Levenberg-Marquardt Algorithm and Bayesian Inference for SABR Implied Volatility Smile Parameter Estimation

Francisco Pedro Furtado Simões

Thesis to obtain the Master of Science Degree in

Mathematics and Applications

Supervisors: Dr. Igor Kravchenko
Prof. Dr. Cláudia Rita Ribeiro Coelho Nunes Philippart

Examination Committee

Chairperson: Prof. Dr. Manuel González Scotto
Supervisor: Dr. Igor Kravchenko
Member of the Committee: Dr. Luciana Salles Barbosa

October 2020

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I would like to express my deepest appreciation to my supervisors Prof. Dr. Cláudia and Dr. Igor for the patience and all the help during the elaboration of this work. They believed in my capabilities and were relentless in providing me with technical support, as well as willing me on along the way to this here end result. I'm very thankful for being able to have had such incredible people as my supervisors.

I would like to thank the team from BNP Paribas, namely, François Giradot, Subhadeep Mazumder, Miguel Ribeiro, Sebastien Brogniez, Maxime Brochard and Abdelkader Challel, for giving me the opportunity to work with them, suggesting the thesis' direction and providing valuable insight to the challenges we faced throughout this work. The partnership with BNP Paribas is the reason this work aims for innovation and is geared towards, hopefully, providing the industry with applications and solutions. The conditions during my time at BNP Paribas were not the easiest due to the current turbulent world state, but everyone's availability to help and work ethic were not wavered despite it. As my first professional experience, I take many valuable lessons with me for life and I'm very grateful for having been part of this team.

I would also like to thank my friends André, Bruno, Diogo, Marta and Mariana, for checking up on my progress, but more importantly, on me. They gave me strength during tough times and cheered me up when I was down, that did not go unnoticed. Thank you from the bottom of my heart.

In my family I find home, this feeling and comfort is greatly important during trying situations, and their support was what kept me going, so thank you to my sisters Rita, Joana and Mariana, and my uncle Jonas for being as supportive and loving as one can be.

No amount of words may express my gratitude towards my parents Carla and Pedro. They sacrificed very much to keep me in this comfortable position allowing me to pursue my academic goals. One could not have chosen better parents. I love you very much mom and dad.

To my grandmother Hortense Furtado Jesus.

Abstract

Under the hypothesis of the underlying asset of an option following the SABR model, we tested the Levenberg-Marquardt algorithm for volatility smile interpolation using the price of European options. This prove to be fast in the simulated framework. However, the real framework sometimes implies uneven distribution of quotes (with respect to relative strikes) or the presence of outliers, these two factors affect significantly the performance of the paremeter estimation using this algorithm. In this work it was also used Bayesian inference as an alternative to mitigate these limitations. We employed this approach on SABR's α parameter with good results.

Keywords

Stochastic Volatility Model; SABR; Bayesian Inference; Levenberg-Marquartdt Algorithm; Financial Mathematics.

Resumo

Sob a hipótese do ativo subjacente a uma opção seguir o modelo estocástico SABR, foi testado o algoritmo Levenberg-Marquardt para interpolação do ”smile” de volatilidade utilizando o preço das opções Europeias. O método demonstrou ser rápido no senário simulado. No entanto, a situação real por vezes implica distribuição não homogénea de dados (em relação aos ”strikes” relativos) e a presença de “outliers”, esses dois fatores afetam significativamente o desempenho de estimação de parâmetros do algoritmo. Neste trabalho foi também aplicada a inferência Bayesiana como alternativa para mitigar essas limitações. Empregámos esta abordagem no parâmetro do modelo SABR com bons resultados.

Palavras Chave

Modelo de Volatilidade Estocástica; SABR; Inferência Bayesiana; Algoritmo Levenberg-Marquardt; Matemática Financeira.

Contents

List of Figures

List of Tables

List of Algorithms

1

Introduction

Contents

The term ”market” makes reference to the financial market, in which securities and derivatives are traded at low transaction costs relative to the assets themselves. The securities include, but are not limited to, public traded companies' stocks and bonds by any company whose willing to issue them. There are two types of popular derivative products, forward contracts and options. In this work we will be focused on the latter and methods involving its contract pricing, ignoring transaction costs.

1.1 Financial Options

A financial option (commonly referred to as option) is a contract between two parties. There are two main types of option contracts: calls and puts. The call contract gives the buyer the right to buy the underlying asset, the put gives the buyer the right to sell. The other party, the seller, has the obligation to sell (call) or buy (put) if the specifications of the contract are met, and the buyer chooses to exercise his right. Each contract has usually the following specific properties:

- The type of contract, whether it is a call option or a put option;
- The specifications of the underlying asset, e.g. quantity and type of asset;
- The strike price, K;
- The duration of the contract or time to maturity, T;
- The contract price.

The strike price is the price at which a specification of the contract can be sold or bought. The contract price is also referred to as the premium of the contract and is usually designated by c in case of a call option contract, or p in case of a put option contract.

When the buyer of the contract may exercise his right, subdivides the types of options further into American options or European options. In the latter the buyer can only exercise his option at the end of the duration of the contract. In contrast, when it comes to American options, the buyer may exercise the option at any time before the terminus of the contract, including at the expiration time. In this work we will only deal with European options, so we will refer to these simply as options.

Usually the underlying assets in these contracts are stocks and their behaviour on the stock market, i.e. their market price at time t, $S(t)$, dictates the contract pricing. This is why modeling their properties is useful to accurately and fairly price option contracts.

We will consider following functions $c \equiv c(K, T, S(0), \sigma)$ and $p \equiv p(K, T, S(0), \sigma)$, where $S(0)$ is the present market value of the underlying of the contract and σ is the volatility of the underlying, and try to model them throughout this work.

1.2 Stochastic Processes

A random variable Y is a function that assigns a number to an event w, $Y(w)$; a stochastic process $\mathbf{X}$ is a function that assigns a number to an event and real variable t, $\mathbf{X}_w(t)$. The latter parameter is often called time, even when it does not correspond to time in the real physical application of the underlying process. Stochastic processes can be seen as a family of random variables indexed by the parameter t,

$$\mathbf{X} = \{\mathbf{X}_t, t \in \Gamma\}, \tag{1.1}$$

where usually $\Gamma = \mathbb{N}_0$ or $\Gamma = \mathbb{R}_0^+$. We note that, given $w \in \Omega$, the collection $\mathbf{X}_w = \{\mathbf{X}_w(t), t \in \Gamma\}$ is a possible realization of the process.

1.2.1 Brownian Motion

The botanist Robert Brown observed that dust particles suspended in the air described an erratic behaviour and their path was random. The family of stochastic processes used to describe these phenomena became known as Brownian motion (BM), that also may be referred to as Wiener processes, in honor of mathematician Robert Wiener who studied the properties of the BM. This kind of behaviour was since discovered in many more instances, for example, an electron under black body radiation, among others. In the financial market, it is used to model the random behaviour of stock prices.

Albert Einstein was one of the people who made an effort to formalize this phenomenon, and one of the assumptions made was that the future behaviour of the particle is only dependent on its position at present time and probability of Δx displacement at time instant Δt. The stochastic processes that exhibit this property are called Markov processes. For instance, we will assume that the future stock prices depend only on the present price and the density probability distribution for the movement of said stock price in the next clock cycle Δt, or in continuous terms, the infinitesimal time change dt, and therefore follow a Markov process.

A stochastic process in continuous time, $B = \{B(t) : t \geq 0\}$ is a BM if the following holds:

- $B(0) = 0$;
- For each $w \in \Omega$, $\{B_w(t), t \geq 0\}$ is a continuous functions of t;
- $B(t) - B(s) \sim N(0, t - s)$, with $Cov(B(t) - B(s), B(s)) = 0, \forall s \leq t$

1.3 The Modeling of Stock Prices

Public companies trade equity in the form of stock shares. The price of each stock depends greatly on the company's product demand, and overall earnings. Its behaviour is practically erratic, though some properties may be extracted from historical data and used to forecast future stock price's trend and volatility.

1.3.1 Stock Price as Geometric Brownian Motion

In this work the stock price process $S = \{S(t) : t \geq 0\}$ will be modelled as a Geometric Brownian motion defined as,

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t), \tag{1.2}$$

where μ is the rate of return and σ is the volatility. The first term is responsible for the drift and the second can be seen as noise to the path of the BM. This is the most widely used model for stock prices. The solution of eq.1.2 is given by:

$$S(t) = S(0)e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B(t)}, \tag{1.3}$$

and therefore the logarithmic returns $ln\left(\frac{S(t)}{S(0)}\right)$ are normally distributed,

$$ln\left(\frac{S(t)}{S(0)}\right) \sim N\left(lnS(0) + \left(\mu - \frac{\sigma^2}{2}\right)t), \sigma^2 t\right). \tag{1.4}$$

1.3.2 Volatility

Roughly speaking, volatility measures the uncertainty associated with the stock price fluctuations. Based on historical data one can estimate this volatility, which, in the case of the GBM, is constant, meaning that it can be estimated by σ, with σ being the standard variation of process $\{S(t), \forall t\}$. Therefore, under this assumption, we would have the following estimate of σ:

$$\hat{\sigma} = \frac{1}{\sqrt{\tau}}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \overline{u})^2}. \tag{1.5}$$

where τ denotes the (constant) time interval between consecutive times t_i and t_{i+1}, for $i = 0, ..., n-1$, where n is the number of historical values of S. So,

$$u_i = ln\left(\frac{S(t_i)}{S(t_{i-1})}\right) \quad \text{for } i = 1, 2, ..., n\ , \tag{1.6}$$

denotes the logarithmic return and its mean is

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{n} u_i. \tag{1.7}$$

If the volatility was constant in time this would be a good enough approximation, but despite some models considering it to be constant, it appears this is not the case.

Later, we will present other models of estimating volatility such as SABR models, in which volatility changes over time, much in the same way as stock prices do.

1.3.3 Interest Rate and The Forward Price of an Asset

Interest rates are fruit of the relation between lenders and borrowers, more specifically with the risk associated with lending money. The higher the risk, the higher is the interest rate. Naturally different credit have different risks and different lenders measure these risks differently, but the market stabilizes the interest rate transversely between all lenders by the way of offer and demand, and risks associated as stated before.

The time value of money is direct product of interest. From an investors point of view, a given amount of money today is more valuable in one year from now, since the investor can take the money to a bank and retrieve it a year later with added interest.

Dictating the rate of return of investments is the risk-free rate, r. In this work the risk-free rate will be considered constant. The forward price, f, of an initial investment S continuously compounded over time period t is:

$$f = Se^{rt} . \tag{1.8}$$

The forward price may not only depend on r, but in this work we consider this approximation to be sufficient for our purposes.

1.4 Black-Scholes Pricing Formula

Black, Scholes and Merton [3] (BSM) predicate some assumptions about the market and stock behaviour. We will not go into further detail on the financial-economic assumptions. The important ones for us are:

- The stock price behaviour is that of a geometric Brownian motion, under the risk neutral measure $\mathbb{Q}$ (for details see Björk [4]);
- No arbitrage opportunities exist;

- The risk-free interest rate, r, is constant.

Under BSM model assumptions, the price of an European option with contract function g and expiration time T is given by

$$V(t,T) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}[g(S(t))] \tag{1.9}$$

where $\{S(t), t \leq T\}$ denotes the process that describes the price of the underlying asset with initial value $S(t)$ at time t.

We refer to Black and Scholes' original paper [5] for a derivation and explanation of this result. In particular for call options we have

$$g(S(T)) = max(K - S(T), 0) \tag{1.10}$$

Then it follows that for a call option, the price at time 0 is

$$V(0,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[(S(T) - K)\mathbb{1}_{S(T)>K}], \tag{1.11}$$

where $\mathbb{1}$ is the indicator function. Since K is a constant, we can separate the equation above into two terms

$$V(0,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[S(T)\mathbb{1}_{S(T)>K}] - e^{-rT}K\mathbb{Q}(S(T) > K), \tag{1.12}$$

where $\mathbb{Q}(S(T) > K)$ is the probability of $S(T)$ being greater than K under measure $\mathbb{Q}$:

$$\mathbb{Q}(S(T) > K) = \mathbb{Q}\left[B(T) > ln\left(\frac{K}{S(0)}\right) - \frac{r - \sigma^2}{2}T\right] = N(d_2), \tag{1.13}$$

since $B(T) \sim N(0,T)$, with $\sigma \equiv \sigma_{BS}$ for the GBM in eq. 1.2. Regarding the first term,

$$\mathbb{E}^{\mathbb{Q}}[S(T)\mathbb{1}_{S(T)} > K] = \int_K^{+\infty} x\mathbb{Q}_{S(T)}(x)|_{S(T)>K}dx = \int_{ln\frac{K}{S(0)}}^{+\infty} S(0)e^x N_z(x)dx, \tag{1.14}$$

where $z = ln\frac{S(T)}{S(0)} \sim N((r - \frac{1}{2}\sigma^2)T, \sigma^2 T)$. Using eqs. 1.4, 1.13 and 1.14, the price of an European call option is given by

$$c_{BS} = S(0)N(d_1) - Ke^{-rT}N(d_2), \tag{1.15}$$

with

$$d_1 = \frac{ln(S(0)/K) + (r + \sigma_{BS}^2/2)T}{\sigma_{BS}\sqrt{T}}, \tag{1.16}$$

$$d_2 = \frac{ln(S(0)/K) + (r - \sigma^2/2)T}{\sigma_{BS}\sqrt{T}} = d_1 - \sigma_{BS}\sqrt{T}. \tag{1.17}$$

The European put option valuation can be computed using the same method or by the put-call parity

$p_{BS} + S(0) = c_{BS} + Ke^{-rT}$, and its value is

$$p_{BS} = Ke^{-rT}N(-d_2) - S(0)N(-d_1). \tag{1.18}$$

2

Volatility Modeling

Contents

In this chapter we focus on the volatility of stock prices, and in particular we address the topic related with the implied volatility and its modeling.

2.1 Black-Scholes' Implied Volatility

How the volatility of a stock price may be estimated was discussed earlier, but in fact, volatility can not be known as it is an intrinsic property of the stock price. We may not need to know the volatility of the stock price to trade fairly, as it is the perceived volatility by the market that matters. That's what traders usually work with and it is known as the implied volatility.

Since volatility is the only parameter of the Black-Scholes pricing *formulae* (eqs. 1.15 and 1.18) that we don't have before hand, it is possible, once we have the option price, to calculate it. Unfortunately, the Black-Scholes result is not invertible so we have to resort to numerical methods to obtain the implied volatility of an option, σ_{BS}, as follows.

Solving eq.1.15 with respect to σ_{BS} is equivalent to finding the zero of a function; in this case we want to find the zero of:

$$g(\sigma_{BS}) = c_{BS}(\sigma_{BS}) - \hat{c}, \tag{2.1}$$

where c_{BS} is the price of the call option under BSM model with known and fixed parameters $(K, T, S(0))$, the parameters K and T are known from the contract and $S(0)$ can be directly observed from the market. $\hat{c}$ is the observed value of the price of the option contract. This zero cannot be found analytically as we have already stated, due to the non-invertibility of the Black-Scholes result. Thus one needs to use numerical methods and there are many proposals in the literature to find roots of functions. In this work we use the Newton-Raphson's method. This method consists on guessing the initial value of x_0 and then employing the following recursive formula:

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}, \tag{2.2}$$

where g' is the derivative of g. This simple implementation of the algorithm should converge to root in $O(n^2)$ [6] and the stopping condition is:

$$\frac{|x_{n+1} - x_n|}{x_{n+1}} < \varepsilon, \tag{2.3}$$

for some predefined ε, in which case it is determined that root of g is approximated by x_{n+1}.

Applying this method to the computation of implied volatility, we consider $x_n \equiv \sigma_{BSn}$ and $g(x_n) \equiv c_{BS}(\sigma_{BSn})$, it follows that:

$$\sigma_{BSn+1} = \sigma_n - \frac{c_{BS}(\sigma_{BSn+1}) - \hat{c}}{\nu}, \tag{2.4}$$

where ν is often referred to as ”vega” and is defined as follows:

$$\nu(\sigma_{BSn}) = \frac{\partial c_{BS}(\sigma_{BSn})}{\partial \sigma_{BSn}} = S_0 \phi(d_1(\sigma_{BSn}))\sqrt{T} \tag{2.5}$$

with $\phi(x)$ being the density function of the standard normal distribution.

2.2 The SABR Model

The insight provided by the Black-Scholes approach to option pricing proved to be very useful due to its simplicity and wide application. Despite that, it relies on some assumptions that aren't very applicable the real world financial market. The one that is interesting for the purposes of this work being the nature and overall behaviour of the volatility, which Black [5] assumes to be a constant. As historical records show, the volatility of the stock prices is rarely constant in time. Therefore one needs to find models where the volatility is itself changing with time.

One possible generalization of the BSM model is the introduction of stochastic volatility. We will be focusing on a model proposed by Hagan *et.al* [1],

$$df(t) = \alpha(t) f(t)^{\beta} dB_f(t), \tag{2.6}$$

$$d\alpha(t) = \nu \alpha(t) dB_{\alpha}(t), \tag{2.7}$$

where β is the constant elasticity of variance valued between 0 and 1, and ν is the sensitivity of the volatility, or, in other words, the volatility of the volatility process. We define $\alpha := \alpha(0)$. Finally, $\{B_f(t), t \geq 0\}$ and $\{B_{\alpha}(t), t \geq 0\}$ are two correlated Brownian Motions, such that:

$$\mathbb{E}\left[dB_f(t) dB_{\alpha}(t)\right] = \rho dt. \tag{2.8}$$

Hence the name SABR: stochastic-α-β-ρ. These are the parameters that comprise the modeling of the volatility surface in parameters $(K, f(0))$.

We are interested in using SABR for volatility surface interpolation, i.e., given a set of prices of a type of derivative, the SABR model is calibrated to fit observed implied volatilities and use them as an interpolation tool for other strikes. For the remainder of this work we shall refer to it as SABR fitting. All of this resides in that different strike prices K result in different volatilities for a given derivative price, and therefore the BSM model is not enough since it considers the volatility to be constant for all strikes.

Let us introduce $I(K, T)$ as SABR's implied volatily function (or more commonly referred to as volatility smile). Its expansion in Taylor series expansion with respect to the maturity T is,

$$I(K,T) = I^0(K)(1 + I^1(K)T) + O(T^2), \tag{2.9}$$

the terms $I^0(K)$ and $I^1(K)$ are, respectively, the zeroth and first order coefficients of the Taylor series expansion. For the applications, several approximations are used. In table 2.1 taken from Obloj [7] several of them are presented.

The approximation *formulae* derived in Hagan *et al.* [1] though important, is problematic in some cases. The *formulae* is known to produce negative prices for the region for options with small strike prices and large maturities [7]. Furthermore, in this formulae, $\lim_{\beta\to 1} I^0(K)$ is inconsistent with $I^0_{\beta=1}(K)$, meaning the I^0 term when $\beta = 1$. Berestycki *et al.* [2] took on deriving the formulae in a more rigorous manner. Obloj [7] computed and summarized the differences of both formulae $I^0(k)$ term in Table 2.1 in which we consider $k := \frac{K}{f(0)}$, as the relative strike. This concept of relative strike not only is useful to simplify the notation of the SABR formulae, but will be used throughout the work as it is describes the at-the-money relative strike as $k = 1$ for any option, and therefore, for European call options $k > 1$ refers to an in-the-money contract and $k < 1$ to an out-of-the-money contract.

	Hagan *et al.* [1]	Berestycki *et al.* [2]	notation
$I^0(1)$	$\alpha K^{\beta-1}$	$\alpha K^{\beta-1}$	
$I^0(k)_{\nu=0}$	$\frac{-ln(k)\alpha(1-\beta)}{f^{1-\beta}-K^{1-\beta}}$	$\frac{-ln(k)\alpha(1-\beta)}{f^{1-\beta}-K^{1-\beta}}$	
$I^0(k)_{\beta=1}$	$-ln(k)\nu/ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)$	$-ln(k)\nu/ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)$	$z = \frac{-ln(k)\nu}{\alpha}$
$I^0(k)_{\beta<1}$	$-ln(k)\nu\frac{\mu}{z}/ln\left(\frac{\sqrt{1-2\rho\mu+\mu^2}+\mu-\rho}{1-\rho}\right)$	$-ln(k)\nu/ln\left(\frac{\sqrt{1-2\rho z+z^2}+z-\rho}{1-\rho}\right)$	$z = \frac{\nu}{\alpha}\frac{f^{1-\beta}-K^{1-\beta}}{1-\beta}$ $\mu = \frac{\nu}{\alpha}\frac{f-K}{(fK)^{\beta/2}}$

Table 2.1: Comparison of $I^0(k)$ term in [1] and [2], with $k = \frac{K}{f(0)}$.

The difference in Table 2.1 resides in the $I^0(k)_{\beta<1}$ term. The Berestycki *et al.* [2] formulation advantages and numerical benefits are demonstrated by Obloj [7].

Next we take a closer look to the impact of each parameter to the smile characteristics[1].

[1] For a more detailed analysis we refer to Crispoldi *et. al* [8]

2.2.1 Impact of α

The α parameter is the underlying's volatility at $t = 0$, that being, α is, intuitively proportional to the implied volatility. It is expected for the smile to go up with the value of α as can be seen in Fig.2.1.

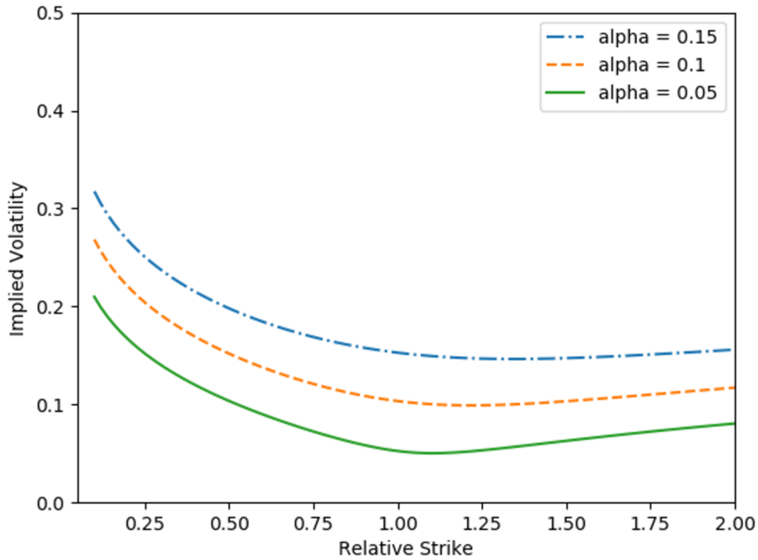


Figure 2.1: Impact of α in the volatility smile.
$\beta = 1$, $\rho = -0.33$, $\nu = 0.25$.

2.2.2 Impact of ν

The parameter ν is the volatility of the process $\{\alpha(t) : t \in [0, T]\}$. It impacts the "curvature" of the smile, the value of the implied volatility increases in the wings, i.e. deep in-the-money and deep out-of-the-money strikes, as ν increases (fig.2.2). The impact at-the-money is not so relevant and may be compensated by α as described above.

2.2.3 Impact of ρ

The correlation of both BMs B_f and B_α, ρ, takes value between -1 and 1, and in most markets is negative, e.g. a downward movement of the forward price f is likely to prompt a upward movement of the volatility α. The intuition of the behaviour of the smile facing changes in ρ doesn't come easy. It's important to note the relation between β and ρ. It appears that the value of β changes the nature of impact that ρ has on the smile. With closer inspection, we can see that, in both Fig.2.3 and Fig.2.4, ρ is slanting or, more commonly used in the literature, affecting the skew of the smile. ´

´

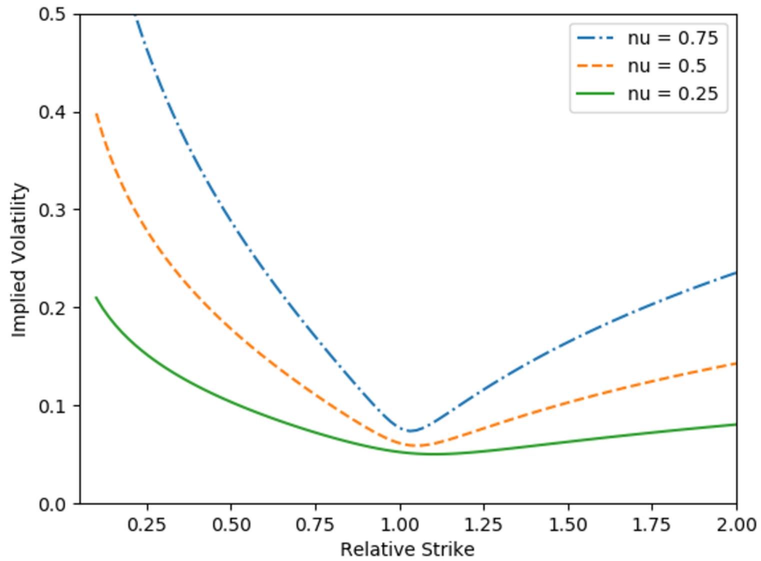


Figure 2.2: Impact of ν in the volatility smile.
$\alpha = 0.05$, $\beta = 1$, $\rho = -0.33$.

2.2.4 Impact of β

This last parameter β is the constant of elasticity of the variance (CEV) and its value is limited to the interval $[0, 1]$.

Its effect on the smile can be seen as similar to the ρ parameter as it affects the smile's skew (fig.2.5). Its value is chosen to accommodate the market's characteristics, and several particular cases present interest:

- $\beta = 0$: here we distinguish between two cases. If $\alpha(t) \equiv \alpha$,then

$$df(t) = \alpha dB_f(t) \tag{2.10}$$

  meaning that $\{f(t), t \geq 0\}$ is a BM scaled by a factor α. Thus f may take negative values (as $f(t) \sim N(0, \alpha^2 t)$), which may not be appropriate. If $\alpha(t)$ depends on t, the same problem holds, but its analysis is more difficult;

- $\beta = 0.5$: this leads to the particular case of the CEV models (see Henry-Labodère [9]). This case does not lead to negative values of $\{f(t), t \geq 0\}$. But there is still one problem: the state 0 is reachable and commonly set as an absorbing state. If the underlying is stock of a company, this means that the company defaulted. The probability that this happens decreases with larger $f(0)$ and increases with larger T (see Crispoldi *et al.* [8]).

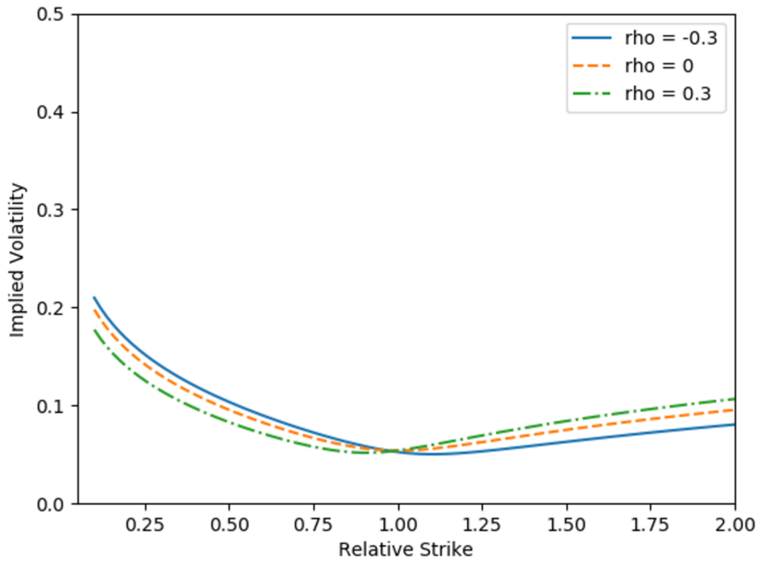


Figure 2.3: Impact of ρ in the volatility smile, $\beta = 1$.
$\alpha = 0.05$, $\nu = 0.25$.

- $\beta = 1$: in case $\alpha(t) \equiv \alpha$, constant in time, then this corresponds to the Brownian motion with zero drift. The advantage of this case is that $\{f(t), t \leq T\}$ will never assume non positive values.

As mentioned before, the effect of β and ρ on the volatility is difficult to distinguish, as it is quite similar: both are responsible for the downward skew in the volatility smile as K increases. The presence of both may lead to an overparametrization, and sometimes one of them is discarded. Along this work we will assume $\beta = 1$.

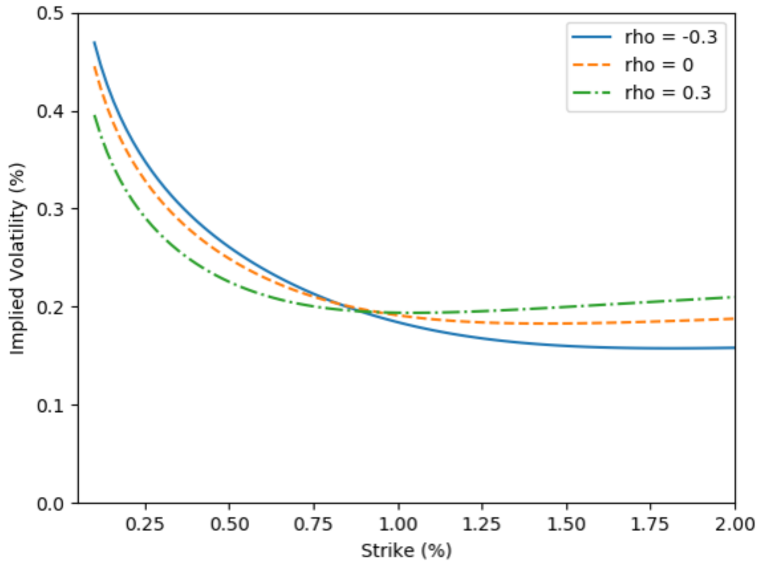


Figure 2.4: Impact of ρ in the volatility smile, $\beta = 0.5$.
$\alpha = 0.05$, $\nu = 0.25$.

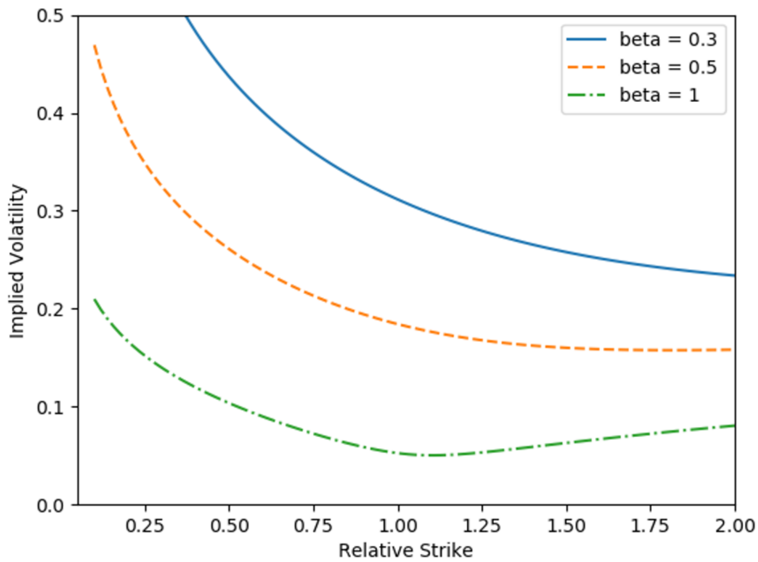


Figure 2.5: Impact of β in the volatility smile.
$\alpha = 0.05$, $\rho = -0.33$, $\nu = 0.25$.

3

Methodology

Contents

3.1 Simulating Quotes and Testing SABR Algorithm

In order to check the accuracy and correctness of our SABR model we can simulate the values of prices of options, for which we know the SABR parameters, and then apply the SABR estimation algorithms, allowing to compare the real values with the estimated ones.

3.1.1 Discretization of the Stochastic Processes

Both stock price and volatility are continuous BMs under the SABR model's assumptions. We use the method bellow to discretize and simulate the values of $\{f(t), t \in [0, T]\}$ and $\{\alpha(t), t \in [0, T]\}$ processes.

Given a differential equation of type:

$$\frac{dP(t)}{dt} = g(P(t)), \tag{3.1}$$

let $t_n = n\Delta t$, where Δt is some fixed time interval, the discretization of eq. 3.1 can be done as:

$$\frac{dP(t)}{dt} \approx \frac{\Delta P(t)}{\Delta t} = \frac{P_{n+1} - P_n}{\Delta t}, \tag{3.2}$$

with P_n denoting P computed at time t_n. Given some initial value $P(0) = P_0$, we set the following iterative formula as an approximation of the process $\{P(t), t \in]0, T]\}$:

$$P_{n+1} = P_n + g(P_n)\Delta t. \tag{3.3}$$

The above method is referred to as Euler's method. Using this construction for SABR model, the discrete approximation to the stochastic differential equation 2.6 is:

$$f_{n+1} = f_n + f_n r\Delta t + e^{-r}(T-t)(1-\beta)\sigma_{n+1} f_n^{\beta-1}\varepsilon_1, \tag{3.4}$$

where $\varepsilon_1 \sim N(0, \Delta t)$, and the volatility (eq. 2.7) is computed as:

$$\sigma_{n+1} = \sigma_n + \sigma_n \nu \varepsilon_2 \tag{3.5}$$

where ε_2 is defined as follows:

$$\varepsilon_2 = \rho\varepsilon_1 + \sqrt{(1-\rho^2)}\varepsilon \tag{3.6}$$

with $\varepsilon \sim N(0, \Delta t)$. This ensures $Cor(\varepsilon_1, \varepsilon_2) = \rho$ and $\varepsilon_2 \sim N(0, \Delta t)$.

3.2 Estimation Algorithms of SABR Model Parameters

In this section we discuss the algorithms used to estimate the parameters of the SABR model. The input parameters of each derivative include

- Type of option or strategy;
- Underlying price at the time of the quote $(S(0))$;
- Time of maturity (T);
- Strike price (K);
- The price of the option (c).

The time to maturity T considered is in years.

Black-Scholes' implied volatility σ_{BS} is computed for each derivative i using the procedure described in section 2.1. That entails that Black-Scholes' price must be computed as part of the iterations of the Newton-Raphson's method. After Black-Scholes' implied volatility is computed we are in a position to estimate the SABR model parameters.

3.2.1 Grid Method

One of the methods for the estimation of parameters is the grid method, that we explain in the following. For each parameter (α, ρ, ν) we define an upper and lower bound and an increment step:

$$(\underline{\alpha}, \overline{\alpha}, \Delta_\alpha), (\underline{\rho}, \overline{\rho}, \Delta_\rho), (\underline{\nu}, \overline{\nu}, \Delta_\nu), \tag{3.7}$$

such that $\frac{\overline{\alpha}-\underline{\alpha}}{\Delta_\alpha} \in \mathbb{N}$, and similarly for the other parameters. Within the values of the grid for each parameter, we chose the ones that solve the minimization problem:

$$(\alpha_{opt}, \rho_{opt}, \nu_{opt}) = \underset{\alpha_i, \rho_i, \nu_i}{argmin} \sum_{j=1}^{N} (I(K_j; \alpha_i, \rho_i, \nu_i) - \sigma_{BSj})^2, \tag{3.8}$$

with,

$$\alpha_i = \underline{\alpha} + i\Delta_\alpha, \quad i = 1, ..., \frac{\overline{\alpha} - \underline{\alpha}}{\Delta_\alpha}; \quad \rho_i = \underline{\rho} + i\Delta_\rho, \quad i = 1, ..., \frac{\overline{\rho} - \underline{\rho}}{\Delta_\rho}; \quad \nu_i = \underline{\nu} + i\Delta_\nu, \quad i = 1, ..., \frac{\overline{\nu} - \underline{\nu}}{\Delta_\nu}.$$

$I(k_j; \alpha_i, \rho_i, \nu_i)$ denotes the volatility of the j^{th} option (with relative strike k_j), assuming that the parameters of the SABR model are $(\alpha_i, \rho_i, \nu_i)$, and σ_{BSj} is the Black-Scholes' implied volatility computed using the j's contract specifications.

3.2.2 Levenberg-Marquardt Algorithm

Let $f(t_i), i = 1, ..., N$ denote the set of observations, $\hat{f}(t_i, \theta)$ the estimate of $f(t_i)$ under the parametrizations $\hat{\theta} = \hat{p}_1, ..., \hat{p}_m$ for the set of parameters $p_j = 1, ..., m$. We may use the criteria of appropriate fit, the chi square statistic[1]:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left[\frac{f(t_i) - \hat{f}(t_i, \theta)}{\sigma_i} \right]^2 \tag{3.9}$$

where σ_i is the measurement error for observation $f(t_i)$.

We can write the above as follows. Denote by $\boldsymbol{f}$ the vector of $f(t_i)$, by $\hat{\boldsymbol{f}}(\theta)$ the vector of $\hat{f}(t_i, \theta)$, and by $\boldsymbol{W}$ the weighting diagonal matrix with $W_{ii} = 1/\sigma_i$ and $W_{ij} = 0$ for $i \neq j$ such that:

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))^\intercal \boldsymbol{W} (\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta})), \tag{3.10}$$

where $A^\intercal$ denotes the transpose of vector A.

The Levenberg-Marquardt algorithm (LMA) results from the combination of two very commonly used numerical methods for minimization: the gradient descent method (GDM) and the Gauss-Newton method (GNM).

At each iteration we will update our initial parameter vector $\boldsymbol{\theta}_0$ by $\boldsymbol{h}$, i.e. $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \boldsymbol{h}$.

The GDM will update $\boldsymbol{\theta}$ towards minimization of χ^2 by means of following the gradient:

$$\partial_{\boldsymbol{\theta}} \chi^2(\boldsymbol{\theta}) = \partial_\theta(\boldsymbol{f}^\intercal \boldsymbol{W} \boldsymbol{f} - 2\boldsymbol{f}^\intercal \boldsymbol{W} \hat{\boldsymbol{f}}(\boldsymbol{\theta}) + \hat{\boldsymbol{f}}(\boldsymbol{\theta})^\intercal \boldsymbol{W} \hat{\boldsymbol{f}}(\boldsymbol{\theta})) = -2(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))^\intercal \boldsymbol{W} \partial_{\boldsymbol{\theta}} \hat{\boldsymbol{f}}(\boldsymbol{\theta}). \tag{3.11}$$

Let $\boldsymbol{J}$ denote the matrix $\boldsymbol{J}_{ij} = \partial_{\theta_i} \hat{\boldsymbol{f}}_{\boldsymbol{j}}$. Then the above equation may be written as follows:

$$\partial_{\boldsymbol{\theta}} \chi^2(\boldsymbol{\theta}) = -2(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))^\intercal \boldsymbol{W} \boldsymbol{J}, \tag{3.12}$$

Furthermore let the parameter update vector $\boldsymbol{h}$ be defined as:

$$\boldsymbol{h} = \alpha \boldsymbol{J}^\intercal \boldsymbol{W} (\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta})), \tag{3.13}$$

where α is the size of the gradient descending step.

The GNM relies on the assumption that $\chi^2(\boldsymbol{\theta})$ is quadratic near the optimal solution and for $\boldsymbol{h}$ ”small” enough, can be expanded in Taylor series as:

$$\chi^2(\boldsymbol{\theta} + \boldsymbol{h}) = \chi^2(\boldsymbol{\theta}) + \boldsymbol{J}\boldsymbol{h} + O(\boldsymbol{h}^2). \tag{3.14}$$

[1] Assuming the residuals are normally distributed

Plugging eq. 3.14 into eq. 3.10, we obtain the following approximation of χ^2:

$$\chi^2(\boldsymbol{\theta}+\boldsymbol{h}) \approx \boldsymbol{f}^\intercal \boldsymbol{W} \boldsymbol{f} - 2\boldsymbol{f}^\intercal \boldsymbol{W} \hat{\boldsymbol{f}}(\boldsymbol{\theta}) + \hat{\boldsymbol{f}}(\boldsymbol{\theta})^\intercal \boldsymbol{W} \hat{\boldsymbol{f}}(\boldsymbol{\theta}) - 2(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))^\intercal \boldsymbol{W} \boldsymbol{J} \boldsymbol{h} + \boldsymbol{h}^\intercal \boldsymbol{J}^\intercal \boldsymbol{W} \boldsymbol{J} h. \quad (3.15)$$

We can argue that the update $\boldsymbol{h}$ that minimizes χ^2 is found from $\partial_{\boldsymbol{h}}\chi^2$, so:

$$\partial_{\boldsymbol{h}}\chi^2(\boldsymbol{\theta}+\boldsymbol{h}) \approx -2(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))^\intercal \boldsymbol{W} \boldsymbol{J} + \boldsymbol{h}^\intercal \boldsymbol{J}^\intercal \boldsymbol{W} \boldsymbol{J}, \quad (3.16)$$

and so:

$$(\boldsymbol{J}^\intercal \boldsymbol{W} \boldsymbol{J})\boldsymbol{h} = \boldsymbol{J}^\intercal \boldsymbol{W}(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta})), \quad (3.17)$$

gives us the update step for the GNM.

As one can see the GDM and GNM update equations, 3.13 and 3.17 respectively, have similar form, and indeed can be combined. This is the approach that Marquardt [10] took, and is at the core of the LMA. A parameter λ is used as a mix factor between the two update equations as follows:

$$(\boldsymbol{J}^\intercal \boldsymbol{W} \boldsymbol{J} + \lambda \boldsymbol{I})\boldsymbol{h} = \boldsymbol{J}^\intercal \boldsymbol{W}(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta})), \quad (3.18)$$

where $\boldsymbol{I}$ is the identity matrix. Initially $\lambda = \lambda_0$, and will update when the quantity:

$$\rho(\boldsymbol{h}) = \frac{\chi^2(\boldsymbol{\theta}) - \chi^2(\boldsymbol{\theta}+\boldsymbol{h})}{\chi^2(\boldsymbol{\theta}+\boldsymbol{h})} \approx \frac{\chi^2(\boldsymbol{\theta}) - \chi^2(\boldsymbol{\theta}+\boldsymbol{h})}{\boldsymbol{h}^\intercal(\lambda \boldsymbol{h} + \boldsymbol{J}^\intercal \boldsymbol{W}(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta})))}, \quad (3.19)$$

is above a set value ε $(\ll 1)$, for a given iteration with step $\boldsymbol{h}$, as $\lambda := max[\lambda/9, 10^{-7}]$, following setting $\boldsymbol{\theta}_i := \boldsymbol{\theta}_{i-1} + \boldsymbol{h}$. Otherwise, $\lambda := min[\lambda \times 7, 10^7]$ and $\boldsymbol{\theta}_i := \boldsymbol{\theta}_{i-1}$, considering the i^{th} iteration of the process. The function $\rho(h)$ measures the improvement of the $\chi^2(\boldsymbol{\theta})$ statistic after update h. This method and specific values are presented in Gavin [11] and we refer to it for further clarification.

There are three proposed criteria of convergence:

- Gradient convergence: max $\left|\boldsymbol{J}^\intercal \boldsymbol{W}(\boldsymbol{f} - \hat{\boldsymbol{f}}(\boldsymbol{\theta}))\right| < \epsilon_1$;
- Parameter convergence: max $|h_i/p_i| < \epsilon_2$;
- χ^2 convergence: $\chi^2/(N-m+1) < \epsilon_3$;

where N is the size of vector f and m is the size of vector θ. As usual, there is also a maximum number of iterations.

LMA is a local algorithm, "good" initial estimates of the vector $\boldsymbol{\theta}$ are essential to assure fast and optimal solution convergence. On the other hand bad initial estimates can lead the algorithm to diverge or converge to some local minimum. Thus we discuss next how we can set these intrinsic estimates of $\boldsymbol{\theta}$ in such a way that the LMA converges to a value near the true one.

3.2.3 First estimation for SABR parameters - Smart Parameters

Gauthier and Rivaille [12] provide an algorithm to find these initial estimates fast and effectively, which they called Smart Parameters. The method is based on the following conjecture: good initial values are such that the SABR smile matches the data implied volatility and its derivative with respect to K at-the-money:

$$\hat{I}(f,K)|_{K=f} = \sigma_{BS}(f,K)|_{K=f}, \tag{3.20}$$

$$\partial_K \hat{I}(f,K)|_{K=f} = \partial_K \sigma_{BS}(f,K)|_{K=f}, \tag{3.21}$$

where $\hat{I}(f,K)$ is the estimation for SABR implied volatility, σ_{BS} are the market observed implied volatilities (or rather, some interpolation of them) and $\partial_K \hat{I}(f,K)|_{K=f}$ denotes the derivative with respect to K of the SABR implied volatility formula at $K = f$. The first equation produces a second degree polynomial in ρ. Gauthier and Rivaille in developing this method assume that both roots of this polynomial are real. The method may not produce results if this does not hold.

Then eq. 3.21 needs to be evaluated for both roots of 3.20:

$$\partial_K \hat{I}(\rho_+(\alpha,\nu); f,K)|_{K=f} = \partial_K \sigma_{BS}|_{K=f}, \tag{3.22}$$

$$\partial_K \hat{I}(\rho_-(\alpha,\nu); f,K)|_{K=f} = \partial_K \sigma_{BS}|_{K=f}, \tag{3.23}$$

where ρ_+ and ρ_- denote the two roots (potentially different) of 3.20. We are now in a position where we have two equations with two unknowns, α and ν (note that β is fixed and f and K are constants for each quote). This system of equations does not have a closed solution, but in principle can be solved numerically.

We implemented this method as smart parameters method's claims seemed quite tempting, it should offer good enough initial approximations for the parameter set α, ρ and ν such that the LMA would fit the smile both fast and accurately, avoiding local *minima*. This prove not to be so easy to implement due to lack of procedural information in the original bibliography [12], and we believe it also presents a real world applicability problem. The method relies on the assumption that both σ_{ATM} and $\sigma'|_{ATM}$ are known or at least well estimated. However we could not find in the literature a way for estimating $\sigma'|_{ATM}$. In an attempt to use this method, we have considered the following procedure. We considered a polynomial regression of the quotes with strike near $k = 1$ (say $k \in [0.8; 1.2]$). We considered its value ATM to be σ_{ATM} and the derivative to be $\sigma'|_{ATM}$. These two quantities will, in the smart parameters method, produce upper and lowers bounds for α given some ν, in the system of equations 3.22 and 3.23, and yield $\rho \equiv \rho(\alpha,\nu)$. If these quantities do not represent well the characteristics of the quote underlying smile, the method will produce disastrous results, e.g. $\rho < -1$ or $\rho > 1$, which in turn renders the pair (α,ρ) also incoherent with the smile. In our view this is due to the difficulty of precisely estimate σ_{ATM}

and $\sigma'|_{ATM}$ from quotes which aren't sufficiently well behaved. This problem only worsens for sparse data with lack of quotes with strike near $k = 1$. This suspicion about the flexibility of the method was confirmed with our numerical results. We could not find, in great majority of the cases, admissible initial estimates using this proposal. So we decided not to use the Smart Parameters method in this work. We conclude that this method may be interesting in the academic setting, but might have application issues in a industry environment. In most cases the user can set as initial parameters the set from the previous SABR fitting. In the beginning of a trading day this can be the parameters of the day before, which will (more often than not) be good enough to initialize fitting.

3.2.4 Bayesian Inference

In this work we propose to estimate SABR's α using Bayesian inference. A decision problem is one where the user chooses a distribution from some family of distributions it prefers for the variable which he'll make observations on. Each observation will affect the variable's distribution as is affects the users decision. Bayesian decision theory or, as it is more commonly known as, Bayesian inference, has as its foundation Bayes' theorem

$$h(\theta|x) = \frac{f(x|\theta)h(\theta)}{f(x)}, \quad \theta \in \Theta \tag{3.24}$$

where Θ is the the parameter space and $h(\theta)$ is the *a priori* distribution of parameter θ, $f(x|\theta)$ is the likelihood of observation $X = x$ given θ, $h(\theta|x)$ is the *a posteriori* distribution of parameter θ after observation $X = x$ and $f(x) = \int_\Theta f(x|\theta)h(\theta)d\theta$, is the predictive distribution of X.

In the particular case that $X \sim N(\theta, \sigma^2)$ and $\theta \sim N(a, b^2)$, we have the following

$$f(x|\theta)h(\theta) = \frac{1}{2\pi\sigma b} exp\left[-\frac{1}{2}\left(\frac{(\theta-a)^2}{b^2} - \frac{(x-\theta)^2}{\sigma}\right)\right], \tag{3.25}$$

and by defining $c := (1/b^2 + 1/\sigma^2)$:

$$f(x) = \int_\Theta f(x|\theta)h(\theta)d\theta = \frac{1}{\sqrt{2\pi c}}\frac{1}{\sigma b} exp\left[-\frac{(x-a)^2}{2(b^2+\sigma^2)}\right], \tag{3.26}$$

which means that predictive distribution of X is $N(a, b^2 + \sigma^2)$. Moreover, the *a posteriori* distribution of θ, following from eq.3.24, is given by

$$h(\theta|x) = \sqrt{\frac{c}{2\pi}} exp\left[-\frac{1}{2}c\left(\theta - \frac{1}{c}\left(\frac{a}{b^2} + \frac{x}{\sigma^2}\right)\right)^2\right], \tag{3.27}$$

and thus

$$\theta|x \sim N(A, B^2), \quad A = \frac{\frac{1}{b^2}a + \frac{1}{\sigma^2}x}{c}, B^2 = \frac{1}{c}. \tag{3.28}$$

In the following section we use these results for inferring about SABR's α.

3.2.4.A Inference on Parameter α in SABR Model

First, we assume that instead of eq.2.9, we consider only the approximation of order 0 , meaning that:

$$I(k,T) = I^0(k) + O(T). \tag{3.29}$$

Then we assume the ATM case for which $k = 1$ (see table 2.1) and $\beta = 1$ which means that:

$$I^0(k) = I^0(1) = \alpha. \tag{3.30}$$

We also assume that we'll make observations on a random variable σ_{BS}, the Black-Scholes' implied volatility of a quote. We have that an observation is $\sigma_{BS} = v$. We assume that

$$\sigma_{BS} \sim N(\alpha, s), \tag{3.31}$$

where, for our purposes, we set $s = 0.1$, this is the "decision" part of the procedure, and may change given the user's belief on the market, in our case this value seems reasonable given the quote simulation method. The *a priori* distribution is also normal and its hyper-parameters are set by the history of past α fittings, i.e. μ_α is the sample mean and σ_α is the sample standard deviation of α values in the past

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha), \tag{3.32}$$

so, following eq. 3.28, the *a posteriori* distribution is

$$\alpha | v \sim N(A, B^2), \tag{3.33}$$

with

$$A = \frac{\frac{1}{\sigma_\alpha^2}\mu_\alpha + \frac{1}{s^2}v}{c}, B^2 = \frac{1}{c}, \tag{3.34}$$

where $c := (1/\sigma_\alpha^2 + 1/s^2)$. This results are presented in more detail in Paulino [13].

4

Results & Discussion

Contents

In this chapter we explain how we get the results for the 2 estimation procedures, LMA and Bayesian inference, and analise them. We consider also their performance under the presence of outliers. All the figures that show these results may be found in appendix.

4.1 SABR quote simulation

In the remaining of this work we will use the simple metric "$error$" to get a sense of fitness of each estimation of the smile $\hat{I}(k, (\hat{\alpha}, \hat{\rho}, \hat{\nu}))$ to the observations $\sigma_{BSi}, i = 1, ..., N$,

$$error = \frac{\sum_i \left| \hat{I}(k_i, (\hat{\alpha}, \hat{\rho}, \hat{\nu})) - \sigma_{BSi} \right|}{N}. \tag{4.1}$$

In this section we analyse the behaviour of the simulation results, in order to assess how close they are from the theoretical ones, using the $error$ of the simulations and/or fittings as a reference. In fig. A.1 we plot the "theoretical SABR" curve along with the implied volatility of the simulated quotes. Each quote is in fact the result of the average of N ($N = 10^2, 10^3, 10^4, 10^5$) simulated processes' payoffs, given the quote's strike. We also include the $error$ for each of the examples.

We proceeded to perform 50 simulations like the ones in fig.A.1 (for each N). For each quote j with forward price of the underlying $f_j(0)$ and strike K_j, we simulate N values of the price of the underlying at maturity $f_j(T)$. Then, we use the average of the payoff of the contract, given $f_j(T)$ and strike K_j, to determine the contract price and to compute the quotes' implied volatility. The sample in for which we compute the $error$ contains 20 quotes similarly to the examples in fig.A.1, and so, $j = 1, ..., 20$. The results of the mean of the $error$ for 50 simulations of these quote samples for each considered value of N, are shown in table 4.1.

Number of Simulations	mean $error$	$error$ standard deviation
10^2	0.018585	0.004242
10^3	0.008793	0.001099
10^4	0.003379	0.001121
10^5	0.002435	0.000185

Table 4.1: $error$ of 20 quote samples as function of number of f_t and α_t processes simulations for option pricing.

In the case of 10^3 process simulations per quote, the mean $error$ is 0.008793 (see table 4.1), this was our preferred choice of number of simulations per quote, as it introduces a reasonable amount of variability in the implied volatilities of the quotes, as in the real market quotes. In this case, the $error$ represents about 1% of the value of the implied volatility of quotes ATM, with quotes on the wings carrying a bit more of responsibility in this variability as it can be seen in fig.4.1. As presented in the first chapter, this is even more of a problem when $\beta = 0.5$, given that in this case we will have a large proportion of values equal to zero (as zero is an absorbing state, see fig.4.2).

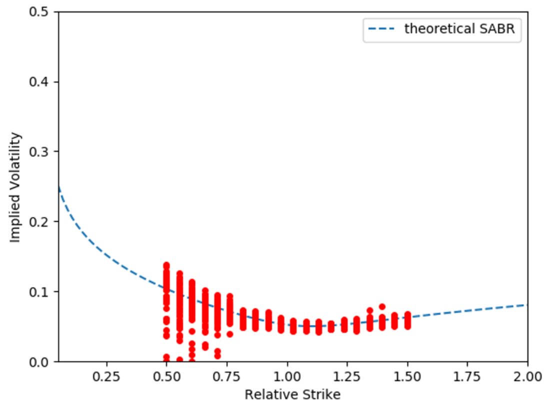


Figure 4.1: 50 samples of a 20 quotes with 10^3 of f_t and α_t simulations for each quote pricing computation.

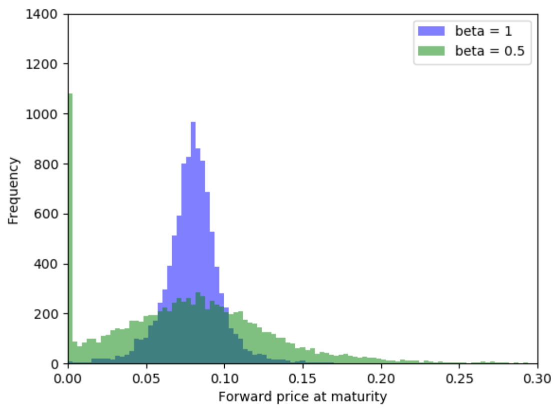


Figure 4.2: f_T distribution for $\beta = 0.5$ and $\beta = 1$. 10^4 process simulations.

Note that the maturity we use throughout this work is quite large which accentuates even more the problem of the high mass at $f_T = 0$.

4.2 LMA Performance

4.2.1 Time performance

In order to ascertain the LMA's time efficiency, we tested it using simulated quotes and compared it to the grid method. The time results in table 4.2 are only referent to the fitting process, not including the quotes' simulation.

We have considered no more than 100 quotes, considering that the information farther in time is not relevant for the fitting process. For the purpose of the interpolation, we consider it to be one day of quotes. The SABR smile is dynamic, it changes throughout the day and even more so from day to day.

	Number of Quotes	10	50	100
Process time (s)	Grid	0.2935	1.473	3.009
	LMA	0.01664	0.05541	0.1180
Initial $error$	LMA	0.03470	0.03389	0.03508
Estimation $error$	Grid	0.006474	0.008181	0.007646
	LMA	0.01314	0.01199	0.01128

Table 4.2: Algorithm time comparison.

Old quotes may not represent the present value of the SABR parameters and for that reason historic data may be only considered until a certain time in the past. In our opinion, it is reasonable to assume that the oldest quotes in a samples with hundreds of quotes are not to be considered for the present estimation.

4.2.2 Comparison between Grid and LMA method

From table 4.2 we conclude that the LMA is faster than the grid method and the estimation errors are comparable. The grid method we presented earlier as a simplification of the calibration suggested by West [14] is as fast as the user wishes in trade for how broad the search of the solutions is. Often it can not be applied in the real world, because it relies to much on user input of the bounds and the information necessary to make the correct guess might not be available on the market. In a setting, too great of a difference between bounds may cause the algorithm to take too much time to compute SABR's smile optimization.

With a maximum iterations possible, the running time of the LMA is limited, but the precision may not be reached and that's the trade off the user has to balance. One parameter that may be overlooked in the method is λ, or rather, its initial value. This parameter can also be tuned for faster convergence, given the user's beliefs as to the most appropriate fitting method. Namely, if the user sets $\lambda >> 1$ then the method used will resemble the GDM in the optimization, otherwise, if $\lambda << 1$, it will be the GNM to take on the dominant role. Anyway, even with some other choice of λ, the LMA performs better than using only one of the GDM or GNM.

4.2.3 Initial Parameter Impact on LMA

As described before, LMA is a local algorithm, and thus initial parameter choices may have a great impact in the quality of the fitting. We need to set a benchmark for the initial parameter we shall use during this work. We computed how a shift[1] of the LMA's initial parameters affects the values of the theoretical parameters. We have simulated 10 quotes and looked at its performance.

[1] The shift is % of theoretical parameter values.

Shift(%)	Mean Initial Error	Mean Estimation Error
0	0.007965	0.008930
5	0.008331	0.009384
10	0.009132	0.009630
50	0.015454	0.010318
90	0.023576	0.011468
130	0.031379	0.012486

Table 4.3: Impact of initial parameters shift in mean error after a LMA fitting on 10 simulated quotes.

The results in table 4.3 show that inputting parameters with up to 10% shift of the real values into the LMA does not impact significantly the $error$. Please note that the convergence criteria of the LMA that is reached the most is the χ^2 minimization. So this result in terms of $error$ may not represent that the LMA is under-performing or not fitting well the smile, it just means that up to a shift of 10% the LMA will not minimize the $error$. If minimizing the error is the objective and the user is certain that his initial parameter estimates are within 10% of the smile's real parameters, we would not recommend using the LMA.

On the other hand, the LMA seems to perform appropriately up to 130% shift in initial parameters with respect to the theoretical ones. From 50% up to 130%, the LMA improves the initial $error$. More than 130% initial parameter shift could be considered, but it is important to note that the SABR's parameters are bounded, e.g. ρ is upper and lower bounded. For the purpose of this work, is sufficient to know from which shift on will the LMA actually improve the fitting with respect to the $error$ (or more informally, from which point on will it have to work). For that reason, in our fittings we input initial parameters that are 50% shifted from the real values of the simulation.

4.3 LMA Results

All throughout this work the simulated quotes correspond to simulation of European call option using, unless specified otherwise, the following parameter values:

- $T = 15$ (time to maturity, in years);
- $f_0 = 0.0801$ (foward price at time $t = 0$);
- $\alpha = 0.05$ (SABR parameter α);
- $\beta = 1$ (SABR parameter β);
- $\rho = -0.33$ (SABR parameter ρ);
- $\nu = 0.25$ (SABR parameter ν)
- $D = 1$ (Discount factor).

For fitting analysis it is not relevant what type of option we are inputting into the fitter, as the algorithm only considers the implied volatility. The specific value of f_0 can also be disregarded in our analysis since all strikes are chosen relative to f_0, and therefore we will only refer to k, the relative strike. The time to maturity T we chose is quite large. This choice will affect the smile in the sense that in-the-money quotes will have greater implied volatility than out-the-money quotes (see Duque and Lopes [15]), $k < 1$ and $k > 1$ respectively in our case. Since we are considering the smile interpolation methods' performance, the shape of the smile will not greatly impact the testing and for that reason we are confident that the maturity choice will not impact our results. The discount factor $D = 1$, or in other words, the assumption of the fixed interest rate $r = 0$, is chosen to simplify quote generating, and the conclusions of this work may still be applicable to quotes under conditions with discount.

Financial option quotes can be easily simulated using eqs.2.6 and 2.7 for SABR's forward price of the underlying asset and its volatility, respectively. To discretize these equations, Euler's method for solving ordinary differential equations proved to be very effective and handle the task. We found that a way of introducing more $error$ on the quotes' implied volatility when simulating the quotes, was to consider a low number of processes $\{f(t), t \in [0,T]\}$ and $\{\alpha(t), t \in [0,T]\}$ for each quote in order to compute $E[f(T)]$. In our case, and considering our SABR parameters and option specifics, we found that 10^3 of the processes above for option price estimation would lead to $error \approx 0.008$. The larger the number of processes used in $E[f(T)]$ for each option, the closer to the Berestycki *et. al* [2] SABR volatility smile the options' implied volatility would be. This result proves that the *formulae* in Berestycki *et. al* correctly describes SABR's dynamics for our considered maturity.

4.3.1 Types of quotes' samples

Given that our Bayesian inference method relies on the assumption that the quotes are at-the-money (or at least near $k = 1$) we decided to produce three types of quotes regarding k:

- full range quotes (FR): k is randomly generated from the distribution $N(1, 0.3)$, considering only $k \in]0, 2[$;
- at-the-money quotes (ATM): k is randomly generated from the distribution $N(1, 0.1)$;
- wing quotes (W): k is randomly generated from either $N(0.7, 0.05)$ or $N(1.3, 0.05)$, with 50% probability each.

Each sample of quotes will contain 20 quotes. Examples of these types of samples of quotes and subsequent LMA fittings can be seen in fig.A.2[2].

[2] In all figures we denote by "theoretical SABR" the curve we obtain when we fix the SABR parameters specified in the beginning of the section

We then proceeded to take into consideration fittings when the quote sample was contaminated with outliers to test the robustness of the fitting procedure. The way we generate a sample with outliers is to create a sample in the way described above (for whatever quote type we choose to consider) and then selecting a quote with 20% probability and inflating its implied volatility by $+0.1$ (see fig.A.3).

We then generated a 100 samples of quotes for each of the types of samples with and without outliers, as described above (full range, ATM and wing quotes).

4.3.2 SABR Parameter results

Performing LMA fittings to these 100 samples generates 100 estimates for the parameter set $\{\alpha, \rho, \nu\}$. For a question of presentation, we illustrate the results obtained for the 100 values of the estimate of each parameter as a form of histogram, for the 3 types of quotes, with and without outliers. The results of $\hat{\alpha}$, $\hat{\nu}$ and $\hat{\rho}$ of these 100 fittings are displayed in figs. A.4, A.5 and A.6, respectively.

4.3.2.A α parameter estimation

Fig. A.4 shows that this estimation method tends to over estimate α in all situations, with and without outliers. The presence of outliers impacts significantly on the histograms, with a clear shift towards larger values of α. Again, α will impact on the shape of the smile, lifting it or lowering it. The contamination of the results lifts the smile, which explains partially this over estimation, in the contaminated case.

In table 4.4 we see that this method will produce mean estimations of α close the theoretical value, i.e. $\alpha = 0.05$. The ATM being the one which is farthest away from this value, with $\overline{\hat{\alpha}} = 0.0596$, and with the remaining types of quotes, FD and W, with similar results as $\overline{\hat{\alpha}} = 0.0538$. It is also remarkable that the standard deviation of $\hat{\alpha}$ in the FD case is greater than the other two. This comes at the expense of the greater amount underestimation cases of α, in comparison with the other two types, which also can be seen in fig. A.4.

When the sample is contaminated the results are quite different. As we can see in table 4.5, the mean α estimations closest to the real value is in the FD case, in which $\overline{\hat{\alpha}} = 0.0619$. The ATM and W cases result in $\overline{\hat{\alpha}} = 0.0646$ and $\overline{\hat{\alpha}} = 0.0631$, respectively. We conclude that the presence of outliers will impact the mean of estimations of α using the LMA.

The difference in $\overline{\hat{\alpha}}$ due the presence of outliers of this value is +0.0081, +0.0050 and +0.0093 for the FD, ATM and W cases, respectively. While it appears that the presence of outliers seems to have the least impact in the ATM case, we note that both with and without outliers it is in this case that the LMA overestimates parameter α the most.

4.3.2.B ν parameter estimation

From fig. A.5 we see that $\hat{\nu}$ is biased, it shows a high frequency in the $[0.06; 0.12[$ bin for all types of quotes, with and without outliers. Its the second largest bin that changes between cases, we can see that this bin is higher in $\hat{\nu}$ for the quotes with the presence of outliers. Despite the bias of the observations, the mean for the uncontaminated quotes is close to the theoretical value of ν, meanwhile, when in the presence of outliers, the mean is higher than expected. These observations are supported by the results in tables 4.4 and 4.5.

The results with respect to this parameter ν are not the most conclusive as the standard deviation of the estimations is of the order of the value parameter. We can see that, for all cases with outliers, the histograms show a distinct 2 frequent classes, neither one containing the true value, except in the ATM case. Fig. A.5 also shows that ν is often under estimated. We note that the underestimation of $\hat{\nu}$ produces a "flat" smile.

We would like to note that, as is shown in tables 4.4 and 4.5, ν is often underestimated in the ATM case.

4.3.2.C Relationship between ν and α estimations

In the case of wing quotes it is expected for the higher $\hat{\nu}$ to produce a decrease in the values of $\hat{\alpha}$. This in fact happened throughout all the types of quote simulations. From fig.4.3 we can see that when $\hat{\rho} < 0$, $\hat{\alpha}$ and $\hat{\nu}$ have negative correlation, their covariance matrix being

$$Cov(\hat{\alpha}, \hat{\nu}) = \begin{bmatrix} 0.0001615 & -0.001048 \\ -0.001048 & 0.01919 \end{bmatrix}, \tag{4.2}$$

considering all 587 quotes (excluding $\hat{\rho} > 0$ estimates). This relation can be also seen in fig. 4.3.

4.3.2.D ρ parameter estimation

The positive values of ρ in the presence of wing quotes is also to be expected. This reflects in skewing the smile in such a way that lower the $k < 1$ implied volatilities and raises the $k > 1$ ones. It affects the smile's overall shape but we think it is not that the reason we see these high values of ρ but rather an imbalance of quote distribution on each "wing" that $\hat{\alpha}$ and $\hat{\nu}$ cannot account for, e.g. both wings have more quotes with lower values of k.

4.3.2.E Relationship between ρ parameter estimation and (α, ν) pair estimations

When ρ is positive it changes the shape of the smile dramatically and not only its skew, therefore it will affect the estimated pair ($\hat{\alpha}$, $\hat{\nu}$). In fig.4.3 we can see just how when $\hat{\rho} > 0$ very few estimates fall

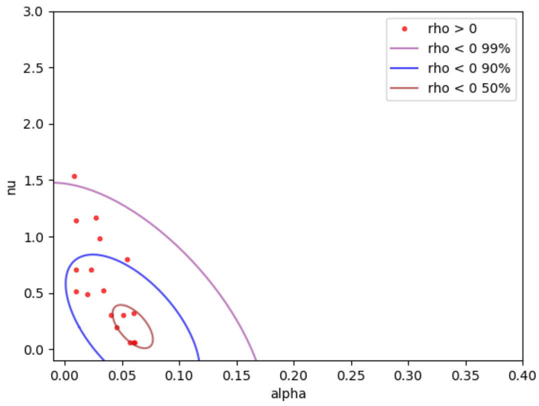


Figure 4.3: ($\hat{\alpha}$, $\hat{\nu}$) estimates when $\hat{\rho} > 0$. Confidence regions of ($\hat{\alpha}$, $\hat{\nu}$) when $\hat{\rho} < 0$. 17 points.

inside the 50% confidence region[3] of (α, ν) of the estimates where $\hat{\rho} < 0$, which constitutes the large majority of estimates and is consistent with $\rho = 0.33$ used to generate the testing quotes. Only 17 out of 600 estimates produce $\hat{\rho} > 0$. It is our understanding that in these cases the LMA converges to local *minima* due to uneven spread of quotes' relative strike. This phenomenon requires additional study and we decided not to represent these simulations' results in tables 4.4 and 4.5.

The results in tables 4.4 and 4.5 are representative of figs. A.4, A.6 and A.5, with the elimination of the cases where $\rho > 0$.

4.4 Bayesian Inference Results

A similar scheme as the one described above for the LMA was executed for the Bayesian inference method, with the exception that the values of the *a priori* distribution of α needed to be estimated before performing the Bayesian inference on the quote samples.

4.4.1 Preliminary phase

We called this the preliminary phase of the Bayesian inference method, see algorithm B.1, and consisted in generating 40 full range quotes, with no outliers, then performing 20 LMA fittings to unique quote samples. This phase is common to every of the 100 fittings used in calculating $\mu_\alpha = \overline{\hat{\alpha}}$ and $\sigma_\alpha = \sigma(\hat{\alpha})$.

In the end we consider $\alpha \sim N(\mu_\alpha, \sigma_\alpha)$. We also consider as a Bayesian inference fitting result $\hat{\rho} \equiv \overline{\hat{\rho}}$ and $\hat{\nu} \equiv \overline{\hat{\nu}}$ from this preliminary phase, given that this method will only produce an estimate for the distribution of α.

[3]We assume a bivariate normal distribution for the two parameters to construct these regions using the results from eq.4.2

4.4.2 Bayesian procedure

After the preliminary phase, similarly to the LMA fitting, we consider 20 quotes from a given quote sample type and perform Bayesian inference on each one successively; see algorithm B.2. In the end we have the *a posteriori* distribution for α. Examples of the fittings are displayed in figs. A.7 and A.8.

With only one parameter to account for, α, we displayed the results in a similar fashion as above, but in this case we also display the results of $\hat{\alpha}$ from LMA fittings for comparison; see figs. A.9 and A.10.

4.4.3 Bayesian Inference parameter s

All in all, our method still performed well and provided us with some insight to its applicability. The most remarkable characteristic of it being that when making a decision on the distribution $\sigma_{BS} \sim N(\alpha, s)$, the parameter s will control the variability of the shape allowed.

As we need to choose a specific value for s, we have chosen to set $s = 0.07$. The reason for this choice is not bullet-proof, meaning that we do not have a clear argument for that. We chose such a value as a kind of middle term/compromise: choosing s too small would lead to results too dependent on the presence of outliers (as fig. 4.4 shows, with the curve corresponding to $s = 0.01$ shifting towards the outliers, when compared with the theoretical one). On the contrary, choosing s too large would not provide us enough flexibility for the estimation of α.

Clearly this question would deserve a more in-depth study, which is beyond the scope of this work.

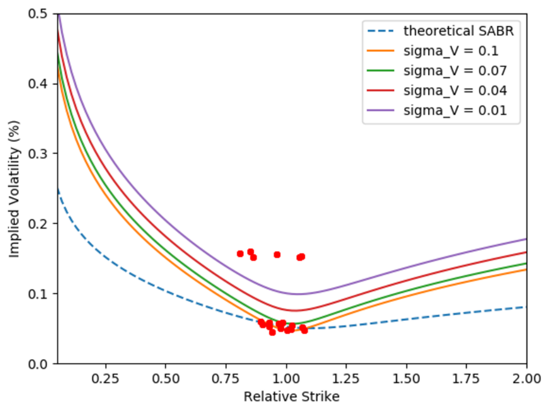


Figure 4.4: Illustration of Bayes decision s influencing the robustness of α to ATM outliers. $\hat{\rho} = -0.19215$, $\hat{\nu} = 0.46975$

4.5 Bayesian Inference and LMA comparison

As we can see from fig. A.9, the results of the Bayesian estimation when no outliers are present in the sample lead to quite concentrated *a posteriori* distribution for α, given the quotes. In all cases the estimated *a posteriori* mean of α is around 0.05; see also table 4.6.

When we consider the contaminated sample this value increases, which is as expected (as all the contaminated values are in the direction of larger implied volatility). Comparing both estimation methods (LMA and Bayesian), the results are interesting and encouraging. From fig. A.10 we see that the LMA method seems to be more sensitive to the presence of outliers and thus less robust than the Bayesian one. The difference in $\hat{\mu}_\alpha$ when outliers are introduced is $+0.0044$, $+0.0030$ and $+0.0049$, for the FD, ATM and W cases, respectively. Comparing these values with the ones of the LMA, which are larger for all cases, we conclude that the Bayesian inference method is more robust in the presence of outliers when it comes to the estimation of α.

		mean	5-percentile	95-percentile	standard deviation
FD	$\hat{\alpha}$	0.0538	0.017	0.0678	0.0151
	$\hat{\rho}$	-0.344	-0.6249	-0.1606	0.198
	$\hat{\nu}$	0.2014	0.1069	0.5173	0.1373
ATM	$\hat{\alpha}$	0.0596	0.0483	0.0672	0.0065
	$\hat{\rho}$	-0.3841	-0.795	-0.1637	0.2388
	$\hat{\nu}$	0.1541	0.0733	0.2615	0.0638
W	$\hat{\alpha}$	0.0538	0.0484	0.0579	0.0076
	$\hat{\rho}$	-0.3748	-0.6672	-0.1767	0.2114
	$\hat{\nu}$	0.1819	0.1088	0.4023	0.1363

Table 4.4: Results of 100 LMA fittings from 100 samples (20 quotes each). Quotes with no outliers.

		mean	5-percentile	95-percentile	standard deviation
FD	$\hat{\alpha}$	0.0619	0.0355	0.082	0.015
	$\hat{\rho}$	-0.3939	-0.6644	-0.169	0.1917
	$\hat{\nu}$	0.2302	0.0937	0.3845	0.1327
ATM	$\hat{\alpha}$	0.0646	0.0443	0.0805	0.0125
	$\hat{\rho}$	-0.3528	-0.5982	-0.1683	0.1938
	$\hat{\nu}$	0.1909	0.0996	0.3507	0.1074
W	$\hat{\alpha}$	0.0631	0.038	0.0804	0.012
	$\hat{\rho}$	-0.3476	-0.6049	-0.1818	0.1685
	$\hat{\nu}$	0.2462	0.0787	0.642	0.1942

Table 4.5: Results of 100 LMA fittings from 100 samples (20 quotes each). Quotes with outliers.

		mean	min	max	standard deviation
FD	$\hat{\mu}_\alpha$	0.0553	0.0377	0.0695	0.0101
	$\hat{\sigma}_\alpha$	0.0068	0.0026	0.013	0.0037
ATM	$\hat{\mu}_\alpha$	0.053	0.0333	0.0659	0.0105
	$\hat{\sigma}_\alpha$	0.0064	0.0025	0.0129	0.0036
W	$\hat{\mu}_\alpha$	0.0524	0.0313	0.0676	0.0123
	$\hat{\sigma}_\alpha$	0.006	0.0025	0.0125	0.0034

Table 4.6: Results of 100 Bayes inference estimations from 100 samples (20 quotes each). Quotes with no outliers.

		mean	min	max	standard deviation
FD	$\hat{\mu}_\alpha$	0.0597	0.0423	0.0756	0.0113
	$\hat{\sigma}_\alpha$	0.0067	0.0023	0.0129	0.0037
ATM	$\hat{\mu}_\alpha$	0.056	0.0304	0.0709	0.0125
	$\hat{\sigma}_\alpha$	0.0067	0.0022	0.0131	0.0038
W	$\hat{\mu}_\alpha$	0.0573	0.0357	0.0711	0.0107
	$\hat{\sigma}_\alpha$	0.0058	0.0025	0.0125	0.0033

Table 4.7: Results of 100 Bayesian inference estimations from 100 samples (20 quotes each). Quotes with outliers.

5

Conclusion

Contents

This work aimed to provide an alternative to a fitting algorithm for the SABR model's volatility smile fitting to financial market option quotes that performed better in the presence of outliers. We found that our Bayesian inference approach to fitting the SABR's volatility smile works well in the presence of outliers, better in fact than the LMA which can produce undesirable results in some settings.

5.1 The LMA for Fitting SABR's Volatility Smile to Option Quotes

The Levenberg-Marquardt algorithm with appropriate initial estimates is a very fast and effective way of fitting SABR's volatility smile for a well behaved set of option quotes. Reasons we identified for this algorithm not to produce reasonable parameter estimates are: uneven distribution of quotes' strikes in the sample and the presence of outliers. An unreasonable result in our case came in the form of (α, ρ, ν) estimates where $\rho > 0$. We simulated data with negative correlation coefficient ρ between processes $\{W_f(t), t \in [0, T]\}$ and $\{W_\alpha(t), t \in [0, T]\}$. A positive estimate $\hat{\rho}$ skews the smile, and this comes at the expense of overestimating the parameters ν and α to fit the sample, as can be seen in fig.4.3.

5.2 Bayesian Inference Method for SABR's α Estimation

Bayesian inference allows us to make a decision on the samples distribution $\sigma_{BS} \sim N(\alpha, s)$ and using Bayes' theorem we'll have $\alpha \sim N(\mu_\alpha, \sigma_\alpha)$. When a new quote's implied volatility $\sigma_{BS} = v$ is observed, we have an *a posteriori* distribution of α. This is our α estimate in this method. The way we compare it to the LMA estimations $\hat{\alpha}$ is to consider μ_α, from the distribution of α.

This method relies on informed user's choices, i.e. the value of s is set by the user and will have a high impact on how the *a posteriori* distribution of α will tolerate a presence of outliers (fig.4.4). This is a very important characteristic of this approach and provides the user with flexibility on choosing how the method reacts to quotes not compliant with the SABR's parameters that are set before the observation, i.e. $\hat{\rho}$, and $\hat{\nu}$.

The main limitation of this method is the estimation of SABR parameters ρ and ν. These are estimated using historic data and some optimization algorithm's estimations of ρ and ν, and this can hamper the reactiveness of our approach to some cases. The estimation of the pair (ρ, ν) is also dependent on the performance of the algorithm used to fit the SABR's volatility smile to the historical data in the preliminary phase. This choice has to be considered depending on the samples characteristics and overall algorithm performance. In our case we chose the LMA, and it faired well with the task, and even if some historic smile estimations were less good, the estimates $\hat{\rho}$, $\hat{\nu}$ were taken as mean values of all fittings and could accommodate for that.

Due to the nature approach's assumptions we would recommend using this method for a market where outliers in the at-the-money region are observed. Namely, in the fast electronic trading, where we do not wish to calibrate the model in its totality, but only partially adjust the α parameter. This approach allows us to "ignore" some of the ill placed trades during the adjustment process.

5.3 Possible Extensions of the Bayesian Approach

The method could be implemented for α with less restrictive assumptions, starting by considering $I^0(k)$ where $k \neq 1$. This step involves more accurate computation for a relationship of α with I. The Bayesian inference considered only on α requires a good fitting algorithm for the other parameters in the preliminary phase. In our case that is LMA, an algorithm that has its flaws, and its results, as discussed before, are highly subjective to data quality (distribution on k, mean $error$, etc.) and so, some bad fittings on parameters ρ and ν in the preliminary phase may have substantial impact on the Bayesian inference on α.

It would also be possible to perform Bayesian inference on the parameters ρ and ν. This is a more demanding task given that these parameters do not have a simple relationship with SABR implied volatility formula as does α.

Bibliography

[1] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward, “Managing smile risk,” *Wilmott Magazine*, vol. 1, pp. 84–108, 01 2002.

[2] H. Berestycki, J. Busca, and I. Florent, “Computing the implied volatility in stochastic volatility models,” *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1352 – 1373, 07 2004.

[3] R. Merton, “The theory of rational option pricing,” *Bell J Econ Manage Sci*, vol. 4, pp. 141–183, 03 1973.

[4] T. Björk, *Arbitrage Theory in Continuous Time*, 12 2019, ch. 7.4.

[5] F. Black and M. Scholes, “The pricing of options and corporate liabilities,” *Journal of political economy*, vol. 81, no. 3, pp. 637–654, 1973.

[6] B. Civek and S. Kozat, “Efficient Implementation Of Newton-Raphson Methods For Sequential Data Prediction, volume = PP, journal = IEEE Transactions on Knowledge and Data Engineering, doi = 10.1109/TKDE.2017.2754380,” 01 2017.

[7] J. Obloj, “Fine-tune your smile: Correction to Hagan et al,” *Quantitative Finance Papers*, vol. 35, 08 2007.

[8] C. Crispoldi, G. Wigger, and P. Larkin, *SABR and SABR LIBOR Market Models in Practice: With Examples Implemented in Python*. Springer, 2016.

[9] P. Henry-Labordère, *Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing*, 01 2008.

[10] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[11] H. Gavin, “The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems,” 2013.

[12] P. Gauthier and P.-Y. Rivaille, "Fitting the smile, smart parameters for SABR and Heston," *SSRN Electronic Journal*, 10 2009.

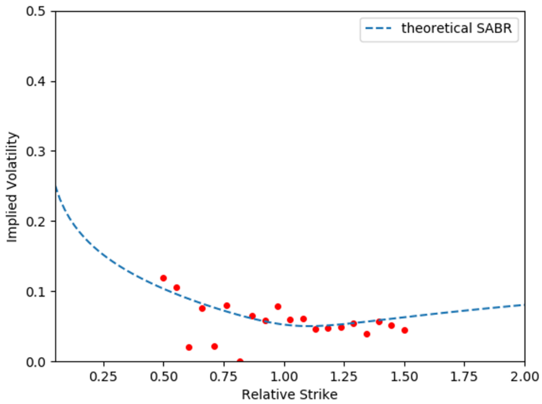
[13] C. D. Paulino, M. A. Amaral Turkman, B. Murteira, and G. L. Silva, *Estatística bayesiana*, 2nd ed. Lisboa: Fundação Calouste Gulbenkian, 2018.

[14] G. West, "Calibration of the SABR model in illiquid markets," *Applied Mathematical Finance*, vol. 12, no. 4, pp. 371–385, 2005.

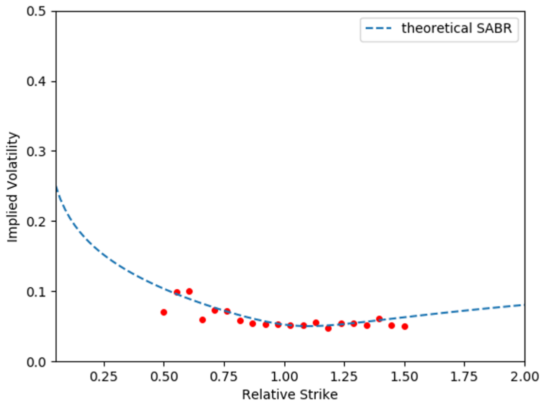
[15] J. Duque and P. Lopes, "Maturity and volatility effects on UK smiles or dying smiling?" *Portuguese Economic Journal*, vol. 2, pp. 173–193, 12 2003.

A

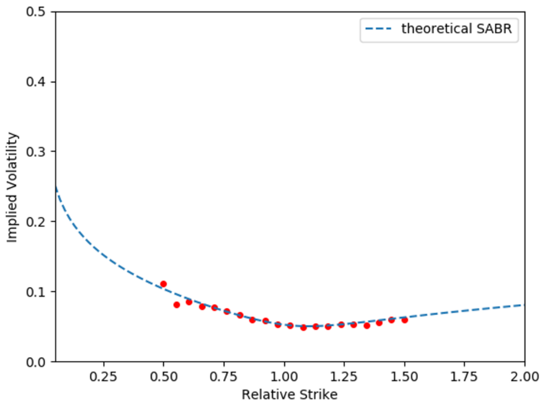
Figures and Plots



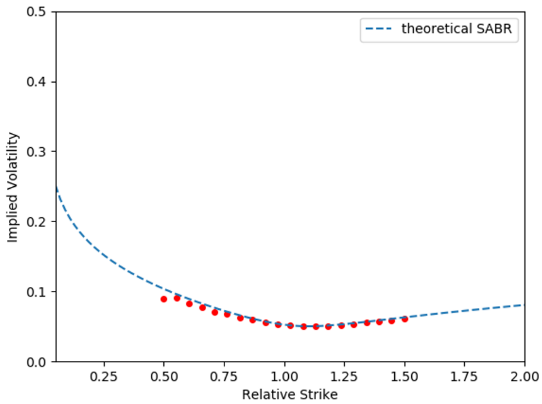
(a) 10^2 process simulations per quote.



(b) 10^3 process simulations per quote.



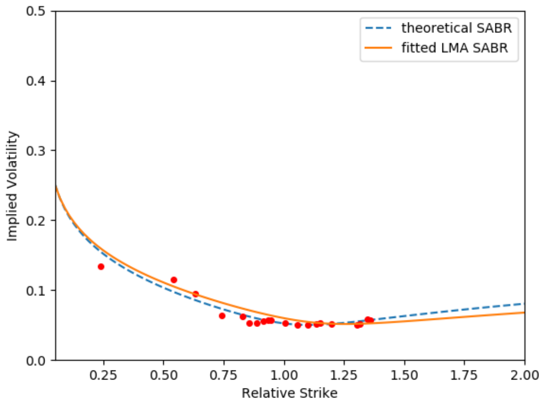
(c) 10^4 process simulations per quote.



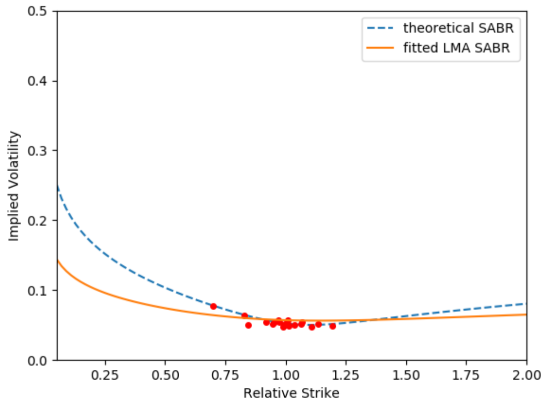
(d) 10^5 process simulations per quote.

Figure A.1: Quote simulation with varying number of f_t and α_t processes simulations with the parameters presented in section 4.3

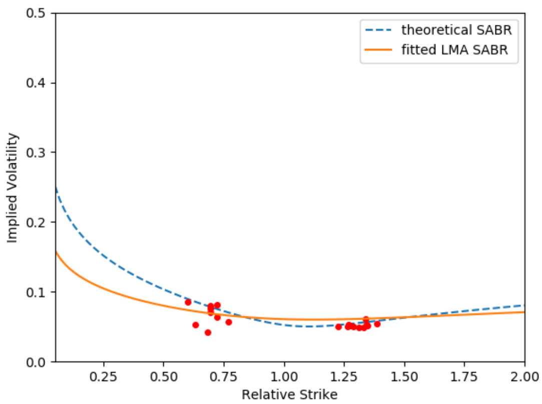
.



(a) full range quotes

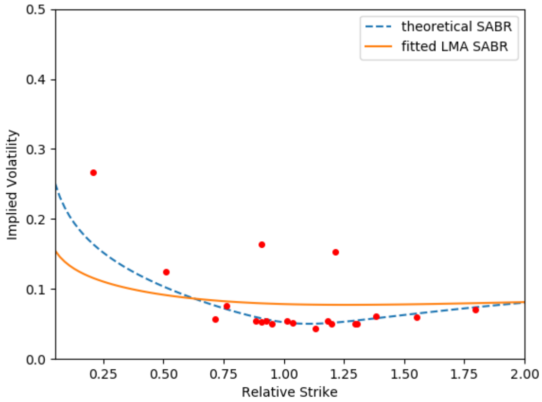


(b) ATM quotes

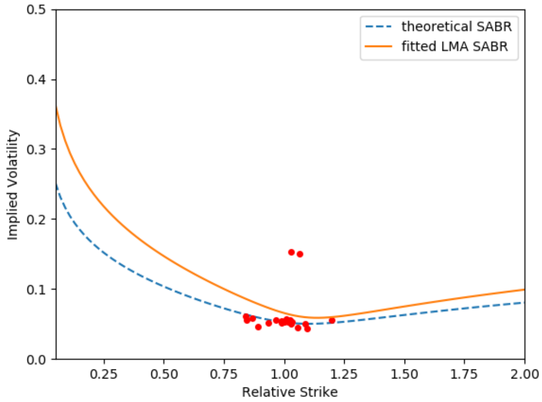


(c) wing quotes

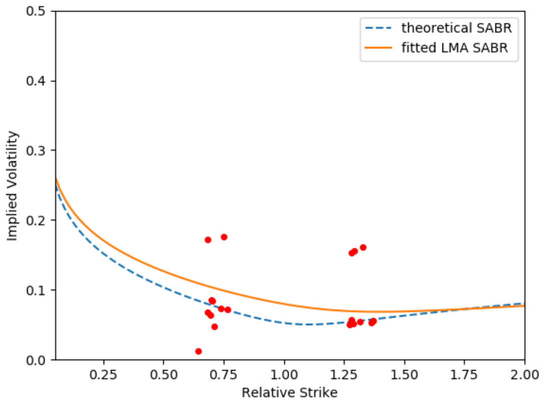
Figure A.2: Example LMA fitting result, no outliers



(a) full range quotes

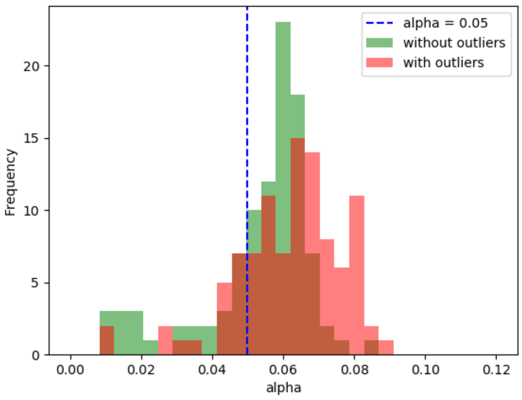


(b) ATM quotes

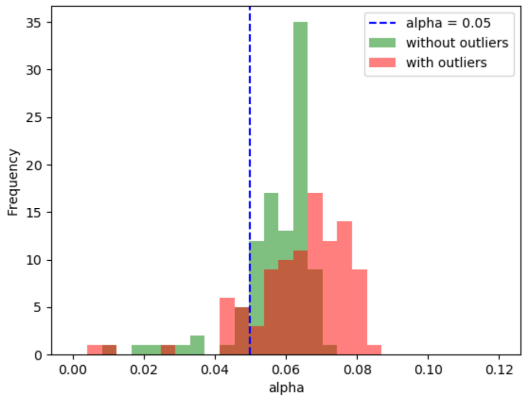


(c) wing quotes

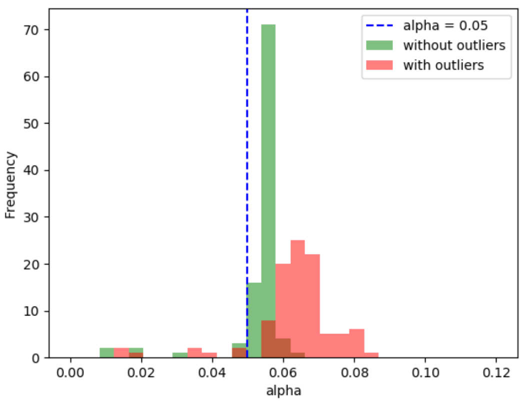
Figure A.3: Example LMA fitting result, with outliers.



(a) full range quotes

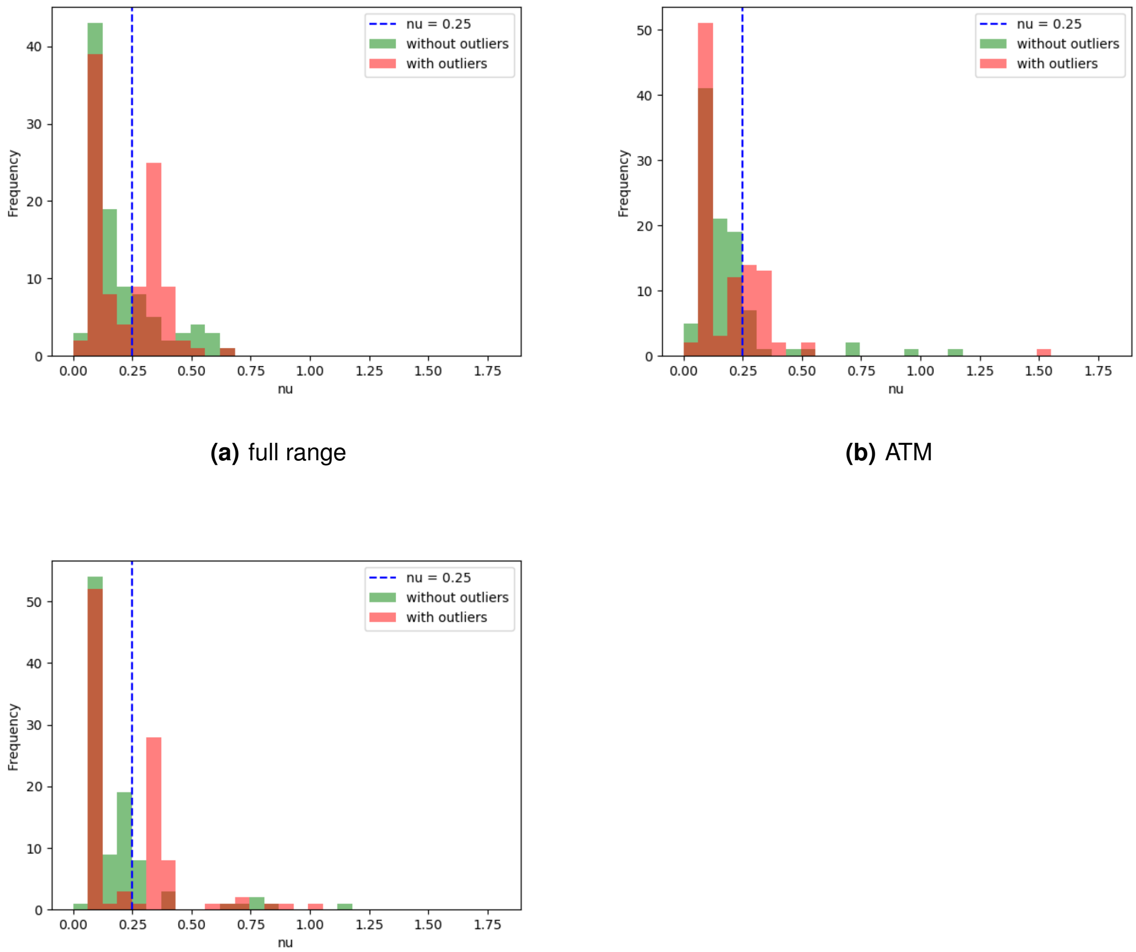


(b) ATM quotes



(c) wing quotes

Figure A.4: Results of $\hat{\alpha}$ from 100 LMA fittings following 100 quote simulations (20 quotes each).

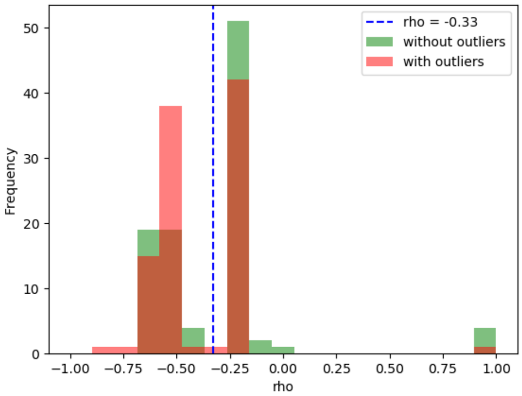


(a) full range

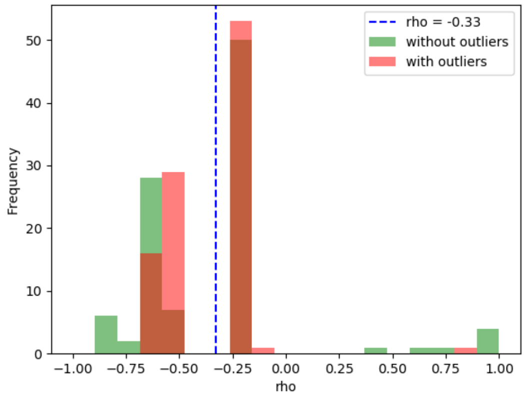
(b) ATM

(c) Wings

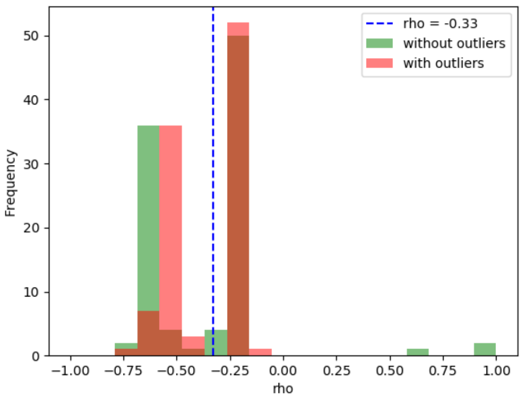
Figure A.5: Results of $\hat{\nu}$ from 100 LMA fittings following 100 quote simulations (20 quotes each).



(a) full range quotes

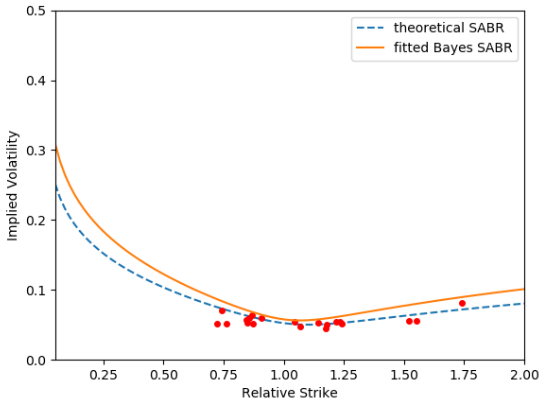


(b) ATM

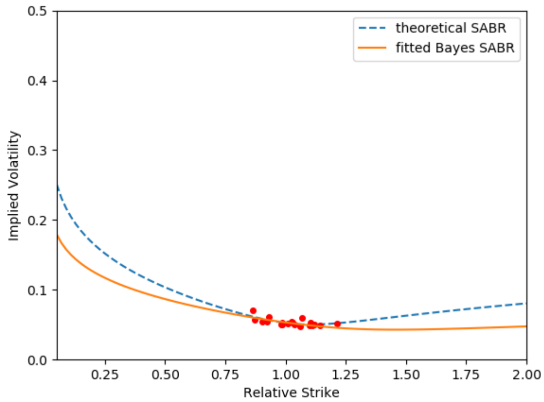


(c) Wings

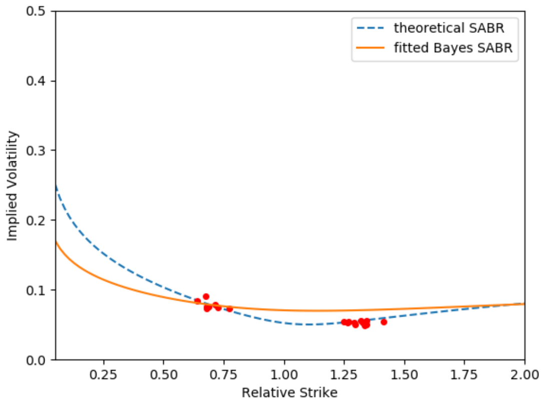
Figure A.6: Results of $\hat{\rho}$ from 100 LMA fittings following 100 quote simulations (20 quotes each).



(a) full range quotes

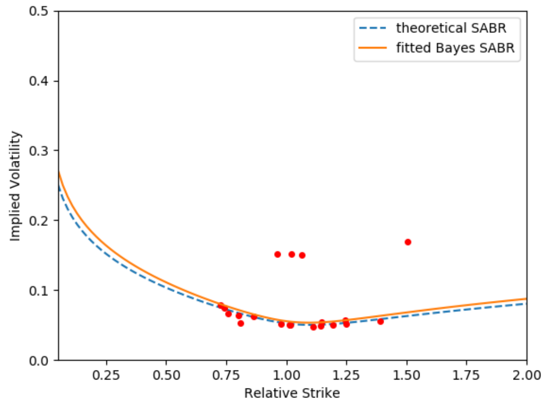


(b) ATM quotes

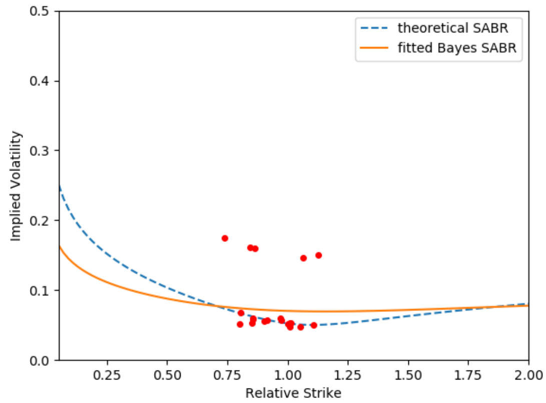


(c) wing quotes

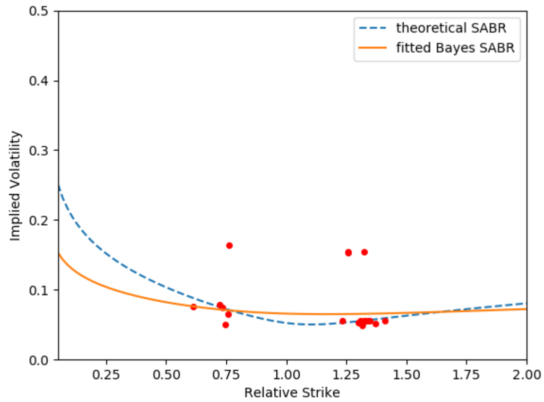
Figure A.7: Example of Bayesian inference results, no outliers



(a) full range quotes

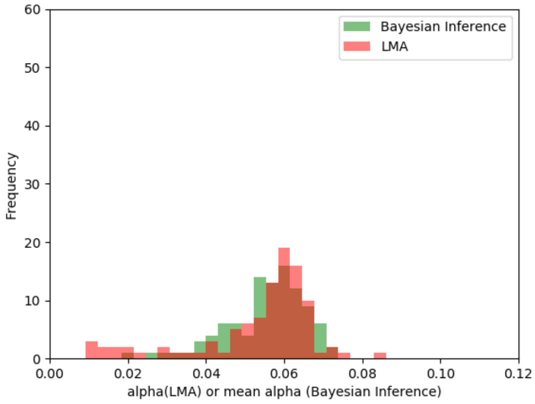


(b) ATM quotes

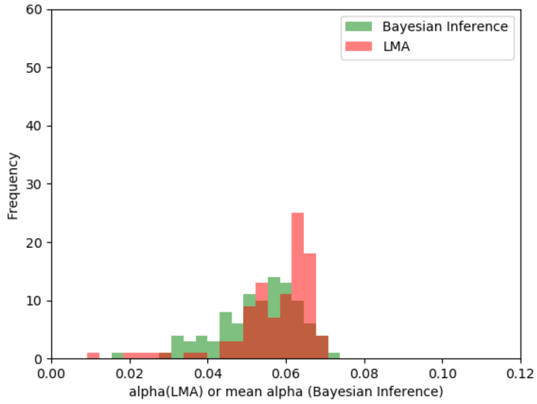


(c) wing quotes

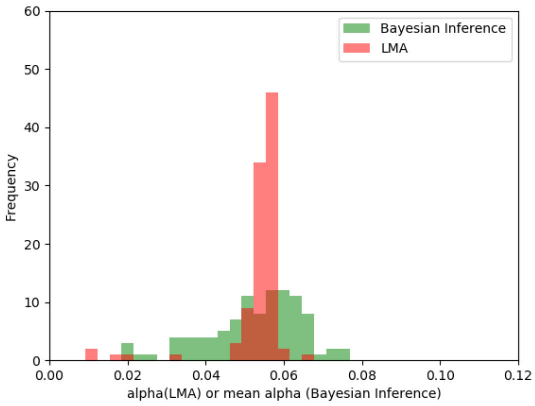
Figure A.8: Example of Bayesian inference results, with 20% of outliers.



(a) full range quotes

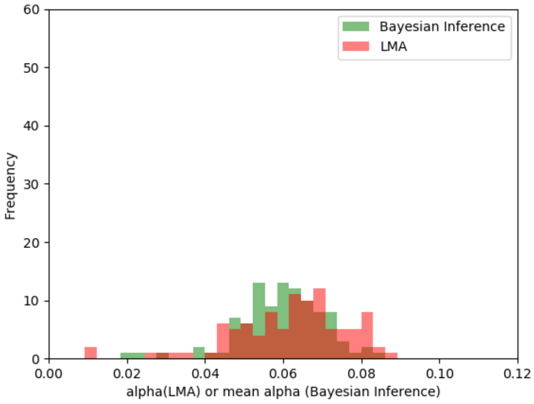


(b) ATM quotes

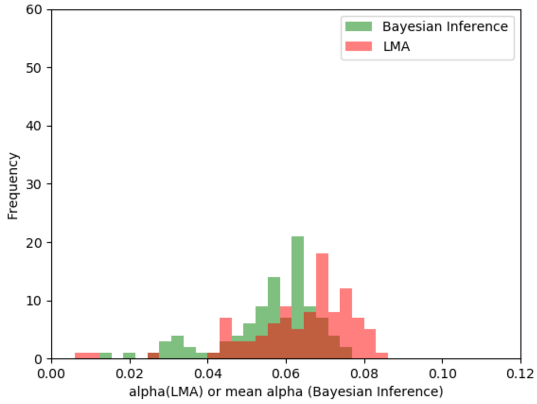


(c) wing quotes

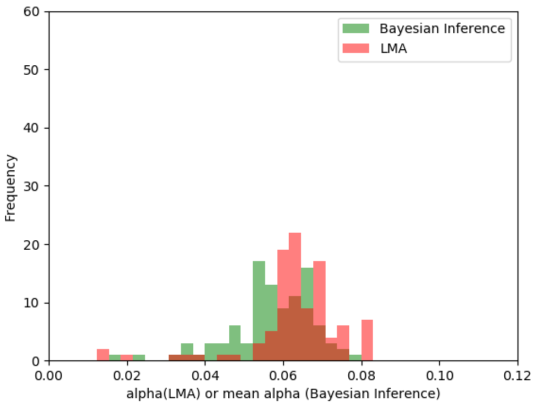
Figure A.9: Results of $\hat{\mu}_\alpha$ from 100 Bayeians inference estimations following 100 quote simulations (20 quotes each), plus a comparison with the LMA's $\hat{\alpha}$ results.
Quotes with no outliers.



(a) full range quotes



(b) ATM quotes



(c) wing quotes

Figure A.10: Results of $\hat{\mu}_\alpha$ from 100 Bayesian inference estimations following 100 quote simulations (20 quotes each), plus a comparison with the LMA's $\hat{\alpha}$ results.
Quotes with outliers.

B

Pseudo-Code

B.1 Preliminary Phase

Algorithm B.1: Preliminary parameter estimation

Result: $\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$, $\hat{\rho}$, $\hat{\nu}$
Quotes $\leftarrow$ 40 simulated quotes
for $1 \leq i \leq 20$ **do**
  quotesToFit $\leftarrow$ 20 quotes from Quotes starting from the ith
  α_i, ρ_i, ν_i $\leftarrow$ Levenber-Marquardt(quotesToFit)
end
$\hat{\mu}_\alpha \leftarrow \overline{\alpha}$
$\hat{\sigma}_\alpha \leftarrow \sigma(\alpha)$
$\hat{\rho} \leftarrow \overline{\rho}$
$\hat{\nu} \leftarrow \overline{\nu}$

B.2 Bayesian procedure

Algorithm B.2: Bayesian procedure

Result: $\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$
quotes $\leftarrow$ 20 quotes of some type
$\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$$\leftarrow$ $\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$ from preliminary phase
for $0 \leq$ i < 20 **do**
  $\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$ $\leftarrow$ BayesianInference(quotes(i), $\hat{\mu}_\alpha$, $\hat{\sigma}_\alpha$)
end