versions of delta, which are currency pair specific. This is due to the possibility of computing the sensitivity of the option price to either the

[2]Constant interest rates are assumed in this Section only, for ease of notation and compatibility with the Black-Scholes formula.

spot or the forward value, and the possibility to express this sensitivity in two different currencies. The different delta types are 'pips spot', 'percentage spot', 'pips forward', 'percentage forward' and 'simple delta', and are defined as

$$\Delta_{S;pips} = \omega e^{-r^f \tau} N(\omega d_1), \qquad \Delta_{S;\%} = \omega e^{-r^d \tau} \frac{K}{S_t} N(\omega d_2),$$
$$\Delta_{F;pips} = \omega N(\omega d_1), \qquad \Delta_{F;\%} = \omega \frac{K}{F_{t,T}} N(\omega d_2),$$
$$\Delta_{simple} = \omega N(\omega d),$$

where $\omega \in -1, 1$ depending on the option being a put or a call respectively, and $d = \ln(F_{t,T}/K)/\sigma\sqrt{\tau}$. The word 'pips' refers to expressing the sensitivity of the option price to the change in spot, both in in CCY2 per CCY1 units. The word 'percentage' refers to expressing the sensitivity of the value of the option in *% foreign terms* to the change in spot in *% foreign terms*:

$$\Delta_{S;\%} \equiv \lim_{\Delta S_t \to 0} \frac{\Delta V_t / S_t}{\Delta S_t / S_t}, \qquad \Delta_{F;\%} \equiv \lim_{\Delta F_{t,T} \to 0} \frac{\Delta V_t / F_{t,T}}{\Delta F_{t,T} / F_{t,T}},$$

where V_t denotes the value of the option expressed in CCY2 per CCY1 units, and the Δ symbols in the subscripts refer to 'the change in' the spot or forward value. Lastly, simple delta is not a delta in the sense of sensitivity with respect to changes in the spot or the forward, but it is used as a simplified measure of moneyness lying somewhere in between the two forward deltas $\Delta_{F;pips}$ and $\Delta_{F;\%}$.
With respect to the quoting of prices, it is characteristic of the FX market that:

- The market quotes implied volatilities instead of prices;
- Prices are quoted per delta instead of per strike; and
- The market quotes the ATM volatility directly, and it quotes the 25 and 10 percent delta put and call volatilities indirectly through the prices of risk reversal (RR) and strangle (STR) options.

The strike corresponding to the ATM delta is computed by either setting it equal to the forward (the ATMF convention), or, setting it such that that the straddle (a combination of a long put and long call with the same strike) is *delta-neutral*, i.e.

$$\Delta_{Call}(K) + \Delta_{Put}(K) = 0.$$

For all conventions, we have closed form expressions for the value of the strike (see Clark (2011)):

$$K_{ATMF} = F_{t,T},$$
$$K_{DNS;pips} = F_{t,T} \exp\left(\frac{1}{2}\sigma^2\tau\right),$$
$$K_{DNS;\%} = F_{t,T} \exp\left(-\frac{1}{2}\sigma^2\tau\right).$$

We can see from the formulas for the different types of delta that the quoted options in the order $-10\Delta, -25\Delta, ATM, 25\Delta$ and 10Δ correspond to increasing strike values, which means that in addition to the ATM options, the market quotes OTM put and call options. As all deltas are a function of moneyness, quotation by delta simplifies making statements about the risk-neutral distribution of the returns of the underlying. This is especially true for both pips deltas as well as simple delta, which are symmetric in moneyness.
The α-percent risk reversal is defined as the difference in volatility of the α-percent delta call and put:

$$RR_\alpha \equiv \sigma(\Delta_{+\alpha}) - \sigma(\Delta_{-\alpha}), \tag{7}$$

where the negative percentage $-\alpha$ refers to the put option. The risk reversal measures the skewness of the implied risk-neutral density of the returns of the underlying relative to the normal distribution. The α-percent smile strangle is given by the average of the corresponding put and call options minus the ATM volatility:

$$STR_\alpha \equiv \frac{\sigma(\Delta_{+\alpha}) + \sigma(\Delta_{-\alpha})}{2} - \sigma_{ATM}. \tag{8}$$

The smile strangle is linked to the kurtosis of the implied risk-neutral distribution relative to the normal distribution, as can be seen in Figure 4. The risk reversals and strangles can be converted to implied volatilities by the following formula:

$$\sigma_{\Delta_{\pm\alpha}} = \sigma_{ATM} \pm \frac{1}{2} RR_\alpha + STR_\alpha. \tag{9}$$

Lastly, we define several basic option trading strategies, which are needed for Section 2. A *call spread* refers to the combination of going long a call option on an underlying with strike price K_1, and going short a call option on the same underlying with a different strike price K_2. If $K_1 < K_2$, then the portfolio aims to profit from a rise in the underlying up to a certain upper bound while reducing costs through the short call; in this case, the portfolio is referred to as a *bull call spread*. In the rest of this thesis we will use the shortened phrase 'call spread' to refer to a 'bull call spread'.
A *butterfly spread* is the strategy of buying a call option with strike K_1, going short two calls at a higher strike K_2 and going long another call option with highest strike K_3. The strategy aims to profit at times investors believe the underlying is unlikely to move far from its current value.
Lastly, a *calendar spread* refers to the strategy of going short a call option with time to expiry τ_1 and going long a call option with longer time to expiry τ_2 with the same strike price. The strategy aims to profit when investors believe the underlying will become more volatile after expiration of the short option, making the long option extra valuable relative to the short option.

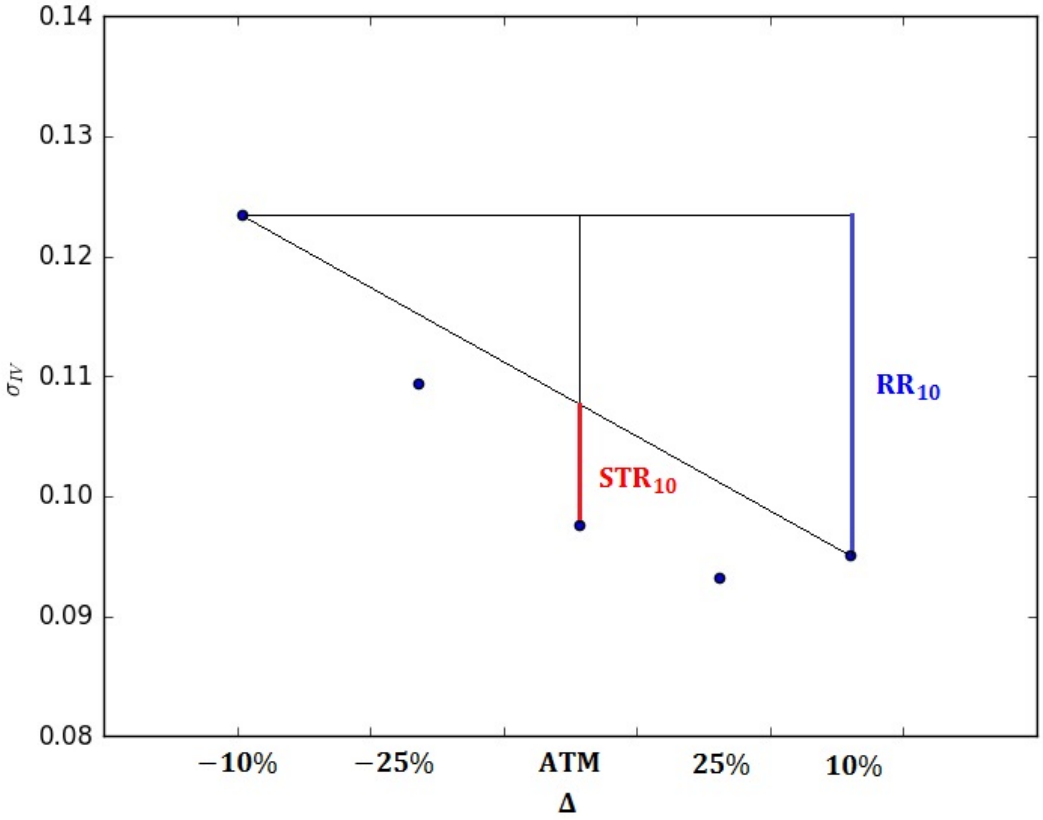

Figure 4: Implied volatility smile for the EUR/USD currency pair with a time to expiry of one year as of 15-06-2016, sorted by delta. The risk reversal and the strangle corresponding to a delta of 10 are shown.

2 Necessary and sufficient no-arbitrage conditions

Several no-arbitrage conditions on the shape of the call price function are known, most of which have been derived by Merton (1973). Hodges (1996) has derived equivalent conditions in terms of the implied volatility, which are nonlinear, as opposed to the call price conditions. This non-linearity is caused by the fact that implied volatility is a nonlinear inverse of the call price function. It is thus easier to implement arbitrage constraints on the call price function than on the implied volatility, which is why we focus here on conditions on the call price function. At the end of this Section, we provide some background on the implied risk-neutral density and the concept of *local volatility*, both crucially tied to the volatility surface.

2.1 Necessary no-arbitrage conditions

According to the First Fundamental Theorem of Asset Pricing, the absence of arbitrage in a market coincides with the existence of an equivalent martingale measure $\mathbb{Q}$, characterized by a corresponding density q_t, under which the value of a derivative is equal to its discounted expected payoff. Under the assumption of deterministic domestic and foreign interest rates $r^d(t)$ and $r^f(t)$, the time t price of a European call is given by

$$C(K,\tau) = e^{-\int_t^T r^d(s)ds}\mathbb{E}_{\mathbb{Q}}\left[(S_T-K)^+\right] = e^{-\int_t^T r^d(s)ds}\int_0^\infty (S_T-K)^+ q_t(S_T)dS_T. \tag{10}$$

From equation (10), we can derive the well-known monotonicity and convexity properties of the call price function. By positivity and integrability to one of the density $q_t(S_t)$, we have

$$-e^{-\int_t^T r^d(s)ds} \leq \frac{\partial C}{\partial K} = -e^{-\int_t^T r^d(s)ds}\int_K^\infty q_t(S_T)dS_T \leq 0. \tag{11}$$

Next, convexity of the call price function follows by taking the second derivative with respect to strike:

$$\begin{aligned}\frac{\partial^2 C}{\partial K^2} &= -e^{-\int_t^T r^d(s)ds}\frac{\partial}{\partial K}\int_K^\infty q_t(S_T)dS_T \\ &= e^{-\int_t^T r^d(s)ds}\frac{\partial}{\partial K}\int_\infty^K q_t(S_T)dS_T \\ &= e^{-\int_t^T r^d(s)ds} q_t(K) \geq 0.\end{aligned} \tag{12}$$

Convexity can also be shown by noting its equivalence to the absence of butterfly spread arbitrage. To see this, we write the second order derivative as the limit of a second difference in the call price with respect to strike divided by the squared change in strike, and let the change in strike go to zero,

$$\frac{\partial^2 C}{\partial K^2} = \lim_{\Delta K\to 0}\frac{C(K+\Delta K,\tau) - 2C(K,\tau) + C(K-\Delta K,\tau)}{(\Delta K)^2} \geq 0,$$

where positivity follows from the fact that the numerator is equivalent to a butterfly spread of infinitisimal width.

With respect to the time to expiry, it follows by equivalence to American option prices in the absence of dividends that call prices should be monotonically increasing. When there are nonzero dividends, or in the FX case, nonzero foreign interest rates, monotonicity for a fixed level of strike no longer holds for European call prices. In the case of nonzero dividends it has been derived by several authors independently (Reiner (2000), Gatheral (2004)) that call prices should be monotonically increasing in the time to expiry dimension for a fixed level of forward moneyness $\kappa = K/F_{t,T}$, a property that also holds for the total implied variance $w = \sigma_{IV}^2 \tau$. We provide here the proof from Gatheral (2004).

Proposition 1. (Gatheral, 2004)
If dividends are proportional to the value of the underlying, the volatility surface is free of calendar spread arbitrage if and only if

$$\frac{\partial C(\kappa,\tau)}{\partial \tau} \geq 0 \quad \textit{for all } \kappa > 0 \textit{ and } \tau > 0, \tag{13}$$

or equivalently, for the total variance it holds that

$$\frac{\partial w(\kappa,\tau)}{\partial \tau} \geq 0 \quad \textit{for all } \kappa > 0 \textit{ and } \tau > 0.$$

Proof. Let $(X_t)_{t\geq 0}$ be a martingale, $L \geq 0$ and $0 \leq t_1 < t_2$. Then the inequality

$$\mathbb{E}\left[(X_{t_2} - L)^+\right] \geq \mathbb{E}\left[(X_{t_1} - L)^+\right] \tag{14}$$

is standard. For any $i = 1, 2$, let C_i be options with strikes K_i and time to expiries T_i. Suppose now that the two options have the same forward-moneyness, i.e.

$$\kappa \equiv \frac{K_1}{F_{t,T_1}} = \frac{K_2}{F_{t,T_2}}.$$

Then, if dividends are proportional to the value of the underlying and if the market is arbitrage-free, the process $(X_t)_{t\geq 0}$ defined by $X_t \equiv S_t/F_{t,T}$ for all $t \geq 0$ is a martingale and

$$\frac{C_2}{K_2} = \frac{\mathbb{E}_{\mathbb{Q}}\left[(S_{T_2} - K_2)^+\right]}{K_2} = \frac{\mathbb{E}_{\mathbb{Q}}\left[(X_{T_2} - \kappa)^+\right]}{\kappa} \geq \frac{\mathbb{E}_{\mathbb{Q}}\left[(X_{T_1} - \kappa)^+\right]}{\kappa} = \frac{\mathbb{E}_{\mathbb{Q}}\left[(S_{T_1} - K_1)^+\right]}{K_1} = \frac{C_1}{K_1},$$

where the first and last equalities follow from the First Fundamental Theorem of Asset Pricing, and the inequality follows from (14).
So, if dividends are proportional to the value of the underlying, arbitrage-free option prices are non-decreasing in time to expiry when keeping the forward-moneyness fixed. We can express the Black-Scholes formula in the form $C^{BS}(\kappa, w(\kappa,\tau))$ with C^{BS} strictly increasing in its second argument, from which it follows that for fixed κ, the total variance $w(\kappa,\cdot)$ must be non-decreasing.

□

It is important to note, that the theorem holds only for moneyness values $\kappa > 0$; the point $\kappa = 0$ corresponds to strikes of 0, where the value of the call option must be equal to $e^{-\int_t^T r^f(s)ds} S_t$. If it were not were not equal, then borrowing the amount $e^{-\int_t^T r^f(s)ds} S_t$ to buy the option with strike 0 and expiry T replicates the forward $F_{t,T}$ but leads to a

different price. The option price at strike 0 thus provides an example of decreasing option prices along the time to expiry dimension for fixed strikes when the foreign interest rate is positive.
Lastly, we find the following necessary bounds on the call prices by arbitrage arguments.

Proposition 2. (Call price bounds for European FX options)
The arbitrage-free call price function is bounded by

$$\left(e^{-\int_t^T r^f(s)ds}S_t - e^{-\int_t^T r^d(s)ds}K\right)^+ \leq C(K,\tau) \leq e^{-\int_t^T r^f(s)ds}S_t. \tag{15}$$

Proof. Clearly the price of any option must be non-negative, as it provides a right, but not an obligation to exercise. Now suppose that

$$e^{-\int_t^T r^f(s)ds}S_t - e^{-\int_t^T r^d(s)ds}K > C(K,\tau).$$

At time t, borrow $e^{-\int_t^T r^f(s)ds}S_t - e^{-\int_t^T r^d(s)ds}K$ from a domestic bank, use this amount to buy the call option $C(K,\tau)$ and sell the forward $F_{t,T}$ for 0. Define the positive remaining amount as $R \equiv e^{-\int_t^T r^f(s)ds}S_t - e^{-\int_t^T r^d(s)ds}K - C(K,\tau) > 0$. At time t, borrow K to exercise the option, and sell the underlying S_T for the agreed forward price $e^{\int_t^T (r^d(s)-r^f(s))ds}S_t$. Use the proceeds to pay back the loan K, as well as the first loan which is now worth $e^{\int_t^T (r^d(s)-r^f(s))ds}S_t - K$. The trading strategy yields costs of 0 at time T, and we still have the profit of R attained at time t.
Suppose now that

$$C(K,\tau) > e^{-\int_t^T r^f(s)ds}S_t.$$

At time t, sell the call option, put the proceeds on a domestic bank account and buy the forward for 0. At expiration time T, use the interest accumulated proceeds, which are worth more than $F_{t,T} = e^{\int_t^T (r^d(s)-r^f(s))ds}S_t$, to buy the underlying S_T for the agreed forward price. Sell the underlying for the exercise price if the call is exercised. Independent of the decision to exercise, the strategy ensures a certain profit. □

2.2 Sufficient no-arbitrage conditions

Carr and Madan (2005) note that it can be hard to determine whether a given set of option prices is arbitrage-free as there is only limited information available.

"*To certify that a given set of asset prices are arbitrage-free in continuous time, the structure of the possible price paths must be specified a priori. For example, one must specify whether the possible price paths are purely continuous, pure jump, or a combination of the two. One must also specify whether the possible price paths display finite or infinite variation over time. Since one can only observe past prices in practice and then only discretely, it is difficult to have any confidence in any particular structure which is imposed on future paths. Nonetheless, a certification that a given set of asset prices is arbitrage-free in the traditional sense does require this specification of the nature of the possible price paths.*"

- Carr and Madan (2005)

One solution to this problem is to alter the technical definition of an arbitrage opportunity. By reducing the size of the information set upon which trading strategies can rely, it becomes easier to certify that a given set of asset prices is free of these restricted arbitrage opportunities. For example, Carr, Geman, Madan, and Yor (2003) introduce the concept of *static arbitrage*, a simplified version of arbitrage in which the information set on which trading strategies may rely is reduced.

Definition 1. (Static arbitrage)
A static arbitrage opportunity is an arbitrage opportunity for which the trading strategy can depend only on time and the current price of the underlying. In particular, positions cannot depend on past prices or their path properties.

An example of a static arbitrage opportunity is a negatively priced butterfly spread: as the payoff of a butterfly spread is non-negative, its price should also be non-negative.
The concept of *dynamic arbitrage* complements static arbitrage to form the set of all arbitrage opportunities.

Definition 2. (Dynamic arbitrage)
A dynamic arbitrage opportunity is an arbitrage opportunity that requires future trading of instruments in order to ascertain a riskless profit.

An example of a dynamic arbitrage opportunity would be exploiting an underpriced option by continuously delta hedging, in a world in which all assumptions of the Black-Scholes model hold.
Because in reality, prices are decomposed in bid and ask prices, even if a dynamic arbitrage opportunity would be available in terms of mid prices, it is unlikely that such an opportunity is profitable in practice as any dynamic arbitrage trading strategy would have to overcome the bid-ask spread multiple times. For this reason, the focus of this thesis with respect to modeling the volatility surface directly is on the exemption of static arbitrage, in line with most current research.
Several sufficient conditions have been derived for the exemption of static arbitrage, although under the assumption of zero interest rates and dividends. For example, Carr and Madan (2005) have derived a set of necessary and sufficient conditions on a rectangular grid of call prices to ensure no static arbitrage, and more recently, Roper (2010) has derived a similar set of conditions on the call price function, which are also transformed into a set of conditions on the implied volatility surface. We follow here Carr and Madan (2005).

Theorem 2. (Carr and Madan, 2005)
Under the assumption of zero interest rates and dividends, a rectangular grid of European option call prices in the strike-time to expiry plane is free of static arbitrage if and only if

1. *All adjacent call spreads are nonnegatively priced;*
2. *All adjacent butterfly spreads are nonnegatively priced; and*
3. *All adjacent calendar spreads are nonnegatively priced.*

Conditions 1 and 2 in Theorem 2 are implied by meeting their continuous version equivalents in the form of conditions (11) and (12) respectively. To cope with nonzero interest rates and dividends (foreign interest rates), condition (13) must be met in order to exclude calendar spread arbitrage. In addition, methods that directly estimate the call price function should impose the necessary bounds on the option prices given in 15.

2.3 Implied risk-neutral density and local volatility

Equation (12) implies that it is possible to compute the implied risk-neutral density q_t when a continuum of call prices is available in the strike dimension. We find the following formula for the implied risk-neutral density:[3]

$$q_t(S_T)\Big|_{S_T=K} = e^{\int_t^T r^d(s)ds}\frac{\partial^2 C}{\partial K^2}. \tag{16}$$

From (16) it immediately follows that non-convexity of the call price function is equivalent to a negative implied risk-neutral density, providing yet another justification for why call prices must be convex in strike. It is also through equation (16), that any method specifying the call price surface (or volatitility surface), makes assumptions on the risk-neutral dynamics of the underlying, be it implicitly. In this light, it would be realistic if the call price function is at least two times continuously differentiable (C^2) in strike, as the corresponding implied risk-neutral density would then be continuous.
The information provided by the volatility surface, the call price surface and the implied risk-neutral densities for all time to expiries combined are equivalent, and using the knowledge of all the implied risk-neutral densities conditional upon a fixed combination (S_t, t), Dupire (1993) derived a unique diffusion process that generates these densities. The diffusion process mentioned is a state-dependent diffusion of the following form

$$dS_t = r(t)S_t dt + \sigma_{loc}(S_t, t)S_t dt, \tag{17}$$

in which $\sigma_{loc}(S_t, t)$ denotes the *local* volatility. As shown by Dupire (1993), under the assumption of zero interest rates, the following expression for $\sigma_{loc}(S_t, t)$ can be derived using the Fokker-Planck (Kolmogorov-forward) differential equation, which links the evolution of the corresponding density over time to the coefficients of the generating diffusion:

$$\sigma_{loc}(K, \tau) = \sqrt{\frac{2\partial C/\partial \tau}{\partial^2 C/\partial K^2}}.$$

In the case of constant foreign and domestic interest rates, we can show (see e.g. Clark (2011)) that the local volatility is given by

$$\sigma_{loc}(K, \tau) = \sqrt{2\frac{\partial C/\partial \tau + (r^d - r^f)K\partial C/\partial K + r^f C)}{K^2\partial^2 C/\partial K^2}}, \tag{18}$$

[3]The result is often attributed to Breeden and Litzenberger (1978), who derived it independently, but 70 years after Bronzin; see equation (17.a) on page 51 in Bronzin (1908).

and the implied volatility version of the local volatility is given by

$$\sigma_{loc}(K,\tau) = \frac{\sigma_{IV}^2 + 2\tau\sigma_{IV}\frac{\partial\sigma_{IV}}{\partial\tau} + 2(r^d - r^f)K\tau\sigma_{IV}\frac{\partial\sigma_{IV}}{\partial K}}{(1 + Kd_1\tau\frac{\partial\sigma_{IV}}{\partial\tau}K)^2 + K^2\tau\sigma_{IV}(\frac{\partial^2\sigma_{IV}}{\partial K^2} - d_1\tau(\frac{\partial\sigma_{IV}}{\partial K})^2)}, \tag{19}$$

where

$$\begin{aligned} d_1 &\equiv \frac{\ln(S_t/K) + (r^d - r^f + 1/2\sigma_{IV})\tau}{\sigma_{IV}\sqrt{\tau}}, \\ \sigma_{IV} &= \sigma_{IV}(K,T). \end{aligned}$$

Equations (18) and (19) show why the call price function and the volatility surface are required to be to be $C^{2\times1}$ when using local volatility as a pricing method: failing to fulfill the $C^{2\times1}$ requirement leads to bumps in the local volatility surface that have no financial economic justification.

3 Arbitrage-free surface methods

In this Section we describe several methods for constructing the arbitrage-free volatility surface. We first discuss two methods that specify a more realistic process for the risk-neutral dynamics of the underlying than (2), namely the SABR model and the lognormal-mixture model, followed by two methods that directly model the volatility surface: constrained moving least-squares and arbitrage-free SVI.
For fair comparison, we tried minimizing the mean absolute error (MAE) with respect to the volatilities (preferred) or the call prices when possible. As all methods except constrained moving least-squares can result in non-convex optimization problems, we first tried to use local optimization with a 'smart' choice of initial parameters, and compared the results to using a global optimization algorithm combined with a local optimization algorithm to refine the solution found. If the results did not improve noticeably, we chose to report the results when using the local optimization methods to gain calibration speed. Otherwise, we chose to report the results corresponding to the hybrid global/local optimization strategy. We thus assign more weight to a model's ability to fit the data than to calibration time in the model comparison, because calibration time depends on the choice of software, hardware and implementation, whereas fitting potential of a model depends only on the implementation (through the choice for starting values and optimization algorithm); model fitting results can thus easier be generalized than calibration time results.

3.1 SABR

The SABR model (short for *stochastic-$\alpha\beta\rho$* model) introduced by Hagan et al. (2002), is a two-factor stochastic volatility model for which a truncated expansion formula for the implied volatility exists. The model is given by

$$\begin{aligned} dF_t &= \alpha_t F_t^{\beta} dW_t, \\ d\alpha_t &= \nu \alpha_t d\tilde{W}_t, \\ \mathbb{C}\mathrm{orr}[W, \tilde{W}] &= \rho, \end{aligned}$$

where ρ denotes the correlation between the two Brownian motions W and $\tilde{W}$, ν denotes the volatility of the volatility (*vovol*) and $\alpha = \alpha_0$ denotes the initial value of the stochastic volatility. The exponent β can be chosen to produce a stochastic normal model ($\beta = 0$), a stochastic lognormal model ($\beta = 1$) or the stochastic Cox-Ingersoll-Ross model ($\beta = 1/2$). The other parameters have clear interpretation with respect to the resulting implied volatility smile and surface: the initial volatility a_0 influences the level, the correlation parameter ρ influences the skew and the vovol parameter ν influences the curvature of the smile.
SABR is particularly popular in fixed-income markets, where assets have only a single expiration date (e.g. swaption markets), as the method is known to be capable of fitting separate volatility smiles well. We will calibrate the model to market prices of all quoted time to expiries using only one parametrization, in order to construct the volatility surface instead of separate smiles. In this way, we ensure that the resulting prices are in line with the time to expiry condition (13). It is of interest to see whether the SABR model

is capable of fitting entire surfaces as opposed to separate volatility smiles.
As the original expansion formula from Hagan et al. (2002) is known not to converge as $\beta \to 1$, as well as providing a poor approximation for small strikes and large time to expiries, we present here the corrected formula for the volatility in terms of log-forward moneyness $x \equiv \ln F/K$ and time to expiry τ as proposed in Oblój (2008):

$$\sigma(x,\tau) = I^0(x)\left(1 + I^1(x)\tau\right) + O(\tau^2),$$

where

$$I^0(x) = \begin{cases} \alpha K^{\beta-1}, & \text{if } x = 0, \\ \frac{x\alpha(1-\beta)}{F^{1-\beta}-K^{1-\beta}}, & \text{if } \nu = 0, \\ \frac{\nu x}{\ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)} & \text{with } z = \begin{cases} \frac{\nu x}{\alpha}, & \text{if } \beta = 1, \\ \frac{\nu(F^{1-\beta}-K^{1-\beta})}{\alpha(1-\beta)}, & \text{if } \beta < 1, \end{cases} \end{cases}$$

and

$$I^1(x) = \frac{(1-\beta^2)\alpha^2}{24(FK)^{1-\beta}} + \frac{\rho\nu\alpha\beta}{4(FK)^{(1-\beta)/2}} + \frac{(2-3\rho^2)\nu^2}{24}.$$

Calibration
The authors note that the choice of the exponent parameter β usually does not have much influence on the resulting implied volatilities; it is either chosen equal to one of $0, 1/2, 1$, or estimated by using the following approximation of the ATM volatility:

$$\ln \sigma_{ATM} \approx \ln \alpha - (1-\beta)\ln F.$$

In this way, we can estimate β by linear regression on a time series of $\ln \sigma_{ATM}$ and $\ln F$. As we choose to model the volatility surface instead of separate smiles, it is unclear which ATM volatility to use, and since the influence of the β parameter is small, we simply set it to $1/2$.
There are two ways to estimate the remaining parameters α, ν and ρ: we can either estimate all parameters at the same time, or we can estimate ρ and ν directly and compute α from the resulting estimates and the ATM volatility. Again, as we are estimating the surface instead of separate smiles, we choose to estimate the remaining three parameters at the same time.
As starting values, we use the ATM vol of the middle quoted time to expiry for α and set the initial vovol parameter ν equal to a moderate value of 0.3. For the initial value of the correlation parameter ρ, we used the 25 delta risk reversal of the middle quoted time to expiry divided by the maximum of the 25 delta put and call volatility to ensure an absolute bound of 1. We used the interior-point implementation in `Matlab`'s `fmincon` function to minimize the MAE with respect to the market volatilities. As the results were satisfactory, we chose here not to use global optimization algorithms.

3.2 Lognormal mixture model

Brigo and Mercurio (2002b) introduce a method to price options when the risk-neutral density of the underlying is based on a mixture of known basic densities, of which the

lognormal choice for basic density is treated as most important example. The lognormal mixture model yields excess kurtosis compared to the normal distribution, but is unable to generate skewness. Therefore, Brigo and Mercurio extend the model to a mixture of shifted lognormal distributions, where the assumption is made that the risk-neutral density of the underlying is equal to the weighted average of known densities of N given diffusion processes

$$\begin{aligned} S_t^i &= \beta_i e^{\mu t} + X_t^i, \;\; i = 1, ..., N, \\ dX_t^i &= \mu X_t^i dt + \sigma_i(t) X_t^i dW_t, \end{aligned}$$

where $\mu \equiv r^d - r^f$ denotes the interest rate differential and $\sigma_i(t)$ denotes the deterministic and possibly time-varying volatility. The form of the shift term $\beta e^{\mu t}$ has been chosen such that it is the most general affine transformation of the S_t^i's that preserves the correct risk-neutral drift. The assumed dynamics lead to the following SDEs:

$$dS_t^i = \mu S_t^i dt + \sigma_i(t)(S_t^i - \beta_i e^{\mu t}) dW_t, \;\; i = 1, ..., N,$$

and the corresponding risk-neutral density of the underlying S is given by

$$q_t(x) \equiv \frac{d}{dx}\mathbb{Q}_t(S_T \leq x) = \sum_{i=1}^{N} \lambda_i \frac{d}{dx}\mathbb{Q}_t(S_T^i \leq x) = \sum_{i=1}^{N} \lambda_i q_t^i(x), \tag{20}$$

where the q_t^i's are the probability densities of the separate diffusion processes S^i and the λ_i's are strictly positive constants such that $\sum_{i=1}^{N} \lambda_i = 1$. The λ_i's can be interpreted as the probability that the underlying S will take on the values of the shifted diffusion component S^i. The density of the underlying is thus a convex combination of the densities corresponding to the N separate diffusion processes that are given by[4]

$$\begin{aligned} q_t^i(x) = &\frac{1}{(x - \beta_i e^{\mu t}) V_i(t) \sqrt{2\pi}} \exp\left[-\frac{1}{2V_i^2(t)} \left(\ln\left(\frac{x - \beta_i e^{\mu t}}{S_0 - \beta_i} \right) - \mu t \frac{1}{2} V_i^2(t) \right)^2 \right], \\ &x \geq \beta_i e^{\mu t}, \end{aligned}$$

where $V_i(t) \equiv \sqrt{\int_0^t \sigma_i^2(u) du}$ denotes the integrated volatility. It is important to note here, that distribution formed by the mixture of densities is not the same as the distribution corresponding to the random variable formed by the weighted average of the corresponding separate diffusion processes, even if the weights are the same. For example, when the basic densities are normal, the former leads to distribution with excess kurtosis and (usually) multiple modes, whereas the latter leads to a normal distribution again.
Application of the Fokker-Planck equation on density (20) yields the local volatility process

$$dS_t = \mu S_t dt + \sqrt{\frac{\sum_{i=1}^{N} \lambda_i \sigma_i^2(t)(S_t - \beta_i e^{\mu t})^2 q_t^i(S_t)}{\sum_{i=1}^{N} \lambda_i q_t^i(S_t)}} dW_t. \tag{21}$$

[4]For consistency with Brigo and Mercurio (2002b), we will use here S_0 to denote the current value instead of S_t.

SDE (21) must be regarded as defining some candidate dynamics that leads to the density (20), and can be used to price exotic options.
We can use the First Fundamental Theorem of Asset Pricing to show that option prices based on a convex combination of densities for the underlying leads to the same convex combination of option prices based on the individual densities. For example, in case of a call option we have

$$\begin{aligned}
C(K,\tau) &= e^{-\int_t^T r^d(s)ds}\mathbb{E}_{\mathbb{Q}}\left[(S_T-K)^+\right] \\
&= e^{-\int_t^T r^d(s)ds}\int_0^\infty (S_T-K)^+ q_t(S_T)dS_t \\
&= e^{-\int_t^T r^d(s)ds}\int_0^\infty (x-K)^+ \sum_{i=1}^N \lambda_i q_t(x)dS_t \\
&= \sum_{i=1}^N \lambda_i e^{-\int_t^T r^d(s)ds}\int_0^\infty (x-K)^+ q_t(x)dS_t \\
&= \sum_{i=1}^N \lambda_i C_i(K,\tau).
\end{aligned}$$

The above result is particularly beneficial when the basic densities are lognormal, as then the mixture of densities pertains closed form expressions for option prices. The formula for the call option price based on the density (20), assuming $S_0-\beta_i>0$ and $K-\beta_i e^{\mu\tau}>0$ for all densities i and strikes of interest K, is therefore[5]

$$\begin{aligned}
C(K,\tau) &= e^{-r^d\tau}\sum_{i=1}^N \lambda_i\left[(S_0-\beta_i)e^{\mu\tau}N\left(d_1^i\right)-(K-\beta_i e^{\mu\tau})N\left(d_2^i\right)\right], \\
d_1^i &\equiv \frac{\ln\left(\frac{S_0-\beta_i}{K-\beta_i e^{\mu\tau}}\right)+(\mu+\frac{1}{2}\eta_i^2)\tau}{\eta_i\sqrt{\tau}}, \\
d_2^i &\equiv d_1^i-\eta_i\sqrt{\tau}, \\
\eta_i &\equiv \frac{V_i(T)}{\sqrt{T}} = \sqrt{\frac{\int_0^T \sigma_i^2(u)du}{T}}.
\end{aligned}$$

For the functional form of the integrated volatilities $\eta_i(T)$, the authors choose

$$\eta_i(T)=\eta(T;a_i,b_i,c_i,d_i)=a_i+b_i\left[1-\exp\left(-\frac{T}{d_i}\right)\right]\frac{d_i}{T}+c_i\exp\left(-\frac{T}{d_i}\right).$$

Calibration
When calibrating the mixture of an N-component model, the parameter vector to optimize over $\boldsymbol{\theta}\equiv(\lambda_1,...\lambda_{N-1},\beta_1,...\beta_N,a_1,...,a_N,b_1,...b_N,c_1,...,c_N,d_1,...,d_N)$ has dimensionality $(N-1)+N+4N=6N-1$. The number of available parameters to fit the

[5]See formula 2.17 and the corresponding derivation in Brigo and Mercurio (2002a) for the call price formula corresponding to a single displaced diffusion.

data can be made arbitrary large by choosing the number of basic densities N arbitrary large, but we must make a trade-off between calibration speed and data fitting quality. We therefore choose $N = 2$ basic densities, in line with Brigo and Mercurio (2002b), who recommend the use of either two or three basic densities.
As criterion function, we choose to minimize the MAE with respect to the call prices. In addition, we include large penalties in the criterion function to ensure that for all basic densities i, all strikes of interest K and all time to expiries τ_j for which market prices are available the following conditions hold:

$$\begin{aligned} S_0 - \beta_i &\geq 0, \\ K - \beta_i e^{\mu\tau} &> 0, \\ v_{i,j} &\geq 0, \text{ where } v_{i,j} \equiv \eta_i(\tau_j), \text{ and} \\ v_{i,j+1} &\geq v_{i,j}\sqrt{\tau_j/\tau_{j+1}}, \end{aligned}$$

where the third condition is included to avoid imaginary values for the functions σ_i.
Unfortunately, the optimization problem posed by fitting the option prices to the data is a non-convex optimization problem, and Brigo and Mercurio (2002b) therefore propose to use a global search algorithm with few parameters after which the search can be refined with a local search optimization algorithm. We choose to fit the model initially using the Genetic Algorithm method implemented in `Matlab`'s `ga` function with all parameters, and refine the solution with the BFGS method. By performing global optimization on all parameters, we try to find out how well the model can fit the data when using two basic densities. Such calibration procedure will result in relatively long calibration times, but is consistent with our choice of assigning more weight to fitting the data than on calibration speed in the model comparison.

3.3 Constrained moving least-squares

Glaser and Heider (2012) propose to construct the call price surface free from static arbitrage by using moving least-squares (MLS) subject to no-arbitrage constraints. The method is a local approximation method, i.e. a different function is estimated depending on the point in the strike-time to expiry plane of interest. The local nature of the method ensures the method is efficient. The implied volatility surface is obtained by numerically inverting the Black-Scholes formula.
Suppose that the values of a function $f : \Omega \to \mathbb{R}$ with $\Omega \in \mathbb{R}^2$ are given by a discrete set of points $X = \{x_1, ..., x_n\} \in \Omega$. The idea of MLS is to solve for every $x \in \Omega$ a weighted least-squares fit. For a given $x \in \Omega$, the value of the approximant is given by the solution of

$$\begin{aligned} \min_{p \in \Pi} \quad & \frac{1}{2} \sum_{j \in I(x)} \Phi_\delta(x - x_j) \cdot (p(x_j) - f(x_j))^2 \\ \text{subject to} \quad & (Bp) \leq 0, \end{aligned} \tag{22}$$

where Π is a suitable polynomial space, $\Phi_\delta : \Omega \to \mathbb{R}$ is a weighting function with parameter δ, $I(x)$ denotes a set of nearest neighbors of x and B is an operator used to impose linear constraints on the polynomial p. Glaser and Heider define the polynomial space $\Pi \equiv (1, K, K^2, \tau)$ generated by the monomials $1, K, K^2$ and τ, resulting in the smallest

polynomial space that satisfies $\Pi \subset C^{2\times 1}$.
For the call price function, note that for fixed (K^*, τ^*), one can represent any $C(K,\tau) \in \Pi$ as

$$c(\kappa,\tau) = a_0 + a_1(K-K^*) + a_2(\tau-\tau^*) + \frac{a_3}{2}(K-K^*)^2, \tag{23}$$

with unique $\boldsymbol{a} = (a_0, a_1, a_2, a_3) \in \mathbb{R}^4$, so that any call price function can be seen as a vector in $\mathbb{R}^4$. This is convenient, as this representation allows for imposing linear constraints on the call price function.
Glaser and Heider (2012) restate time to expiry condition (13) as

$$C(Ke^{(r^d-r^f)\Delta\tau}, \tau+\Delta\tau)e^{r^f\Delta\tau} - C(K,\tau) \geq 0,$$

and show that taking the limit of $\Delta\tau$ to zero, one finds that

$$\frac{\partial C}{\partial \tau} + (r^d - r^f)K\frac{\partial C}{\partial K} + r^f C(K,\tau) \geq 0. \tag{24}$$

To impose arbitrage constraints (11), (12), (24) and (15), one can define for fixed (K,τ) on Π the operator $B_{K,\tau} : \Pi \to \mathbb{R}^5$ by imposing [6]

$$B_{K,\tau}c \leq \boldsymbol{b},$$

where

$$B_{K,\tau}c = \begin{pmatrix} \frac{\partial C(K,\tau)}{\partial K} \\ -\frac{\partial^2 C(K,\tau)}{\partial K^2} \\ -\frac{\partial C}{\partial \tau} - (r^d - r^f)K\frac{\partial C(K,\tau)}{K} - r^f C(K,\tau) \\ -C(K,\tau) \\ C(K,\tau) \end{pmatrix},$$

$$\boldsymbol{b} \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\left(e^{-\int_t^T r^d(s)ds}S_t - e^{-\int_t^T r^f(s)ds}K\right)^+ \\ e^{-\int_t^T r^f(s)ds}S_t \end{pmatrix},$$

which can be represented by means of equation (23) as a matrix $\boldsymbol{B} \in \mathbb{R}^{5\times 5}$,

$$\boldsymbol{B}_{K,\tau}\boldsymbol{a} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -r^f & -(r^d - r^f)K & -1 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{a} \leq \boldsymbol{b}. \tag{25}$$

[6] Glaser and Heider (2012) only impose the lower bound $C(K,\tau) \geq 0$, which is not tight enough, as can be seen from the bounds on the call price function in (15).

The call price function and the corresponding restrictions are discussed, but we still need to provide details on several quantities from optimization problem (22). First, the following exponential weighting formula is used to weight the observations:

$$\phi_\delta(x) = \exp(-\|S \cdot x\|_2^2/\delta^2),$$

where δ is a parameter and $\boldsymbol{S} \in \mathbb{R}^{2\times 2}$ is a scaling matrix. For any pair (K, τ) we define the corresponding index set of nearest neighbors

$$I(K,\tau) \equiv \left\{ i \in \{1, ..., N\} \middle| \quad \left\| \boldsymbol{S} \cdot \begin{pmatrix} K_i - K \\ \tau_i - \tau \end{pmatrix} \right\|_2 < \delta \right\}. \tag{26}$$

The index set I contains only those indices of observed calls having similar strike and time to expiry as the point of interest. The scaling matrix S is used to relate strike and time to expiry, which are given different length scales and quantities. For every $j \in I(K,\tau)$, let $\Delta_K^j \equiv K_j - K$ and $\Delta_\tau^j \equiv \tau_j - \tau$ denote the differences in strike and time to expiry of the observed call prices relative to the pair (K,τ). We also introduce the positive definite diagonal weighting matrix

$$\boldsymbol{\Phi} \equiv \operatorname{diag}\left(\phi\left(\left\| S \cdot \begin{pmatrix} K_j - K \\ \tau_j - \tau \end{pmatrix} \right\|_2 \right) \right) \in \mathbb{R}^{|I(K,\tau)|\times|I(K,\tau)|},$$

the evaluation matrix

$$\boldsymbol{A} \equiv \begin{pmatrix} 1 & \Delta_K^1 & \Delta_\tau^1 & \frac{1}{2}(\Delta_K^1)^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta_K^{|I(K,\tau)|} & \Delta_\tau^{|I(K,\tau)|} & \frac{1}{2}(\Delta_K^{|I(K,\tau)|})^2 \end{pmatrix} \in \mathbb{R}^{|I(K,\tau)|\times 4},$$

and the data vector

$$\boldsymbol{y} \equiv (C_j)_{j\in I(K,\tau)} \in \mathbb{R}^{|I(K,\tau)|}.$$

Using the matrix $\boldsymbol{B}$ from (25), we can now formulate equation (22) as a linearly constrained linear least square problem:

$$\begin{aligned} \min_{\boldsymbol{a}\in\mathbb{R}^4} \quad & \frac{1}{2}\|\boldsymbol{\Phi}^{1/2}(\boldsymbol{A}\boldsymbol{a} - \boldsymbol{y})\|_2^2 \\ \text{subject to} \quad & \boldsymbol{B}\boldsymbol{a} \le \boldsymbol{b}, \end{aligned} \tag{27}$$

with $\boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2} = \boldsymbol{\Phi}$. Optimization problem (27) allows for a QP representation (see Rasmussen (2016), Appendix A.1):

$$\begin{aligned} \min_{\boldsymbol{a}\in\mathbb{R}^4} \quad & \frac{1}{2}\boldsymbol{a}'(\boldsymbol{A}'\boldsymbol{\Phi}\boldsymbol{A})\boldsymbol{a} - (\boldsymbol{A}'\boldsymbol{\Phi}\boldsymbol{y})'\boldsymbol{a} \\ \text{subject to} \quad & \boldsymbol{B}\boldsymbol{a} \le \boldsymbol{b}. \end{aligned} \tag{28}$$

Determination of S and δ

To construct appropriate index sets, Glaser and Heider (2012) suggest the following. First, prescribe how many strike and time to expiry values per direction should be contained

in the index set $I(K,\tau)$. Then, identify the necessary distance from the evaluation point (K,τ) for the strike and maturity direction to achieve this preset. Let χ_1 be the distance in the strike direction and χ_2 be the distance in the maturity direction. Define the scaling matrix by

$$\boldsymbol{S} \equiv \begin{pmatrix} \frac{1}{\chi_1} & 0 \\ 0 & \frac{1}{\chi_2} \end{pmatrix}.$$

Setting $\delta = 1$ the index set is then determined by the ellipse

$$I(K,\tau) = \left\{ i \in \{1,...,N\} \middle| \frac{(K_i - K)^2}{\chi_1^2} + \frac{(\tau_i - \tau)^2}{\chi_2^2} < 1 \right\}.$$

The cardinality of the index set is fully controlled by the parameters χ_1 and χ_2. Because the degree of clustering of the observation points in the $(K,\tau)-$plane varies, the distances $\chi_1(K,\tau)$ and $\chi_2(K,\tau)$ are dependent on (K,τ). Glaser and Heider (2012) find by means of simulation that a good performance of the algorithm is obtained by determining $\chi_1(K,\tau)$ and $\chi_2(K,\tau)$ so that market data on three maturity levels with about 10 data points per time to expiry levels are included in each index set $I(K,\tau)$.

Calibration
To calibrate the model to observed prices, we use `Matlab`'s `quadprog` function set to its 'interior-point-convex' algorithm applied on QP (28).

3.4 Arbitrage-free SVI

Gatheral and Jacquier (2014) find that it is seemingly impossible to determine conditions for the exemption of butterfly arbitrage on the original SVI formulation (5). Another drawback of formula (5) is that its parameter values are not very intuitive. The SVI Jump-Wings (SVI-JW) parameterization provides a more intuitive representation of the volatility smile, given in terms of the original SVI parameters by

$$\begin{aligned}
v_t &\equiv \frac{a + b(-\rho m + \sqrt{m^2 + c^2})}{t} \geq 0, && \text{(ATM variance)}, \\
\psi_t &\equiv \frac{b}{2\sqrt{w_t}} \left(-\frac{m}{\sqrt{m^2 + c^2} + \rho} \right) \in \mathbb{R}, && \text{(ATM skew)}, \\
p_t &\equiv \frac{1}{\sqrt{w_t}} b(1 - \rho) \geq 0, && \text{(left (put) wing slope)}, \\
c_t &\equiv \frac{1}{\sqrt{w_t}} b(1 + \rho) \geq 0, && \text{(right (call) wing slope)}, \\
\tilde{v}_t &\equiv \frac{1}{t}(a + bc\sqrt{1 - \rho^2} \geq 0, && \text{(minimum implied variance)}.
\end{aligned} \tag{29}$$

Similarly, the parameters of the original SVI model can be expressed in terms of SVI-JW parameters, which allows for computation of implied variances using the SVI-JW model. To attain arbitrage-free volatility surfaces, the authors first derive that for yet another class of SVI volatility surfaces called *surface SVI* (SSVI), the absence of butterfly arbitrage is guaranteed for all time to expiries. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth function

such that the limit $\lim_{\tau\to 0}\theta_t\varphi(\theta_\tau)$ exists in $\mathbb{R}$, and let θ_τ denote the ATM total variance at time to expiry τ. The SSVI surface is then defined by

$$w(x,\theta_\tau;\rho)=\frac{\theta_\tau}{2}(1+\rho\varphi(\theta_\tau))x+\sqrt{(\varphi(\theta_\tau)x+\rho)^2+(1-\rho)^2}, \tag{30}$$

with log-forward moneyness $x=\ln(K/F_{t,T})$ and parameter $\rho\in[-1,1]$. Two parametrizations for $\varphi(\cdot)$ are proposed:

$$\begin{aligned}&\varphi(\theta)\equiv\frac{1}{\lambda\theta}\left(1-\frac{1-e^{-\lambda\theta}}{\lambda\theta}\right), && \text{(Heston-like parametrization)},\\ &\varphi(\theta)\equiv\eta\theta^{-\gamma}, && \text{(Power-law parametrization)},\end{aligned}$$

with λ,η and γ parameters that characterize the parametrization. It is shown in Gatheral (2011) that the implied variance skew of the Heston-like parametrization is consistent with the Heston model.
The parameters of the SSVI model allow for a transformation to SVI-JW parameters (29), from which it is derived that for SVI-JW parameter values v_τ,ψ_τ and p_τ we may guarantee a smile without butterfly arbitrage by choosing the remaining two parameters as

$$c_\tau^*\equiv p_\tau+2\psi_\tau \text{ and } \tilde{v}_\tau^*\equiv v_t\frac{4p_\tau c_\tau^*}{(p_\tau+c_\tau^*)^2}. \tag{31}$$

To exclude calendar spread arbitrage, the time to expiry slices are calibrated iteratively from last to first expiry with the inclusion of a penalty term in the criterion function that is active when two slices cross in terms of total variance at the same level of log-forward moneyness x. The entire procedure is summarized in Algorithm 1.

Algorithm 1. Arbitrage-free SVI calibration

1. Fit the SSVI model (30) to the quoted volatilities by minimizing the sum of squared differences.
2. Starting with the SSVI as initial guess in SVI-JW representation, optimize over the SVI-JW parameters v_τ,ψ_τ and p_τ slice by slice starting with the longest time to expiry minimizing the sum of squared differences to the quoted volatilities, with a large penalty term included in the criterion function when crossing either the previous or the next slice in terms of total variance. During and after the optimization, the two remaining parameters c_τ and $\tilde{v}_\tau$ are determined by (31).

Calibration

From (29) and (30) it follows that fitting the SSVI and SVI-JW models to quoted prices does not need to be a convex optimization problem. The use of either local or global optimization to fit the SSVI model did not appear to have a significant impact. For the

SVI-JW fitting this appears to be different; because the MAE for both the volatilities as well as the call prices improved by roughly a factor 2 compared to local optimization, we decided to use the Genetic Algorithm method implemented in `Matlab`'s `ga` function and refined the estimated solution with `Matlab`'s active-set algorithm as implemented by the `fmincon` function. This caused the average calibration times to increase with a factor between 3 and 4.

3.5 Other models considered

The following methods were considered for comparison, but were left out due to time constraints or other reasons.

Benko, Fengler, Härdle, and Kopa (2007) try to approximate the implied volatility surface by calibrating a two dimensional Taylor approximation with respect to the strike and time to expiry to the quoted implied volatilities. Convexity of the call price function is ensured, as well as exemption from calendar spread arbitrage. The method is similar to the constrained MLS method, but applied to volatilities instead of call prices. A resulting drawback is that the method leads to an optimization problem with non-linear constraints, which can be hard to solve.

Fengler (2009) estimates a different call price function for each time to expiry by a natural cubic spline with shape constraints to incorporate restrictions (11), (12) and (13) as well as call price bounds (15). The method cannot be directly applied to FX option price data, because the market prices form an irregular grid in the strike (or moneyness) - time to expiry plane. As the method cannot reprice all market prices without specifying a separate extrapolation formula, we choose to exclude the method from the comparison.

Fengler and Hin (2015) transform the call prices under a homogeneity assumption to impose arbitrage constraints, and estimate the call price surface by the use of a bivariate tensor-product B-spline. It is unclear whether the resulting untransformed call prices are ensured to be free from arbitrage.

4 Simulation of FX option price data

Realistic simulation of FX option price data in the form of volatility surfaces yields two main challenges: choosing a realistic process for the risk-neutral dynamics of the underlying and ensuring consistency with FX market conventions (in particular quoting prices by delta). We focus in this Section on dealing with the second challenge, bypassing the first challenge by proposing a method that can be used for any model of choice for the underlying.

We start by constructing a volatility surface based directly on prices from an underlying model. Real option prices, however, are not solely driven by the dynamics of the underlying, but also by supply and demand in the options market. To make our prices more realistic, we perturb the prices with random noise terms that follow the Uniform distribution. The sizes of noise terms are inspired by what we know of bid-ask spread tendencies; the errors will be larger as the absolute value of moneyness of the options increases and for specific time to expiries of which the options are usually illiquid. The addition of noise terms to mimic market noise allows us to simulate a large number of realistic FX option price data sets in a short amount of time, which aids testing and comparison of option pricing methods.

4.1 Generating the FX volatility surface without market noise

Given the choice for the dynamics of the underlying and corresponding parametrization, we have available European option prices for all strikes and time to expiries (e.g. by the use of Monte Carlo simulation) from which we can compute the volatility surface $\sigma(K, \tau)$ by numerical inversion; we use the method of Jäckel (2006) to compute accurate implied volatilities. Then, generating prices consistent with FX market conventions means we must provide implied volatilities corresponding to deltas of $-10, -25, 25$ and 10 as well as the ATM volatility. For the ATM volatility, we know how to compute the strike given the convention (ATMF or DNS), and by plugging it into the volatility surface we find the ATM volatility. Computing the other strikes and corresponding volatilities requires more effort as we do not know which strikes and volatilities correspond to the deltas of interest.

When the market quotes the volatility corresponding to a certain delta, we need to solve for the strike given the corresponding volatility σ_{IV}. For example, let us assume simple delta convention,

$$\Delta_{simple} = \omega N(\omega d), \quad d = \frac{\ln(F_{t,T}/K)}{\sigma_{IV}\sqrt{\tau}}, \quad \omega = \begin{cases} 1 & \text{if } \Delta_{simple} \geq 0, \\ -1 & \text{if } \Delta_{simple} < 0. \end{cases}$$

Inverting the formula to make it a function of strike yields the following root-finding problem:

$$K \exp[\sigma_{IV}\sqrt{\tau}N^{-1}(\Delta_{simple}/\omega)/\omega] - F_{t,T} = 0,$$

in which Δ_{simple} is a fixed value from the set $\{-10, -25, 25, 10\}$. This root-finding problem has closed form solution[7]

$$K = \exp[-\sigma_{IV}\sqrt{\tau}N^{-1}(\Delta_{simple}/\omega)/\omega]F_{t,T}. \tag{32}$$

[7]Ignoring here the intractability of $N^{-1}(\cdot)$, for which accurate numerical values are available.

When generating prices based on the dynamics of the underlying, we can construct a similar root-finding problem to find the strikes and implied volatilities corresponding to specific values of delta, by simply replacing the fixed implied volatility σ_{IV} by the volatility surface $\sigma_{IV}(K,\tau)$:

$$f(K) \equiv K \exp[\sigma_{IV}(K,\tau)\sqrt{\tau}N^{-1}(\Delta_{simple}/\omega)/\omega] - F_{t,T} = 0. \tag{33}$$

The root-finding problem no longer has a closed form solution as $\sigma_{IV}(K,\tau)$ now varies with strike, but we can still solve for the strike numerically by using a root-finding algorithm. Figure 5 shows how function (33) looks like for a one-years option with a simple delta value of 25% and a particular parametrization of the Bates model. The root of the function is seen to be unique. This argument can be applied for any delta convention of choice, and we provide the root functions for the other delta conventions in Table 1. Figure 6 shows an example of a generated FX-volatility surface.

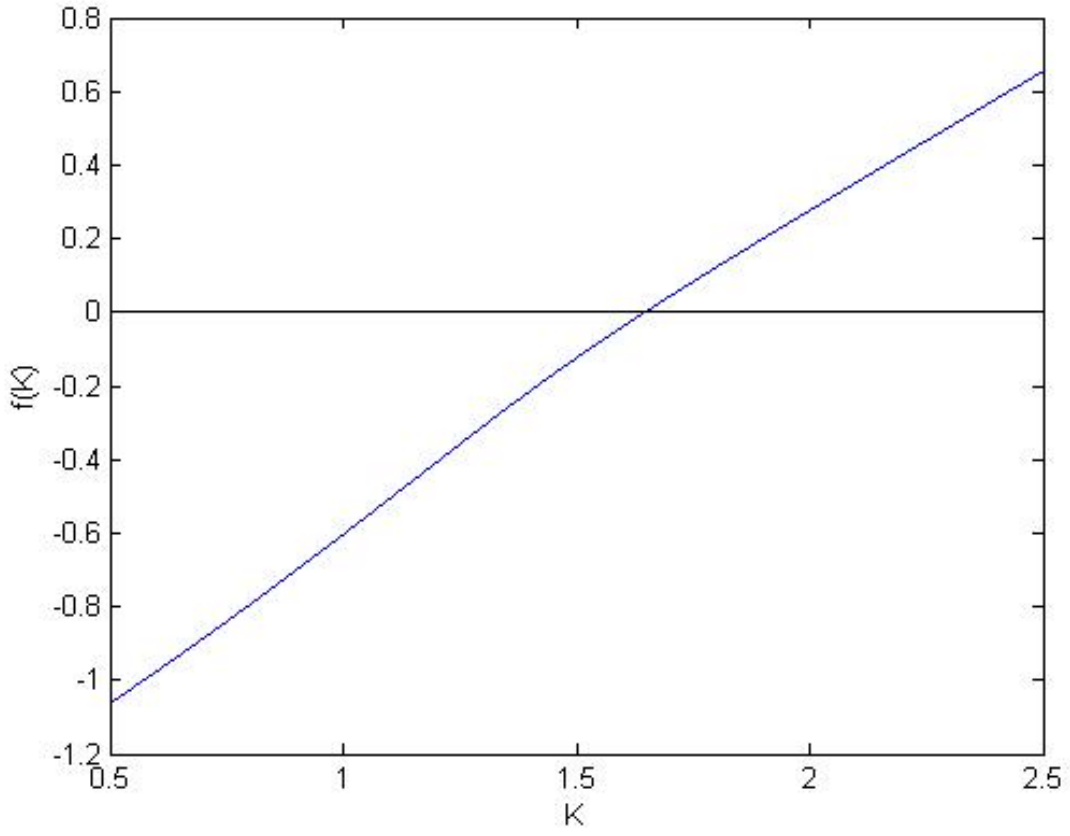

Figure 5: Example of the root function given in (33) for a one-years option with a simple delta of 25% and the moderate smile parametrization of the Bates model given in Section 5.2.

Delta convention	Root function $f(K)$
Spot pips	$K\exp\left[\sigma_{IV}(K,\tau)\sqrt{\tau}N^{-1}\left(e^{r^f\tau}\Delta_{S;pips}/\omega\right)/\omega-\frac{1}{2}\sigma_{IV}^2(K,\tau)\tau\right]-F_{t,T}$
Forward pips	$K\exp\left[\sigma_{IV}(K,\tau)\sqrt{\tau}N^{-1}\left(\Delta_{F;pips}/\omega\right)/\omega-\frac{1}{2}\sigma_{IV}^2(K,\tau)\tau\right]-F_{t,T}$
Spot percentage	$K\exp\left[\sigma_{IV}(K,\tau)\sqrt{\tau}N^{-1}\left(\frac{e^{r^d\tau}S_t\Delta_{S;\%}}{\omega K}\right)+\frac{1}{2}\sigma_{IV}^2(K,\tau)\tau\right]-F_{t,T}$
Forward percentage	$K\exp\left[\sigma_{IV}(K,\tau)\sqrt{\tau}N^{-1}\left(\frac{F_{t,T}\Delta_{F;\%}}{\omega K}\right)+\frac{1}{2}\sigma_{IV}^2(K,\tau)\tau\right]-F_{t,T}$

Table 1: Root functions for the other FX delta conventions.

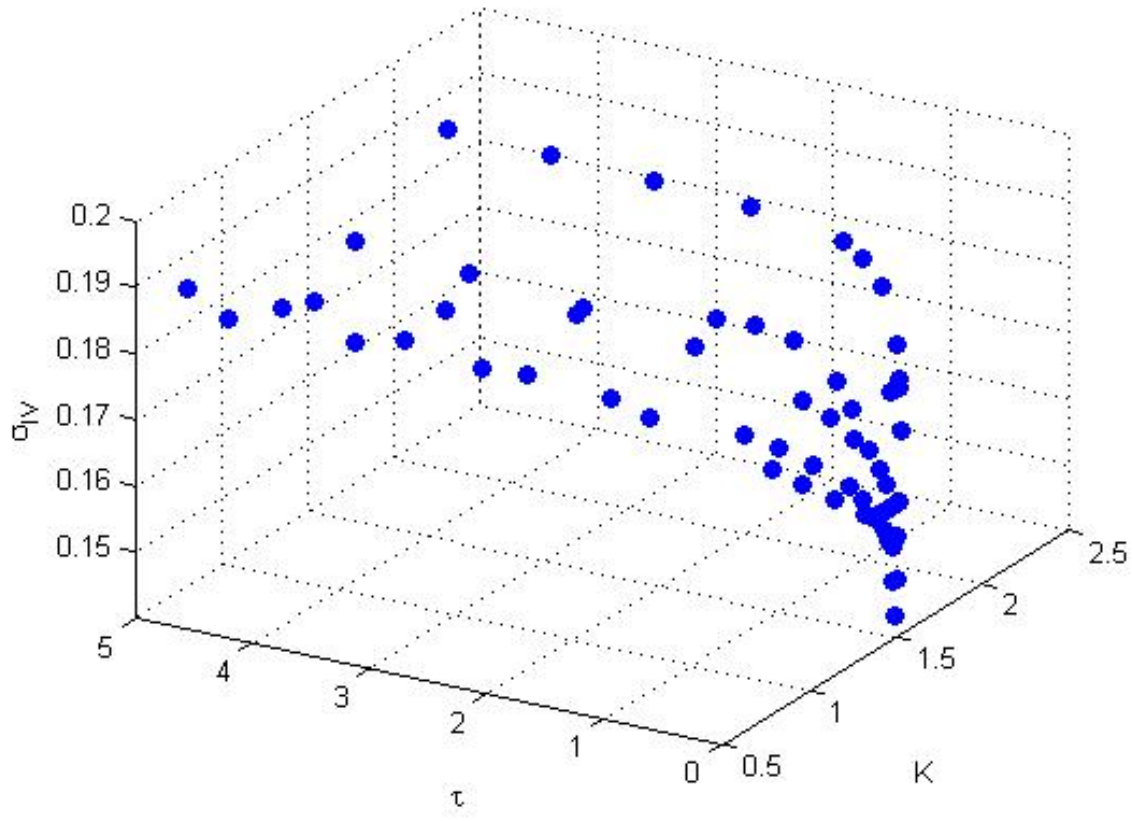

Figure 6: Example of a generated FX volatility surface, constructed using the root function given in (33) and the moderate smile parametrization of the Bates model given in Section 5.2. The value of the underlying is set to 1.5.

4.2 Perturbing the surface

Given a volatility surface based on dynamics of the underlying, we still need to add noise terms to make the prices more realistically. As such noise is caused by supply and demand, we decide to let the size of the perturbations be inspired by what we know of the tendencies of bid-ask spreads observed in the market. The basic volatility surface provides us with five implied volatilities per time to expiry corresponding to specific deltas and the ATM strike, but real FX market data is quoted in terms of the ATM volatilities, risk reversals and strangles. Thus, if we want to perturb the data in a realistic way, we must transform the implied volatilities into risk reversal and strangle quotes using formulas (7) and (8), and perturb those quotes instead.

In general, we know that out of the money options are less liquid then in the money options, which are less liquid than at the money options. By perturbing the ATM volatilities, the risk reversals and the strangles separately, we are ensured that the ATM volatility is subject to less noise than the other volatilities. The non-ATM volatilities always correspond to OTM options (puts or calls), and we know from the definitions of the delta conventions that the 10% delta volatilities are more out of the money than the 25% volatilities. We propose therefore to let the perturbation on the risk reversal and strangle quotes increase with the largest absolute value of log-moneyness $|\ln(K/F_{t,T})|$ of the corresponding two volatilities, a quantity that approximates the percentage of moneyness.[8]

Regarding the time to expiry dimension, we know that options with short time to expiries (shorter than 1M) and options with long time to expiries (2Y and on) have larger bid-ask spreads than options with moderate time to expiries. Therefore, we propose to increase the size of the perturbations by a constant factor c for the short and longer time to expiries, say $c = 2$. Alternatively, we might multiply the quotes at these time to expiries with different factors to make the noise even more realistic.

We can ensure that the volatilities will remain inside a chosen bid-ask spread by drawing noise terms from the Uniform distribution; if the bid ask spread is for example $-0.01/0.01$ on top of the quotes, then we simulate random errors $\tilde{U} \equiv 2 \times 0.01 \times U - 0.01$, where $U \sim \text{Uniform}[0, 1]$. These noise terms will form the basic noise terms to perturb the quotes, possibly multiplied by a factor corresponding to the option's moneyness or time to expiry as mentioned. After perturbing the ATM volatilities, risk reversals and strangles, we convert these perturbed quotes back to volatilities using (9), and recompute strikes using the closed form expression (32). The entire procedure is summarized in Algorithm 2.

[8] A remark is in order regarding the bid-ask spread conventions in the FX market. Many parties quote a bid-ask spread only for the ATM volatility, and quote the risk reversals and strangles by midprice. The rationale is that the options computed using OTM volatilities are worth less than the ATM option, hence using the bid-ask spread from the ATM volatility to express uncertainty would have a larger effect on these OTM options. However, this reasoning does not take into account the fact that OTM options have a lower vega than ATM options and that OTM option prices are smaller in absolute value than ATM options. Thus, quoting this way is likely to result in less uncertainty on the prices of OTM options than ATM options, which is counterintuitive as OTM options are bets on events less likely than those bet on in ATM options. But even if one would prefer to base simulated prices on such bid-ask spread quotation, risk reversals and strangles should still be attributed more noise than ATM options, as we aim to model market noise stemming from supply and demand, and not the bid-ask spread itself.

An example of a perturbed surface is shown in Figure 7. The added noise will sometimes lead to arbitrage by violation of convexity and (less likely) monotonicity of the call price function as is illustrated by Figure 8. Such violations are not uncommon in real data, caused by the use of midprices or data of bad quality (see Fengler (2009)). Nonetheless, it is possible to remove these arbitrage occurrences from the simulated data by redrawing errors whenever for given strikes $K_1 < K_2 < K_3 < K_4 < K_5$, at a fixed time to expiry the inequalities

$$-e^{-\int_t^T r(s)ds} \leq \frac{C(K_i,\tau) - C(K_{i-1},\tau)}{K_i - K_{i-1}} \leq \frac{C(K_{i+1},\tau) - C(K_i,\tau)}{K_{i+1} - K_i} \leq 0, \;\; i = 2,3,4,$$

are violated, or when the ATM calls are not increasing in time to expiry when ATMF convention is used. Such a redrawing procedure is unlikely to make the data more realistic, and we mention it only for the sake of completeness; we will not apply it in the simulation study in the next Section.

Algorithm 2. Simulation of FX option price data

1. Compute unperturbed implied volatilities corresponding to the deltas of interest, by choosing a model for the underlying and solving equation (33) numerically;
2. Transform the implied volatilities to quotes for the risk reversals and the strangles using equations (7) and (8);
3. Define a bid-ask spread $-\beta/\beta$. Draw random noise terms using the Uniform$[0,1]$ distribution, and transform them by $\tilde{U} \equiv 2 \times \beta \times U - \beta$. Multiply the transformed errors by $1 + |\ln(K/F_{t,T})|$. Multiply the errors for short time to expiries (e.g. shorter than 1M) and long time to expiries (e.g. 2Y and on) by a constant of choice c. Add the resulting noise terms to the ATM implied volatilities, the risk reversals and the strangles separately;
4. Transform the perturbed ATM implied volatilities, risk reversal and strangle quotes back to implied volatilities using equation (9);
5. Compute the strikes corresponding to the perturbed implied volatilities by using equation (32).

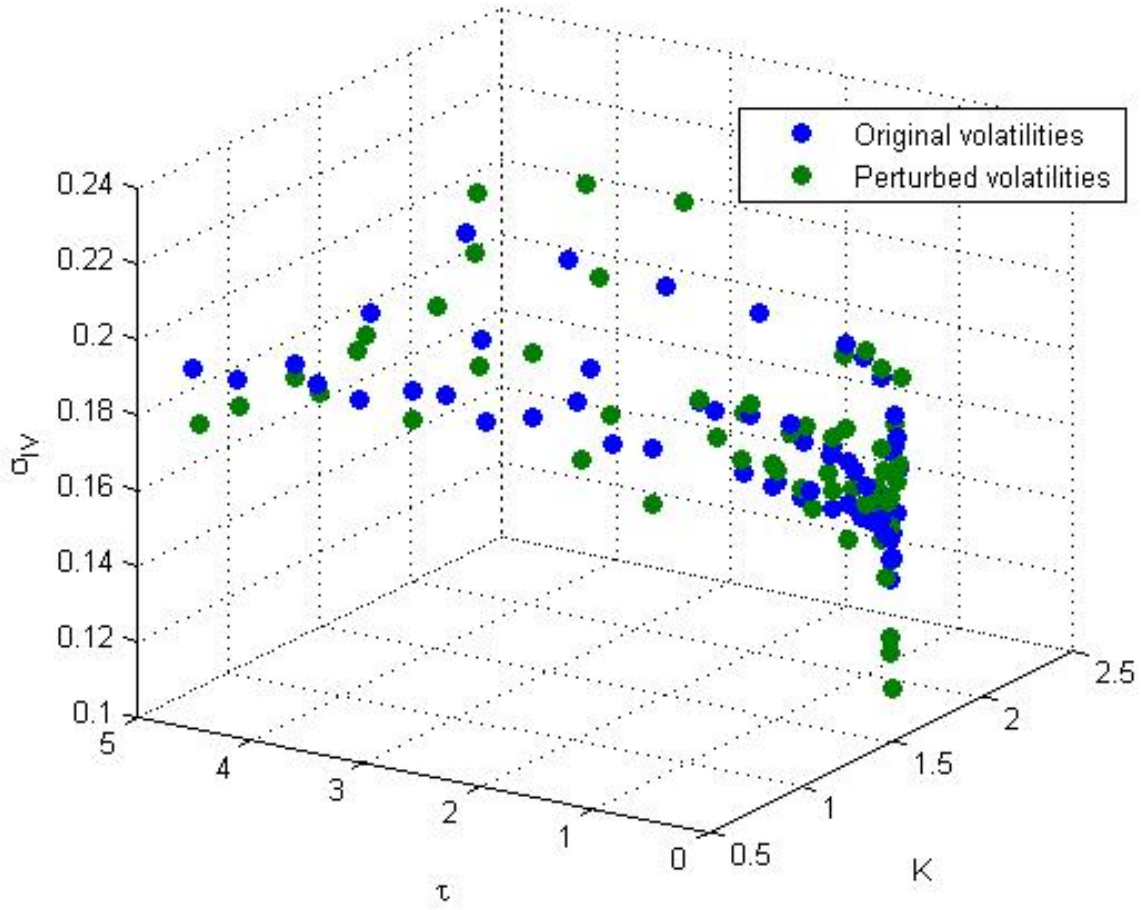

Figure 7: Example of a perturbed FX volatility surface plotted along its original surface based on the moderate smile parametrization of the Bates model given in Section 5.2. The value of the underlying is set to 1.5.

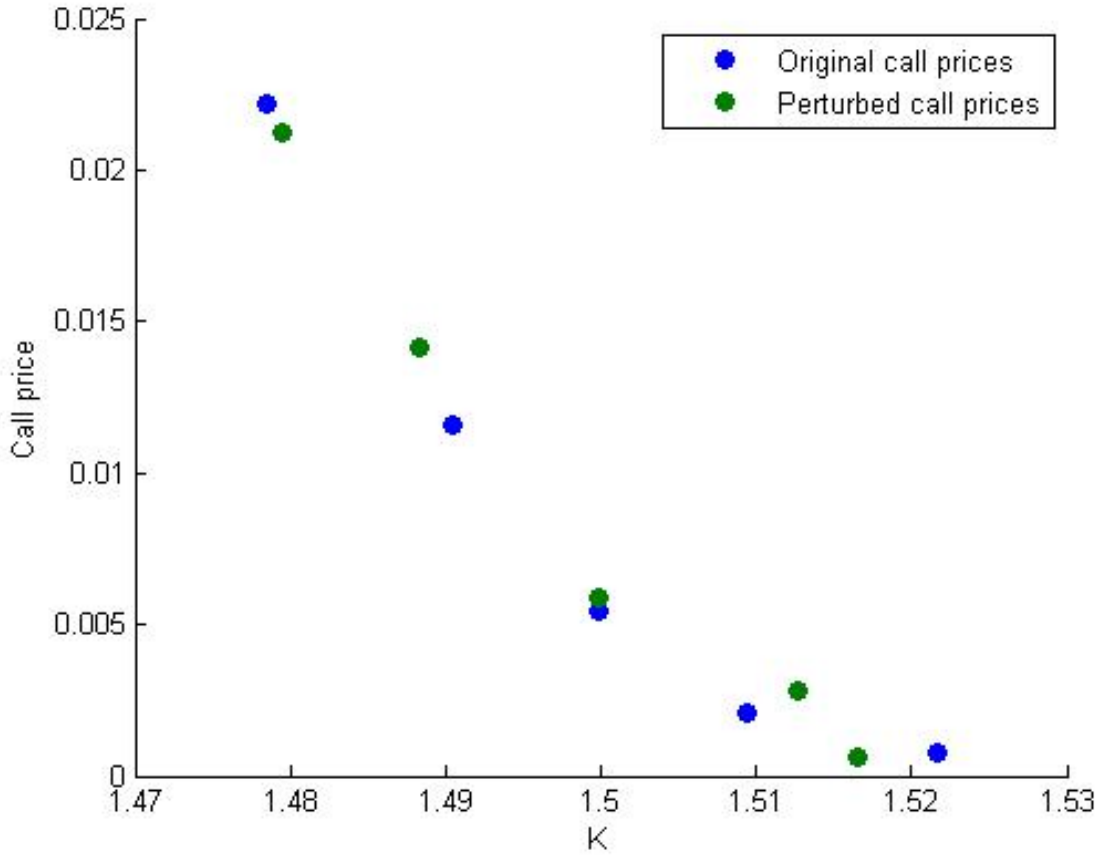

Figure 8: Original and perturbed option prices with a time to expiry of one day and the value of the underlying set to 1.5, simulated using the moderate smile parametrization of the Bates model and remaining parameter choices from Section 5.2. The perturbed prices are arbitrageable, as the call prices are not in convex order.

5 Simulation study

In order to ascertain which of the arbitrage-free methods is *optimal* for constructing the FX volatility surface, we simulate FX option price data in the form of volatility surfaces using the method proposed in Section 4, and use the simulated prices to calibrate the models discussed in Section 3. We use the Bates model (4) for the risk-neutral dynamics of the underlying, as it is capable of replicating nearly all characteristics of the volatility surface seen in the FX market. In particular, the inclusion of a jump term allows the model to create pronounced smile effects even for short time to expiries, which is not possible when using regular stochastic volatility models.
We use four different parameter choices of the Bates model for which we simulate the FX volatility surface 100 times and report the average mean absolute error (MAE) and maximum absolute deviation (MAD; in our case, the maximum of the absolute errors) for the call prices and implied volatilities to measure overall fit and worst cases respectively, as well as their standard deviations (in parentheses). The statistics that measure the fit in terms of call prices are complementary to the statistics concerned with implied volatilities, because it does not need to be the case that a method dominates in both quantities. The reason is that the sensitivity of the call price to changes in the volatilities, which can be approximated by an option's vega, depends on the moneyness of the option: ATM options are most sensitive to changes in the volatility, followed by ITM options and OTM options are least sensitive to volatility changes. In light of the axiom of no-arbitrage, it is important when quoting prices to remain within the bid and ask prices given by the market; as the market quotes options in terms of implied volatilities, we will assign more weight to fitting the implied volatilities than to fitting the call prices in determining which method is optimal.
In addition to the MAE and MAD statistics, we report the average calibration time in seconds required to calibrate the volatility surface and compute all requested prices. We note that the calibration time is dependent on both the hardware and software used, as well as the implementation. In particular, there are other programming languages that are (much) faster than `Matlab`. For this reason, we emphasize that the calibration time results merely show potential in terms of speed of the corresponding model; if a model is found to be fast, then it could be made even faster.
As a benchmark, we include the original (unperturbed) prices from the Bates model. For the arbitrage-free SVI method, we include the results for using the Heston-like SSVI parameterization, as well as the Power-law SSVI parameterization. We note also, that as constrained MLS is a local approximation method, the true calibration time it takes to compute one call price or implied volatility of interest is N times as small as the reported average calibration time when there are N points in the quoted surface.
We now first discuss how unperturbed option prices are computed using the Bates model, followed by a justification for the chosen parameters and the resulting unperturbed data sets. Lastly, we report the results of the simulation study.

5.1 Computing option prices with the Bates model

We compute prices of European options with the Bates model by using its corresponding characteristic function. Bakshi and Madan (2000) give the following formula for the price

of European options when the characteristic function of the underlying is known:

$$C(K,\tau) = e^{-q\tau} S_t \Pi_1 - e^{-r\tau} K \Pi_2,$$

where S_t denotes the value of the underlying at the start of the contract, r and q denote the constant risk-free interest rate and dividend yield (in the FX case, the foreign interest rate) respectively and the values of $\Pi_j, j = 1, 2$, depend on the characteristic function $\phi(\omega)$:

$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left(\frac{e^{-i\omega \ln(K)} \phi(\omega - i)}{i\omega \phi(-i)} \right) d\omega, \tag{34}$$

$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left(\frac{e^{-i\omega \ln(K)} \phi(\omega)}{i\omega} \right) d\omega. \tag{35}$$

Here, $\mathrm{Re}(\cdot)$ denotes the real part of a complex number, and the characteristic function of the Bates model is given by (as in Gilli and Schumann (2011)):

$$\begin{aligned}
\phi(\omega) &= e^{A+B+C+D}, \\
A &= \mathrm{i}\omega \ln(S_0) + \mathrm{i}\omega r \tau, \\
B &= \frac{\theta\kappa}{\sigma^2} \left((\kappa - \rho\sigma\mathrm{i}\omega - d)\tau - 2\ln\left(\frac{1 - g e^{-d\tau}}{1 - g} \right) \right), \\
C &= \frac{\frac{v_0}{\sigma^2}(\kappa - \rho\sigma\mathrm{i}\omega - d)(1 - e^{-d\tau})}{1 - g e^{-d\tau}}, \\
D &= -\lambda\mu_J \mathrm{i}\omega\tau + \lambda\tau \left((1 + \mu_J)^{\mathrm{i}\omega} \exp\left[\frac{1}{2}\sigma_J^2 \mathrm{i}\omega(\mathrm{i}\omega - 1) \right] - 1 \right), \\
d &= \sqrt{(\rho\sigma\mathrm{i}\omega - k)^2 + \sigma^2(\mathrm{i}\omega + \omega^2)}, \\
g &= \frac{\kappa - \rho\sigma\mathrm{i}\omega - d}{\kappa - \rho\sigma\mathrm{i}\omega + d}.
\end{aligned}$$

We use numerical integration, `Matlab`'s `integral` function to be specific, which is an adaptive version of Simpson's rule, to determine the values of the integrals (34) and (35). As numerical integration methods cannot deal with infinite bounds, we choose to set the cutoff point at 200 instead, in line with Gilli and Schumann (2011). The justification is that the integrals (34) and (35) decay rapidly to zero.

5.2 Parameter choices and the generated data

We generate European option prices using four different parameter choices for the Bates model, each representing a different pattern of implied volatility: a moderate smile, a strongly pronounced smile, a pattern corresponding to a right-skewed risk-neutral distribution and a pattern corresponding to a left-skewed risk-neutral distribution. The parameter values are given in Table 2. For the market parameters, the underlying exchange rate S_t is set to 1.50, the domestic risk-free short rate r^d is set to zero to mimic current low interest rates from a European point of view, and the foreign risk-free short

rate r^f is set to 0.02. For the simulation parameters, we choose a bid-ask spread of $-0.01/0.01$ implied volatility and a time to expiry multiplication factor $c = 2$ for the time to expiries shorter than 1M and longer than 1Y. Such a bid-ask spread might be regarded as high for liquid quotes on some currency options, but is justified by noting that we use it to model all market noise, not just the noise stemming from bid-ask spreads. We experimented with different bid-ask spread sizes ($-0.005/0.005, -0.02/0.02$ and $-0.03/0.03$) to see whether this would cause a change in which method performed best, but this was not the case.
The moderate smile parameter choice can be seen as the basic parameter choice; the parameter values corresponding to the other implied volatility patterns have been chosen by adjusting the moderate smile parameter values. Incorporation of left- and right-skewness of the risk-neutral distribution in the volatility smile is done by setting the correlation parameter ρ equal to -0.80 and 0.80 respectively. The initial volatility $\sqrt{v_0}$ and the long-run volatility $\sqrt{\theta}$ are both set to 0.15 such that the prices correspond to implied volatility sizes frequently seen in FX. The volatility of the volatility parameter σ is set to 0.30, to ensure that the stochastic volatility plays a significant role in the determination of the prices. The mean reversion parameter κ is set to 1.50, which corresponds to moderate a half-life time of $\ln(2)/1.50 = 0.46$, i.e. it takes approximately half a year for the process to return to the mean. The jump-intensity parameter λ is set to 0.10, corresponding to a 10% probability of a jump per year. The mean and volatility of the random jump size are set to 0 and 0.10 respectively. For the strongly pronounced smile, we set σ equal to 0.70, and λ and σ_J both equal to 0.30. The generated (unperturbed) volatility surfaces are shown in Tables 3 - 6.

Parameter choice	$\sqrt{v_0}$	$\sqrt{\theta}$	ρ	κ	σ	λ	μ_J	σ_J
Moderate smile	0.15	0.15	0.00	1.50	0.30	0.10	0.00	0.10
Strongly pronounced smile	0.15	0.15	0.00	1.50	0.70	0.30	0.00	0.30
Left-skewed	0.15	0.15	-0.80	1.50	0.30	0.10	0.00	0.10
Right-skewed	0.15	0.15	0.80	1.50	0.30	0.10	0.00	0.10

Table 2: Parameter choices for the Bates model in the simulation study. The market parameters are: $S_t = 1.50, r^d = 0.00$ and $r^f = 0.02$.

τ	-10Δ	-25Δ	ATM	25Δ	10Δ
1D	0.1775	0.1486	0.1431	0.1486	0.1775
1W	0.1596	0.1549	0.1536	0.1549	0.1596
1M	0.1694	0.1589	0.156	0.1589	0.1694
2M	0.1761	0.1614	0.1572	0.1614	0.1761
3M	0.1804	0.1631	0.1578	0.1631	0.1804
6M	0.1869	0.1663	0.1593	0.1663	0.1869
9M	0.1893	0.1683	0.1607	0.1683	0.1893
1Y	0.1900	0.1697	0.1621	0.1697	0.1900
2Y	0.1889	0.1730	0.1666	0.1730	0.1889
3Y	0.1871	0.1747	0.1696	0.1747	0.1871
4Y	0.1858	0.1758	0.1715	0.1758	0.1858
5Y	0.1849	0.1764	0.1729	0.1764	0.1849

Table 3: Generated FX volatility surface, corresponding to the moderate smile parameter choice for the Bates model. The market parameters are: $S_t = 1.50, r^d = 0.00$ and $r^f = 0.02$.

τ	-10Δ	-25Δ	ATM	25Δ	10Δ
1D	0.2099	0.1596	0.1506	0.1596	0.2099
1W	0.2189	0.1773	0.1690	0.1773	0.2189
1M	0.3137	0.2065	0.1833	0.2065	0.3137
2M	0.3618	0.2296	0.1929	0.2296	0.3618
3M	0.3822	0.2458	0.2004	0.2458	0.3822
6M	0.3978	0.2753	0.2185	0.2753	0.3978
9M	0.3960	0.2913	0.2327	0.2913	0.3960
1Y	0.3909	0.3010	0.2443	0.3010	0.3909
72Y	0.3720	0.3171	0.2743	0.3171	0.3720
3Y	0.3608	0.3219	0.2908	0.3219	0.3608
4Y	0.3541	0.3240	0.3008	0.3240	0.3541
5Y	0.3498	0.3251	0.3073	0.3251	0.3498

Table 4: Generated FX volatility surface, corresponding to the strongly pronounced smile parameter choice for the Bates model. The market parameters are: $S_t = 1.50, r^d = 0.00$ and $r^f = 0.02$.

τ	-10Δ	-25Δ	ATM	25Δ	10Δ
1D	0.1853	0.1532	0.1431	0.1443	0.1676
1W	0.1702	0.1604	0.1536	0.1489	0.1480
1M	0.1890	0.1698	0.1559	0.1462	0.1442
2M	0.2008	0.1758	0.1568	0.1435	0.1417
3M	0.2076	0.1794	0.1570	0.1416	0.1411
6M	0.2167	0.1845	0.1569	0.1391	0.1429
9M	0.2192	0.1864	0.1572	0.1391	0.1455
1Y	0.2194	0.1870	0.1577	0.1401	0.1475
2Y	0.2147	0.1868	0.1609	0.1454	0.1518
3Y	0.2100	0.1860	0.1635	0.1498	0.1538
4Y	0.2065	0.1853	0.1655	0.1531	0.1551
5Y	0.2038	0.1847	0.1669	0.1555	0.1561

Table 5: Generated FX volatility surface, corresponding to the left-skewed parameter choice for the Bates model. The market parameters are: $S_t = 1.50, r^d = 0.00$ and $r^f = 0.02$.

τ	-10Δ	-25Δ	ATM	25Δ	10Δ
1D	0.1676	0.1443	0.1431	0.1532	0.1853
1W	0.1482	0.1491	0.1538	0.1606	0.1703
1M	0.1448	0.1469	0.1566	0.1706	0.1898
2M	0.1428	0.1448	0.1581	0.1772	0.2023
3M	0.1425	0.1433	0.1589	0.1814	0.2097
6M	0.1449	0.1418	0.1602	0.1883	0.2206
9M	0.1478	0.1425	0.1614	0.1914	0.2246
1Y	0.1500	0.1439	0.1627	0.1931	0.2260
2Y	0.1549	0.1503	0.1674	0.1951	0.2245
3Y	0.1574	0.1553	0.1708	0.1953	0.2213
4Y	0.1592	0.1589	0.1732	0.1951	0.2186
5Y	0.1606	0.1617	0.1750	0.1949	0.2163

Table 6: Generated FX volatility surface, corresponding to the right-skewed parameter choice for the Bates model. The market parameters are: $S_t = 1.50, r^d = 0.00$ and $r^f = 0.02$.

5.3 Results

Tables 7 - 10 show the results of the simulation study. We comment on the MAE and MAD statistics for each parameter choice separately, whereas we comment on the calibration times altogether as they are similar for all parameter choices.
Table 7 shows the simulation results for the moderate smile parameter choice of the Bates model. Both SVI methods outperform the other methods, including the benchmark, in terms of MAE and MAD for both call prices and volatilities. The SABR model and the lognormal mixture model are close to the SVI models in terms of MAE and MAD values for both call prices and volatilities; the constrained MLS method performs noticeably worse than the other methods. All estimated quantities have a low standard deviation, which implies that it is unlikely that the results depend much on the particular simulation

run.
Table 8 shows the results for the strongly pronounced smile parameter choice. Here, the Power-law version of the SVI model has the lowest MAE and MAE estimates, also dominating the benchmark. The lognormal mixture model and the Heston-like version of the SVI model follow closely with similar MAE and MAD values. The SABR model and the constrained MLS method appear to be (relatively) troubled by the strongly pronounced smile data set, as they are outperformed in terms of volatilities and call prices. The standard deviations are small again.
Table 9 shows the results for the left-skewed parameter choice for the Bates model. Both SVI models outperform the other models with respect to the MAE statistics for the call prices and implied volatilities, where the lognormal mixture model has the lowest MAD values. Similar results are shown in Table 10, which corresponds to the right-skewed parameter choice. Here, the SVI models outperform the other models with respect to the MAE and MAD values for call prices and implied volatilities. The SABR model, the lognormal mixture model and the constrained MLS method all have MAE statistics for the call prices close to the ones of the SVI models. In terms of volatilities however, the SVI models clearly perform better.
The calibration times are similar for all parameter choices of the Bates model: the constrained MLS method is fastest, followed by the SABR model. The SVI models are approximately ten times as slow as the constrained MLS method. The long calibration time of the lognormal mixture model is caused by the use of global optimization for its 11 parameters and was expected, but the comprehensive calibration procedure does not result in the best fit among models.
Overall, the SVI model appears to be the optimal choice based on the simulated data, with the Power-law version slightly outperforming the Heston-like version. The difference between the two variants of the SVI model is largest for data generated with the strongly pronounced smile parameter choice of the Bates model; in other cases, the performance is similar. The constrained MLS method and the SABR model are outperformed by the SVI models with respect to the fitting of the data, which is somewhat compensated by their faster calibration times. Despite its comprehensive calibration procedure, the lognormal mixture model does not appear to be optimal for any parameter choice.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
Bates (benchmark)	0.008 (0.001)	0.073 (0.017)	0.011 (0.001)	0.041 (0.006)	-
SABR	0.004 (0.001)	0.029 (0.009)	0.012 (0.001)	0.042 (0.007)	0.433
Lognormal mixture	0.004 (0.001)	0.027 (0.007)	0.010 (0.003)	0.045 (0.018)	17.695
Constrained MLS	0.007 (0.001)	0.039 (0.008)	0.034 (0.003)	0.154 (0.014)	0.215
SVI (Heston-like)	0.002 (0.001)	0.024 (0.013)	0.005 (0.001)	0.032 (0.012)	2.040
SVI (Power-law)	0.002 (0.001)	0.024 (0.008)	0.005 (0.001)	0.030 (0.010)	2.128

Table 7: Simulation study results for the moderate smile parameter choice using 100 iterations. The average MAE and MAD for the call prices (C) and the volatilities (V) are given, followed by their standard deviations within parenthesis. In addition, the average calibration time over the iterations in seconds is provided.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
Bates (benchmark)	0.007 (0.001)	0.059 (0.016)	0.012 (0.001)	0.047 (0.008)	-
SABR	0.013 (0.001)	0.098 (0.028)	0.052 (0.002)	0.133 (0.008)	0.358
Lognormal mixture	0.006 (0.002)	0.033 (0.009)	0.020 (0.008)	0.080 (0.029)	17.738
Constrained MLS	0.017 (0.001)	0.077 (0.012)	0.094 (0.004)	0.307 (0.007)	0.216
SVI (Heston-like)	0.006 (0.001)	0.034 (0.008)	0.027 (0.002)	0.135 (0.012)	2.070
SVI (Power-law)	0.003 (0.001)	0.032 (0.019)	0.008 (0.002)	0.054 (0.022)	2.229

Table 8: Simulation study results for the strongly pronounced smile parameter choice using 100 iterations. The average MAE and MAD for the call prices (C) and the volatilities (V) are given, followed by their standard deviations within parenthesis. In addition, the average calibration time over the iterations in seconds is provided.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
Bates (benchmark)	0.008 (0.001)	0.071 (0.017)	0.011 (0.001)	0.042 (0.007)	-
SABR	0.005 (0.001)	0.033 (0.008)	0.015 (0.001)	0.058 (0.011)	0.425
Lognormal mixture	0.006 (0.001)	0.032 (0.008)	0.017 (0.003)	0.057 (0.017)	18.849
Cosntrained MLS	0.006 (0.001)	0.035 (0.006)	0.030 (0.002)	0.149 (0.011)	0.205
SVI (Heston-like)	0.004 (0.002)	0.042 (0.050)	0.009 (0.002)	0.058 (0.043)	2.082
SVI (Power-law)	0.004 (0.004)	0.054 (0.077)	0.007 (0.004)	0.062 (0.069)	2.200

Table 9: Simulation study results for the left-skewed parameter choice using 100 iterations. The average MAE and MAD for the call prices (C) and the volatilities (V) are given, followed by their standard deviations within parenthesis. In addition, the average calibration time over the iterations in seconds is provided.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
Bates (benchmark)	0.008 (0.001)	0.082 (0.024)	0.011 (0.001)	0.043 (0.008)	-
SABR	0.006 (0.001)	0.049 (0.012)	0.016 (0.001)	0.062 (0.010)	0.383
Lognormal mixture	0.005 (0.001)	0.028 (0.009)	0.014 (0.002)	0.053 (0.023)	17.498
Constrained MLS	0.008 (0.001)	0.043 (0.008)	0.042 (0.003)	0.170 (0.021)	0.202
SVI (Heston-like)	0.004 (0.001)	0.031 (0.009)	0.009 (0.001)	0.050 (0.014)	2.087
SVI (Power-law)	0.003 (0.001)	0.031 (0.023)	0.006 (0.002)	0.041 (0.026)	2.183

Table 10: Simulation study results for the right-skewed parameter choice using 100 iterations. The average MAE and MAD for the call prices (C) and the volatilities (V) are given, followed by their standard deviations within parenthesis. In addition, the average calibration time over the iterations in seconds is provided.

6 Application to real FX option price data

We apply all models on the EUR/USD option price data set shown in Figure 2 and the USD/JPY data set shown in Figure 9. The domestic and foreign risk-free rates were set to the overnight rate values of $r^d = 0.3\%$ and $r^f = -0.3\%$ for the EUR/USD options and $r^d = -0.1\%$ and $r^f = 0.3\%$ for the USD/JPY options.

Table 11 shows the results for the EUR/USD data set. The Heston-like version of the SVI model has the lowest MAE as well as the lowest MAD with respect to the implied volatilities. The Power-law version of the SVI model outperforms the other methods in MAE with respect to the call prices, although the results are close, and ties the optimal MAD value with constrained MLS. Constrained MLS is optimal in terms of calibration time followed by the SABR model, and the SVI models are approximately ten times as slow as the constrained MLS method. The lognormal mixture model has by far the longest calibration time, as was expected.

Figures 10 - 14 show the resulting call price surfaces and volatility surfaces combined with the market prices.[9] The flexible shapes of the volatility surfaces constructed by the SVI models and the constrained MLS method stand out; the SVI model fits separate slices of implied volatility for each time to expiry to the data, whereas the constrained MLS method fits a different function to each point of interest, which causes the resulting surfaces to be more flexible than those generated by the SABR model and the lognormal mixtures model. The constrained MLS method appears to result in a non-smooth surface, but this non-smoothness disappears when plotting a denser grid of calibrated points. We note also, that calibrating the lognormal mixture model multiple times has led to volatility surfaces with very different shapes; an unwanted property that is caused by non-convexity of the criterion function combined with the large number of parameters.

Table 12 shows the results for the USD/JPY data set. The Heston-like variant of the SVI model again outperforms the other methods with respect to the volatility related statistics, although this time, the MAE and MAD statistics for the call prices are considerably worse. Inspection of the model prices compared to the market prices and volatilities shows that the bad fit is caused by the calibration to the last two time to expiries of four and five years respectively. The SVI Power-law model suffers from the same problems, which causes it to perform worst of all methods in terms of MAE with respect to both call prices as well as volatilities. An explanation for these poor fits to the longer time to expiries is given in Figure 15, which shows that the data is arbitrageable as the call prices are not decreasing in strike. Even after correcting for the constant interest rates assumption by using the actual discount factors, the same violation of arbitrage occurs, although smaller in magnitude. The other models do not appear to have any trouble fitting the longer time to expiries in particular, which could imply that the ability of the SVI model to fit the data is relatively more sensitive to the quality of the data than for other methods. The constrained MLS method performs best in terms of call prices, followed by the lognormal mixture model and the SABR model. The calibration times are similar to those found in the results for the EUR/USD data set.

Figures 16 - 20 show the call price surfaces and volatility surfaces together with the market prices. Particularly remarkable here, is that the slices of implied volatility from the SVI models corresponding to the longer time to expiries show little curvature, whereas

[9] For SVI, linear interpolation between the time to expiry slices is applied to plot surfaces.

the corresponding data shows a pronounced volatility smile, as is shown in Figure 9. This observation coincides with the analysis that SVI is unable to fit the longer time to expiries adequately in the USD/JPY data set.
Overall, the Heston-like variant of the SVI model appears to be the optimal choice for these data sets, since we have chosen to assign more weight to the fitting of implied volatilities than to the fitting of call prices. Noticeably, none of the SVI models were best at fitting the time to expiries of four and five years in the USD/JPY data set, although the fact that this particular data was arbitrageable provides some clemency.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
SABR	0.002	0.009	0.009	0.032	0.428
Lognormal mixture	0.002	0.009	0.013	0.115	18.299
Constrained MLS	0.001	0.005	0.013	0.340	0.222
SVI (Heston-like)	0.001	0.012	0.002	0.014	2.210
SVI (Power-law)	0.000	0.005	0.004	0.051	2.257

Table 11: Estimation results for EUR/USD European option price data as of 15-06-2016, with a spot exchange rate equal to 1.1309. The MAE and MAD for the call prices (C) and the volatilities (V) are given. In addition, the calibration time in seconds is provided.

Method	MAE (C)	MAD (C)	MAE (V)	MAD (V)	Time (s)
SABR	0.388	2.228	0.018	0.098	0.349
Lognormal mixture	0.357	3.601	0.011	0.054	14.782
Constrained MLS	0.163	0.868	0.023	0.269	0.208
SVI (Heston-like)	0.443	4.765	0.006	0.053	2.240
SVI (Power-law)	0.453	3.820	0.044	0.215	2.300

Table 12: Estimation results for USD/JPY European option price data as of 15-06-2016, with a spot exchange rate equal to 102.8. The MAE and MAD for the call prices (C) and the volatilities (V) are given. In addition, the calibration time in seconds is provided.

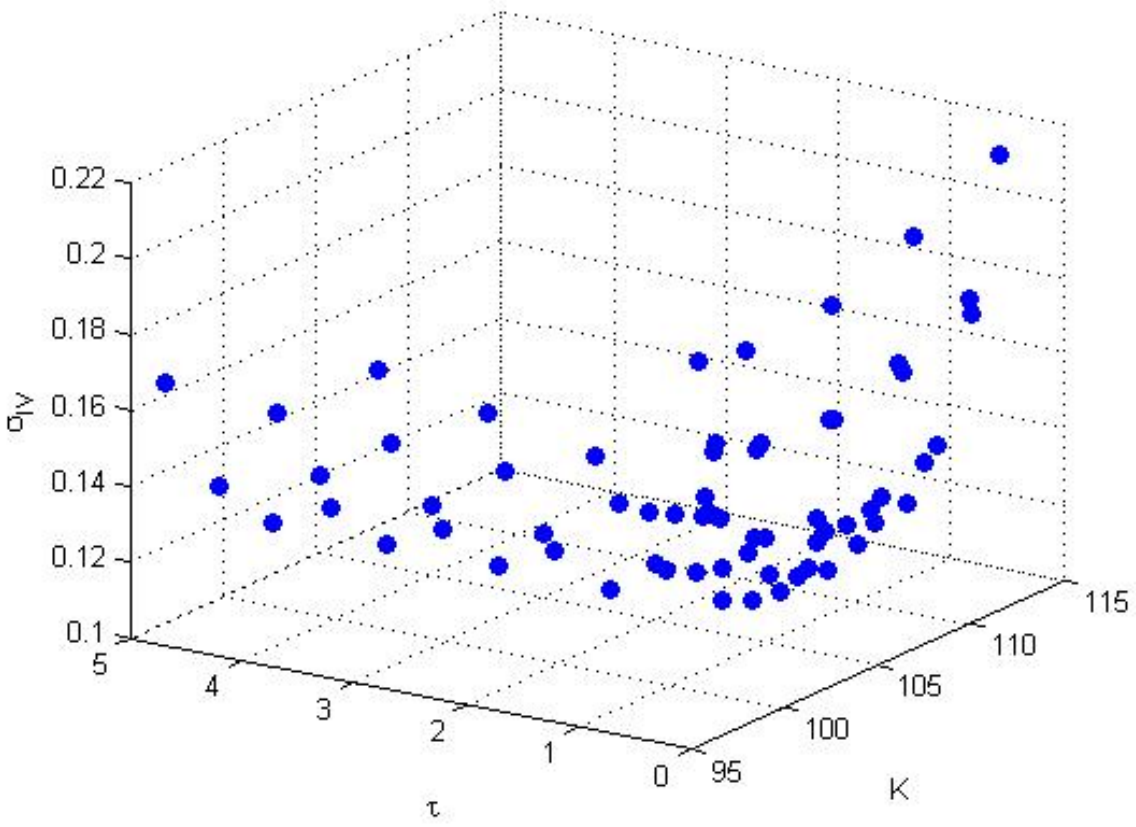

Figure 9: Implied volatilities for the USD/JPY data set with time to expiries ranging from one day until five years as of 16-09-2016, with a spot exchange rate equal to 102.8.

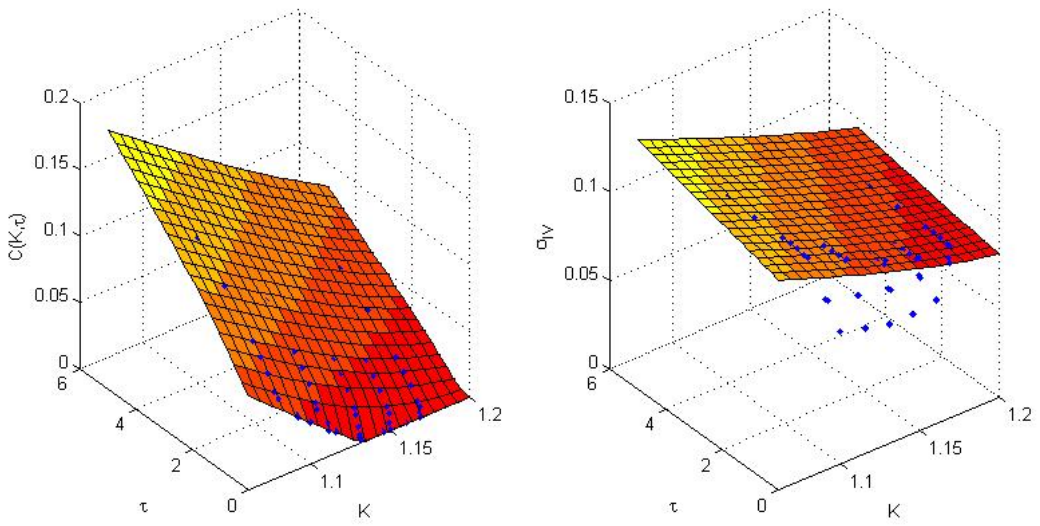

Figure 10: Call price surface (left) and volatility surface (right) for the SABR model applied on the EUR/USD data set.

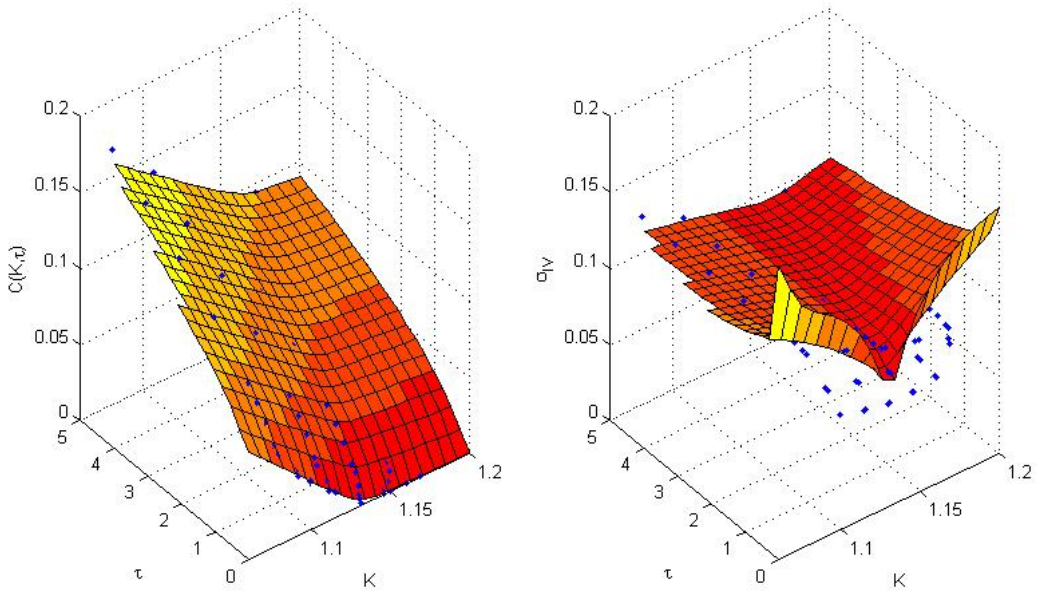

Figure 11: Call price surface (left) and volatility surface (right) for the lognormal mixture model applied on the EUR/USD data set.

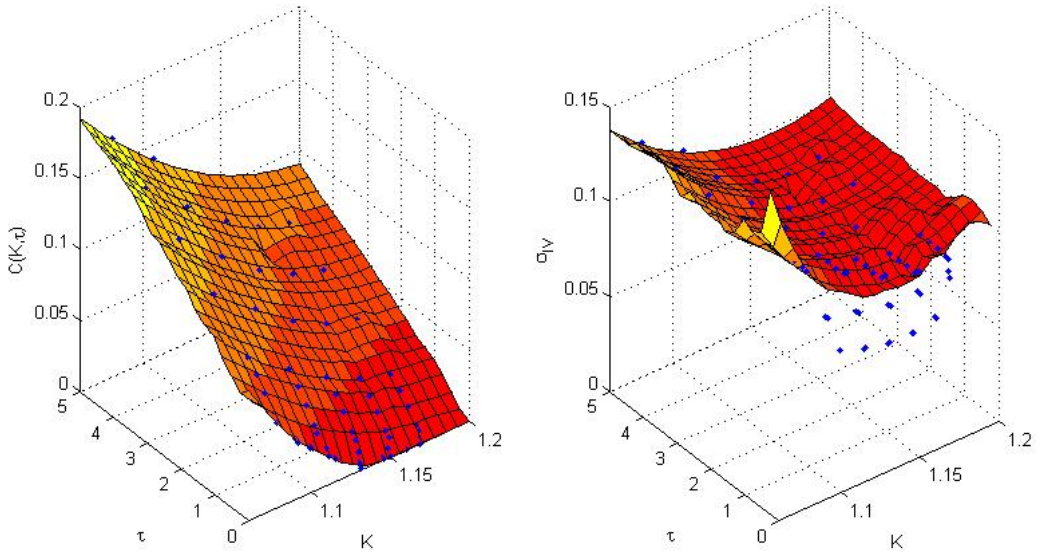

Figure 12: Call price surface (left) and volatility surface (right) for the constrained MLS method applied on the EUR/USD data set.

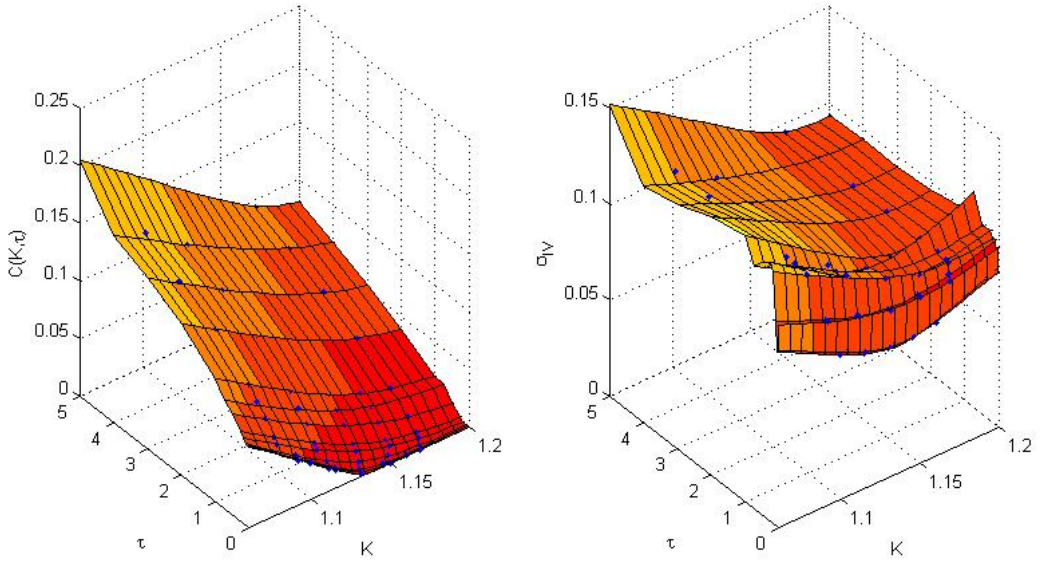

Figure 13: Call price surface (left) and volatility surface (right) for the Heston-like variant of the SVI model applied on the EUR/USD data set.

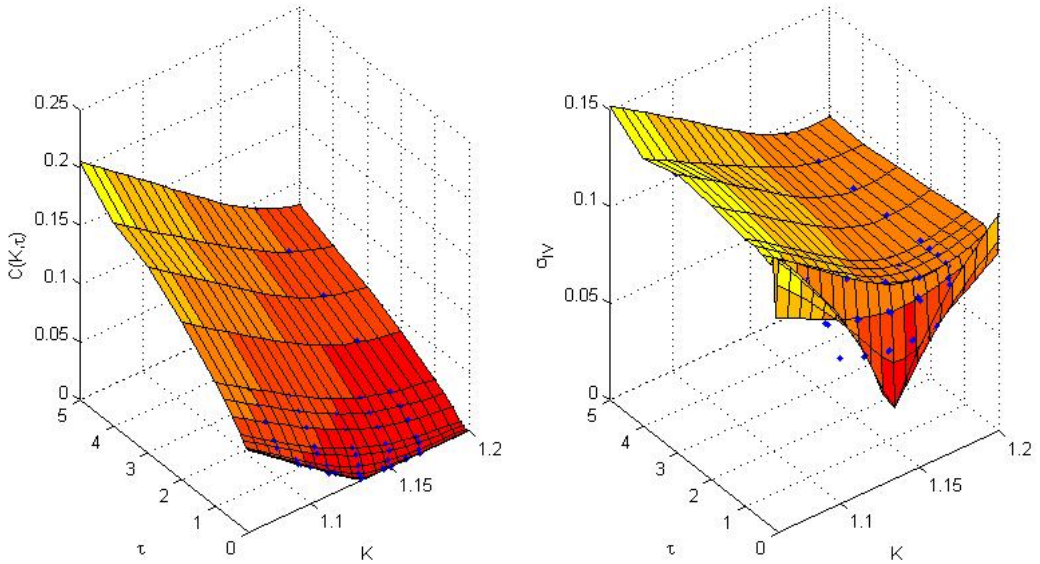

Figure 14: Call price surface (left) and volatility surface (right) for the Power-law variant of the SVI model applied on the EUR/USD data set.

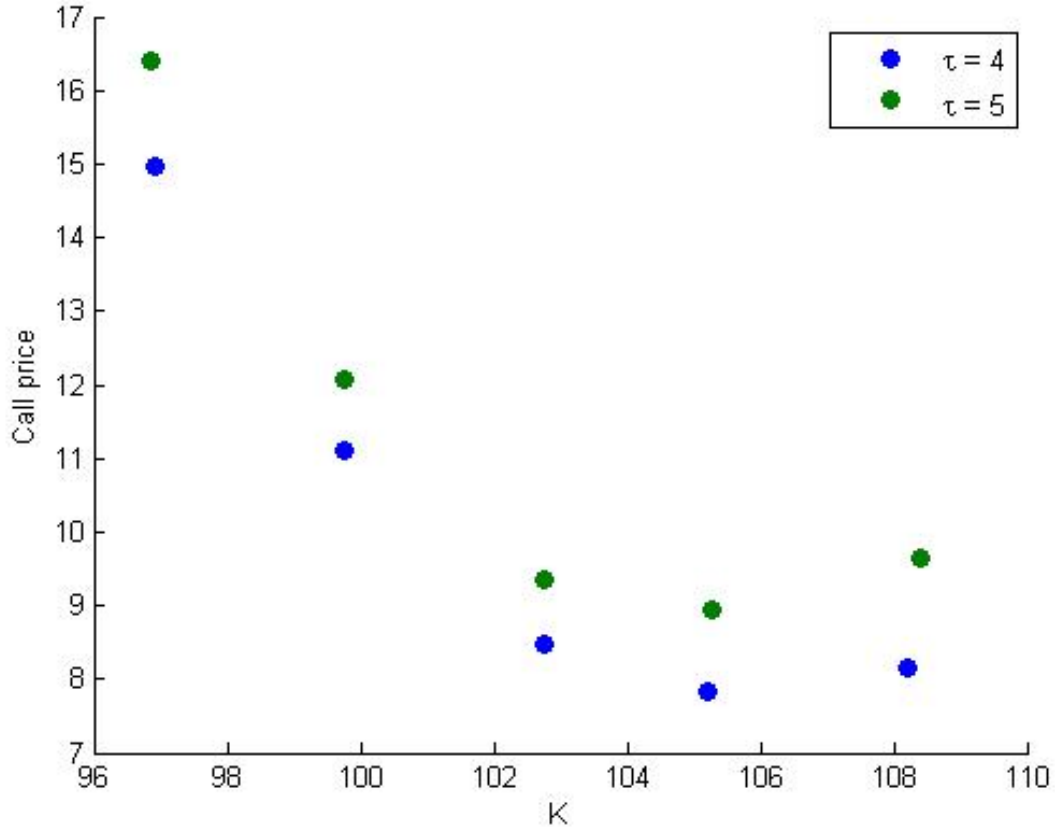

Figure 15: Call prices of the USD/JPY data set for time to expiries of four and five years. The data is arbitrageable, as the call prices are not decreasing in strike.

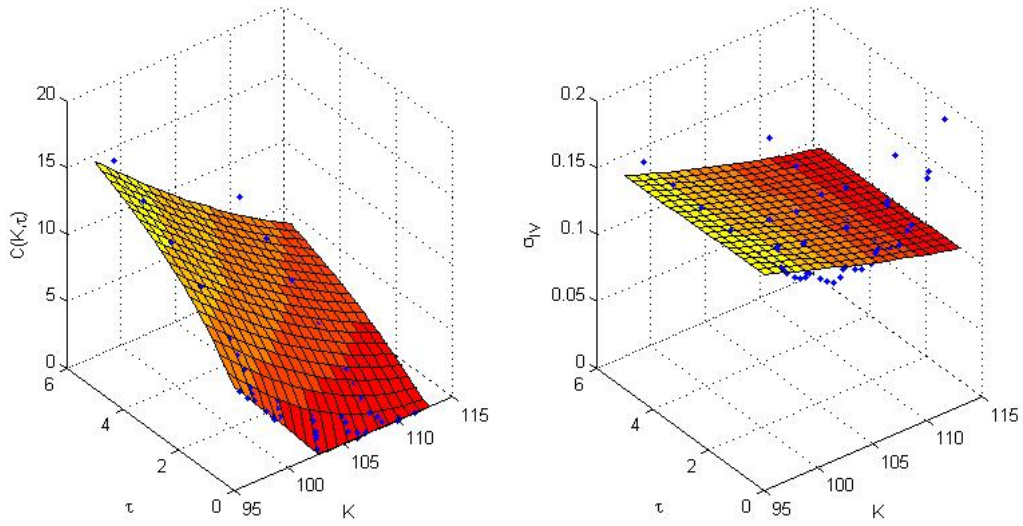

Figure 16: Call price surface (left) and volatility surface (right) for the SABR model applied on the USD/JPY data set.

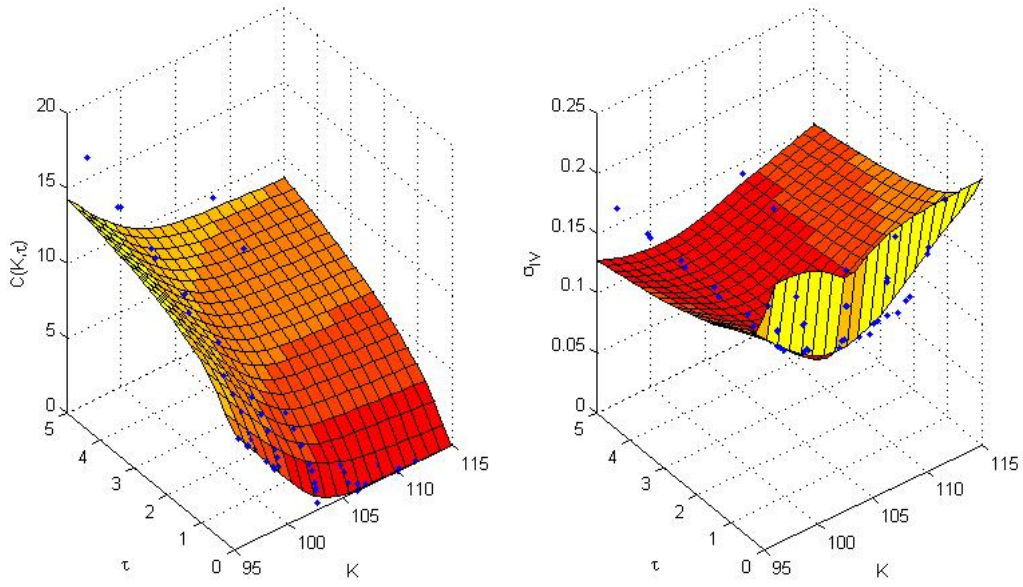

Figure 17: Call price surface (left) and volatility surface (right) for the lognormal mixture model applied on the USD/JPY data set.

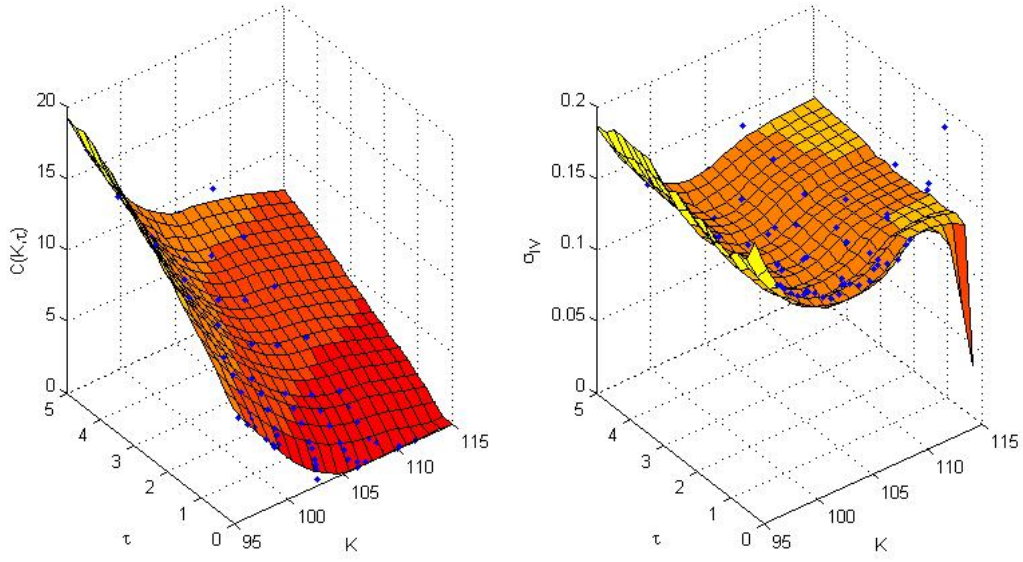

Figure 18: Call price surface (left) and volatility surface (right) for the constrained MLS method applied on the USD/JPY data set.

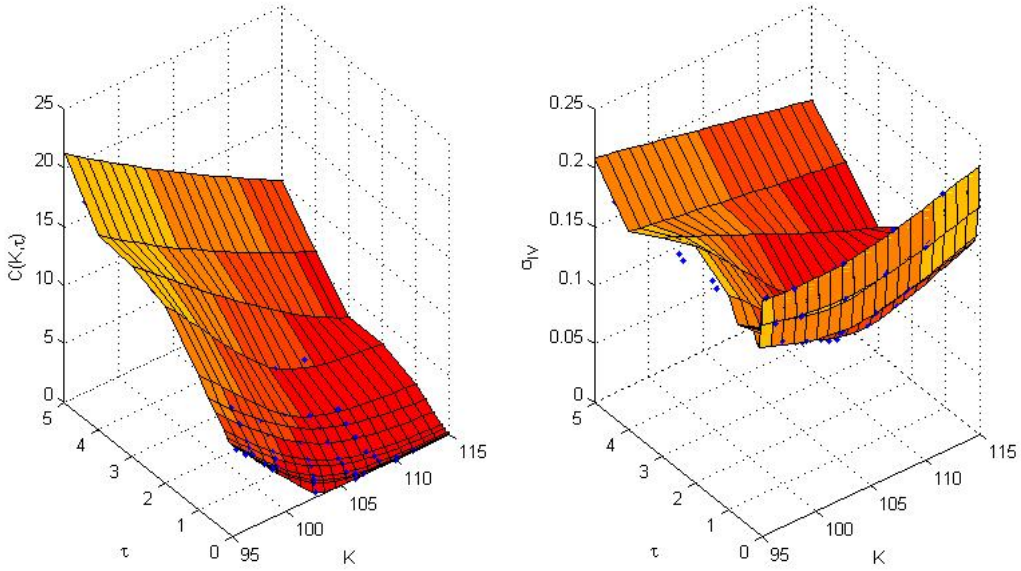

Figure 19: Call price surface (left) and volatility surface (right) for the Heston-like variant of the SVI model applied on the USD/JPY data set.

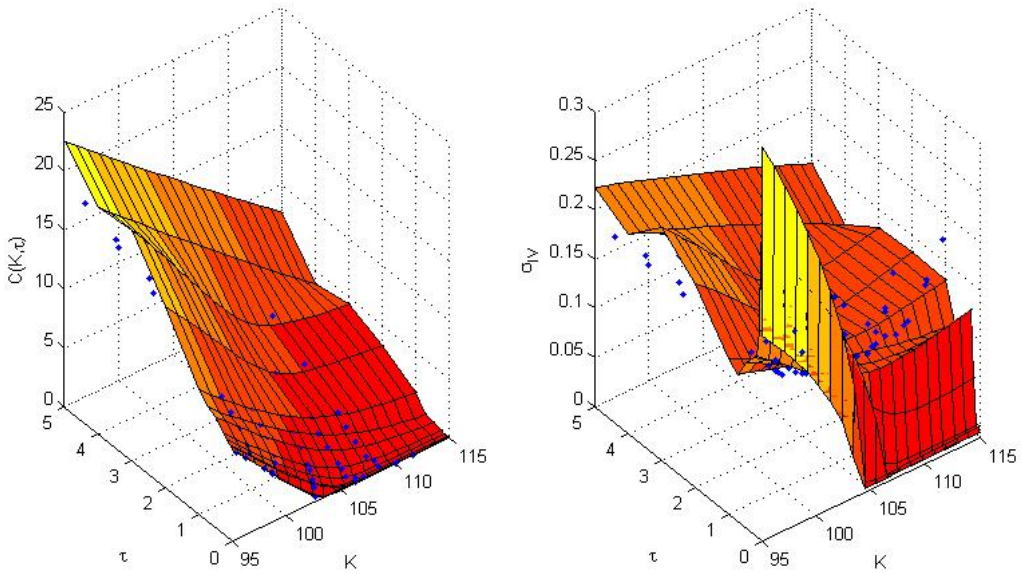

Figure 20: Call price surface (left) and volatility surface (right) for the Power-law variant of the SVI model applied on the USD/JPY data set.

7 Conclusion and future work

We have proposed a method to simulate realistic FX options data, and used it to determine which of several methods is optimal to construct the arbitrage-free FX volatility surface. Based on results from the simulation study, we find that both versions of the SVI model are better able to fit the data than the SABR model, the lognormal mixture model and the constrained MLS method. The Power-law version of the SVI model appears to outperform the Heston-like version, which is most clear on the strongly pronounced smile data; on the other data sets the performance is similar. In the case of the SABR model and the constrained MLS method, the inferior fit compared to the SVI models is compensated somewhat by their superior calibration times. If only a particular value in the surface is of interest, the calibration time of the constrained MLS method compared to the other models improves additionally, as it is a local approximation method. Despite its comprehensive calibration procedure, the lognormal mixture model does not appear to be optimal for any parameter choice.

We applied all models on real FX options data, and found that the Heston-like version of the SVI model outperforms the other models when it comes to fitting the implied volatilities. The performance of both versions of the SVI model is similar for the EUR/USD data set. For the USD/JPY data set, however, the Power-law version of the SVI model performs worst of all methods in terms of fitting implied volatilities and call prices. There, the Heston-like version is optimal with respect to fitting the volatilities, whereas the constrained MLS method outperforms the other methods when considering the call prices. Specifically on the USD/JPY data set both SVI models appear to have trouble fitting the longer time to expiries, which can be explained by the fact that those specific prices were found to be arbitrageable.

Overall, we conclude that the SVI model is optimal for constructing the arbitrage-free volatility surface for FX options. We note hereby that the Power-like version appears to outperform the Heston-like version on strongly pronounced smile data, whereas application to the USD/JPY option price data set implies the opposite. In addition, the constrained MLS method allows for the fastest calibration, being at least ten times faster than the SVI model.

Several matters deserve further investigation. First, with respect to the comparison study, it would be interesting to include the SABR model calibrated to each time to expiry separately. This would require the use of penalization to rule out calendar spread arbitrage, similar to what is done in the SVI method. In addition, a future comparison study could include the constrained cubic spline method of Fengler (2009) appended with some chosen extrapolation method, as well as the tensor product B-spline method of Fengler and Hin (2015). Second, for the proposed simulation method, it would be interesting to analyze the bid-ask spread pattern over the different time to expiries in the FX option market (either per currency pair or overall), to improve the time to expiry based scaling of perturbations. Lastly, it is of interest to investigate the influence of the quality of the data on the ability of the SVI model to fit the data, for example by enlarging the bid-ask spread in the simulation considerably.

References

Bachelier, L. (1900). *Théorie de la spéculation*. Gauthier-Villars.

Bakshi, G., & Madan, D. (2000). Spanning and derivative-security valuation. *Journal of financial economics*, *55*(2), 205–238.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of financial studies*, *9*(1), 69–107.

Benko, M., Fengler, M., Härdle, W., & Kopa, M. (2007). On extracting information implied in options. *Computational statistics*, *22*(4), 543–553.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The journal of political economy*, 637–654.

Boness, A. J. (1964). Elements of a theory of stock-option value. *The Journal of Political Economy*, 163–175.

Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, 621–651.

Brigo, D., & Mercurio, F. (2002a). Displaced and mixture diffusions for analytically-tractable smile models. In *Mathematical financebachelier congress 2000* (pp. 151–174).

Brigo, D., & Mercurio, F. (2002b). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, *5*(04), 427–446.

Bronzin, V. (1908). *Theorie der prämiengeschäfte.* Leipzig: Franz Deuticke.

Carr, P., Geman, H., Madan, D. B., & Yor, M. (2003). Stochastic volatility for lévy processes. *Mathematical Finance*, *13*(3), 345–382.

Carr, P., & Madan, D. B. (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters*, *2*(3), 125–130.

Chen, A. H. (1970). A model of warrant pricing in a dynamic market. *The Journal of Finance*, *25*(5), 1041–1059.

Clark, I. J. (2011). *Foreign exchange option pricing: a practitioner's guide*. John Wiley & Sons.

Cox, J. C., & Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of financial economics*, *3*(1-2), 145–166.

Dupire, B. (1993). Pricing and hedging with smiles. *Proceedings of the AFFI Conference La Boule*.

Einstein, A. (1905). The theory of the brownian movement. *Ann. der Physik*, *17*, 549.

Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, *9*(4), 417–428.

Fengler, M. R., & Hin, L.-Y. (2015). Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. *Journal of Econometrics*, *184*(2), 242–261.

Garman, M. B., & Kohlhagen, S. W. (1983). Foreign currency option values. *Journal of international Money and Finance*, *2*(3), 231–237.

Gatheral, J. (2004). A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. *Presentation at Global Derivatives & Risk Management, Madrid*.

Gatheral, J. (2011). *The volatility surface: a practitioner's guide* (Vol. 357). John Wiley

& Sons.
Gatheral, J., & Jacquier, A. (2014). Arbitrage-free svi volatility surfaces. *Quantitative Finance*, *14*(1), 59–71.
Gilli, M., & Schumann, E. (2011). Calibrating option pricing models with heuristics. In *Natural computing in computational finance* (pp. 9–37). Springer.
Glaser, J., & Heider, P. (2012). Arbitrage-free approximation of call price surfaces and input data risk. *Quantitative Finance*, *12*(1), 61–73.
Hagan, P. S., Kumar, D., Lesniewski, A. S., & Woodward, D. E. (2002). Managing smile risk. *The Best of Wilmott*, 249.
Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, *20*(3), 381–408.
Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, *6*(2), 327–343.
Hodges, H. M. (1996). Arbitrage bounds of the implied volatility strike and term structures of european-style options. *The Journal of Derivatives*, *3*(4), 23–35.
Jäckel, P. (2006). By implication. *Wilmott*, *26*, 60–66.
Kilin, F. (2006). Accelerating the calibration of stochastic volatility models.
Latané, H. A., & Rendleman, R. J. (1976). Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, *31*(2), 369–381.
Lee, R. W. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, *14*(3), 469–480.
Lee, R. W. (2005). Implied volatility: Statics, dynamics, and probabilistic interpretation. In *Recent advances in applied probability* (pp. 241–268). Springer.
Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, 141–183.
Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, *3*(1-2), 125–144.
Merton, R. C. (2014). *Black-scholes: Robert merton on the options pricing model.* `http://www.bloomberg.com/news/articles/2014-12-04/black-scholes-robert-merton-on-the-options-pricing-model`. (Accessed: 2016-09-06)
Oblój, J. (2008). Fine-tune your smile, correction to hagan et al. *Imperial Collge of London Working Paper*.
Rasmussen, L. (2016). *PhD Thesis Computational Finance: on the search for performance.*
Reiner, E. (2000). Calendar spreads, characteristic functions, and variance interpolation. *Mimeo*.
Roper, M. (2010). Arbitrage free implied volatility surfaces. *preprint*.
Rubinstein, M. (1994). Implied binomial trees. *The Journal of Finance*, *49*(3), 771–818.
Sprenkle, C. M. (1961). Warrant prices as indicators of expectations and preferences. *Yale economic essays*, *1*(2), 178–231.
Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359.
Zimmermann, H., & Hafner, W. (2007). Amazing discovery: Vincenz bronzins option

pricing models. *Journal of Banking & Finance*, *31*(2), 531–546.